

Utilità diagnostica: la fase pre-preanalitica

P. Cappelletti

Laboratorio di Patologia Clinica, Dipartimento di Medicina di Laboratorio, Azienda Ospedaliera S. Maria degli Angeli, Pordenone

Riassunto

Per rispondere adeguatamente alla sua natura, cioè “fornire informazioni cliniche”, la Medicina di Laboratorio deve governare l'intero processo che la origina, attiva ed utilizza. Il punto chiave è l'interfaccia clinica-laboratorio, ove si svolge l'attività predominante delle fasi pre e post analitiche. La fase pre-preanalitica contiene tre passaggi essenziali: la formulazione del quesito clinico, la selezione del test e la richiesta dell'esame.

Nella formulazione del quesito clinico, apparentemente banale, si scaricano le incertezze e le contraddizioni concettuali e pratiche della visione metodologica della diagnosi medica, della complessità del processo diagnostico e dell'ambiente clinico e sociale in cui avviene.

Nella selezione dei test la virtù cardinale è (o dovrebbe essere) l'appropriatezza. Tuttavia definire l'uso appropriato degli esami di laboratorio resta un obiettivo largamente eluso, per le difficoltà di individuarne l'essenza, circoscriverne il campo, dividerne i contorni e fondarlo su elementi provati. Sono state messe a punto metodologie di studio dell'inappropriatezza di cui si conoscono potenzialità e limiti e sono state usate per misurare il sovra-utilizzo della diagnostica di laboratorio. Si è resa evidente la centralità del concetto

di *outcome* e si è approntato un coerente background culturale e teoretico che collega indissolubilmente appropriatezza ed *evidence-based medicine*. Ma l'evidenza in Medicina di Laboratorio è debole, a causa del problema dello standard di riferimento, di progetti di studi primari inadeguati, di difficoltà negli studi di *outcome*, di problemi e di *bias* delle revisioni sistematiche in Medicina di Laboratorio. In ogni caso si studiano le strategie di intervento, considerate efficaci solo se combinazioni di interventi educativi, amministrativi e di feed-back, e se ne pesa l'effetto in termini di volume di attività, di costi e di efficacia clinica.

Il ridisegno della richiesta è uno dei cavalli di battaglia nella lotta per l'appropriatezza. È importante accompagnare la diffusione dei formati di richiesta con adeguata condivisione degli aspetti educazionali e di ricerca e adeguati feed-back e *audit*. Si tratta di una incomparabile occasione di colloquio con i clinici e di condivisioni degli obiettivi generali di salute. D'altra parte se il Laboratorio non conosce il motivo della richiesta, la logica diagnostica e le problematiche della decisione clinica susseguenti alle indagini, non è in grado di rispondere adeguatamente al quesito clinico sottostante.

Summary

Diagnostic effectiveness: the pre-preanalytic phase

To answer to its own nature, Laboratory Medicine should maintain the governance of the total testing process. In this framework, the key is the clinic-laboratory interface, where the essential activities of Laboratory Medicine such as pre and post analytical phases occur. The pre-pre-analytical phase consists of three steps: identifying the question, selection of the appropriate test, the test order or request.

Identifying the clinical question seems an obvious matter; nevertheless, all conceptual uncertainties and discrepancies of the methodological vision of the diagnosis, the complexity of the diagnostic process, and

the issues of the clinical and societal background meet each other at that point.

The appropriateness should be the cornerstone of the test selection. Nevertheless, the definition of the appropriate use of laboratory tests remains an unreach goal, because of the difficulties in determining its essence, measuring the field, pointing out the borders, and founding it on solid evidences. Methodologies have been designed to measure the appropriateness and were used to determine the rate of overuse in the laboratory tests requests. The concept of *outcome* became the meter of measure of inappropriateness and defined the intimate relationship between appropriateness and evidence-based medicine. But the evidence in Labora-

tory Medicine is weak because of the issue of gold standard, the designs of primary studies, the problems in *outcome* studies, and biases in systematic studies and meta-analyses. However, intervention strategies have been selected, as combination of educational, administrative and feed-back techniques, and their results in terms of volume of tests, costs, and effectiveness were measured.

Redesigning the request forms is one of main tools in

the battle for appropriateness. This change should be accompanied and underpinned by sharing educational intervention and research results with clinicians. It is an incomparable opportunity of communication with the other side of the medicine. If the Laboratory does not know the reasons of clinical ordering, the diagnostic logic, the issues of clinical decision making following the laboratory testing, it is not able to answer effectively to the actual clinical question.

Dal quesito clinico alla risposta di laboratorio: la fase pre-preanalitica

Per rispondere adeguatamente alla sua natura, cioè “fornire informazioni cliniche”, la Medicina di Laboratorio deve governare l'intero processo che la origina, attiva ed utilizza. Il punto chiave è l'interfaccia clinica-laboratorio¹, amplificata a dismisura dalla struttura a rete, rivolta al clinico ed al paziente, coinvolgente tutta l'equipe, caratterizzata dalla qualità della comunicazione e dalla sua coerenza con l'obiettivo clinico.

La obbligatorietà per la Medicina di Laboratorio di impegnarsi in tutto il percorso di richiesta diagnostica, dalla formulazione del quesito clinico all'applicazione delle azioni conseguenti alle informazioni fornite, pertanto non esclusivamente nella fase analitica, ne è la necessaria conseguenza, prevista esplicitamente dai più recenti standard certificativi².

La più moderna ed articolata descrizione della complessità della Medicina di Laboratorio, la *Nexus vision*³, esplicitamente collaziona le elaborazioni di Lundberg (brain-to-brain loop), di Truchaud (quality loop), di Weggeman (knowledge feedback) e l'approccio GUM (Guidelines for Uncertainty of Measurements) in un modello coerente e comprensivo, basato sulle possibilità tecnologiche che consentono di digitalizzare l'ambiente diagnostico di un paziente finalizzato a supportare le future responsabilità del laboratorio: una risposta al quesito clinico nel contesto della necessità del paziente, dove “contesto” significa che la qualità deve andare oltre “la risposta corretta al paziente giusto” per “assicurarsi che il paziente sia trattato correttamente”.

Nella *Nexus vision* la classica tripartizione dell'iter diagnostico di laboratorio (fase pre-analitica, analitica, post-analitica), viene estesa a 5 fasi, dove la iniziale pre-pre-analitica richiede la descrizione del contesto del paziente e la valutazione della diagnostica di laboratorio comparata con altri strumenti diagnostici, mentre quella post-post-analitica dovrebbe contenere la descrizione del nuovo contesto del paziente, basato sull'integrazione delle nuove informazioni comunque ottenute, e una visione delle diverse interpretazioni e possibili conseguenze.

Lungo queste cinque fasi si sviluppa più compiutamente lo schema di Lundberg⁴, originariamente descritto in 9 passaggi (richiesta, raccolta, identificazione,

trasporto, preparazione, analisi, risposta, interpretazione, azione). In realtà anche l'elaborazione di Lundberg è andata maturando nel tempo e preparando l'attuale concezione di fase preanalitica suddivisa in 2 momenti concettualmente distinti (pre-pre e pre-analitica). Infatti, all'inizio degli anni '80 Lundberg sentì il bisogno di rappresentare con una immagine sintetica ed efficace un concetto apparentemente ovvio, il fatto cioè che le procedure di laboratorio sono finalizzate ad una decisione clinica, e diede a questa icona il nome di “*brain to brain loop*”, “circuitto da mente a mente”, ma 10 anni dopo⁵ presentò una versione aggiornata del suo schema, con una significativa modifica: il primo passaggio (la richiesta dell'esame) è suddiviso in 3 passi e cioè la formulazione del problema clinico, la selezione del test e la richiesta vera e propria. Burlina⁶ commentava molto positivamente l'innovazione, sottolineando la profonda aderenza della Medicina di Laboratorio italiana al concetto che queste fasi sono pienamente oggetto della Medicina di Laboratorio e l'indissolubile legame tra queste operazioni e l'interpretazione dei risultati, l'altro pilastro della sua logica diagnostica.

La formulazione del quesito clinico

La formulazione del quesito clinico è il primo momento interpretativo del medico di fronte ad un paziente che gli si rivolge con il proprio bagaglio di segni e sintomi e la propria storia clinica, quando è possibile raccoglierla, anche quando il problema che lo spinge non viene, per diverse ragioni anagrafiche, patologiche o psicologiche, chiaramente espresso e verbalizzato. La formulazione del quesito clinico appare un passaggio scontato e la sua importanza per la risposta è universalmente sottolineata, ma in realtà qui si concentrano le incertezze e le contraddizioni concettuali e pratiche della visione metodologica della diagnosi medica, della complessità del processo diagnostico e dell'ambiente clinico e sociale in cui avviene.

Secondo visioni diverse ma non confliggenti, i test diagnostici sono strumenti “usati per ridurre l'incertezza circa la presenza o assenza di una malattia” e “l'estensione tecnologica dei sensi del medico”. Per Sackett⁷ l'uso di un test diagnostico implica non solo il processo decisionale per chiarire il processo di malattia (diagnosi), ma anche per predire il futuro della malattia (prognosi) e la risposta all'intervento (monitoraggio).

Anche per Burlina⁸ i tre grandi motivi sono lo “screening” come modalità di indagine per evidenziare malattie o difetti che non sono manifesti chiaramente, la “diagnosi” come specificazione diagnostica (conferma/esclusione, differenziale, etiologica, fisiopatologica, funzionale), valutazione di gravità e formulazione di giudizio prognostico, e il “monitoraggio” come valutazione longitudinale dell’evoluzione della malattia e della terapia. Per Price⁹ i quesiti clinici generici in relazione all’uso dei test diagnostici sono: screening di malattia; *rule-in/rule-out* di una diagnosi; iniziare, aggiustare, sospendere un trattamento; definire l’efficacia e la compliance per un intervento; determinare la prognosi. Nel linguaggio di Goldschmidt¹⁰ il “contesto” è la situazione clinica del paziente, rispetto alla quale è sufficiente sapere se la richiesta di esame è motivata da screening, per diagnosi o dal monitoraggio della malattia e della terapia. In relazione a questi tre diversi contesti è possibile una interpretazione adeguata (*context-fitted*) basata, rispettivamente, sui criteri bayesiani e sulla valutazione temporale dei risultati rispetto agli screening precedenti, sui criteri bayesiani e il quadro clinico (segni e sintomi) del paziente, oppure sulla valutazione temporale dei risultati rispetto ai precedenti.

Tuttavia i motivi della richiesta di esami di laboratorio sono da più di un quarto di secolo indagati con risultati non definitivi. Nel 1980 BG Wertman et al¹¹ in un ampio studio presso i medici generali, indicavano, nella diagnosi, monitoraggio e screening i motivi più frequenti con il 37, 33 e 32% rispettivamente, mentre meno frequenti erano la presenza di dati precedenti alterati (12%), la prognosi (7%), l’educazione (2%) e le ragioni medicolegali (1%). Con una ricerca meno sofisticata ma più estesa, GD Lundberg¹² nel 1983 elencava una serie di ragioni dalle più note e già citate ad altre sempre presenti, almeno sotto traccia (rassicurazione del paziente, della famiglia del paziente, del medico; insicurezza, frustrazione, curiosità, ricerca; per guadagnare tempo; per pressione del paziente e familiari, della letteratura scientifica, di colleghi; per abitudine; per motivi amministrativi; per profitto personale o dell’istituzione). Già nel 1983 MD McNeely¹³ sottolineava il diverso peso delle 3 principali ragioni nei pazienti ambulatoriali, in cui predominano screening (40%) e diagnosi (30%) rispetto al monitoraggio (30%), in confronto ai pazienti ricoverati, nei quali il monitoraggio è largamente predominante (80%) rispetto a screening e diagnosi (10% ciascuno).

Ma, per avere una risposta che gli consenta di trarre una conclusione ed iniziare un’azione nello specifico paziente, il medico curante ha la necessità di individuare dettagliatamente il problema clinico e gli strumenti per risolverlo. Si tratta di un momento decisionale idiografico, descrivibile per esempi e studiato dalla psicologia empirica e sociologica, in cui peraltro si applicano le regole e i vincoli operativi della medicina nomotetica, derivata dalle scoperte scientifiche e codificata nosograficamente. È il punto di fusione tra costruito

biologico e concettuale di malattia, tra “malato” e “malattia”.

Non esiste, probabilmente, una descrizione generale del metodo per risolvere il problema diagnostico del medico curante, perché non esiste una teoria accettata universalmente della risoluzione dei problemi dell’essere umano. La complessità ed irriducibilità dell’evento diagnostico e prognostico ha indotto ad utilizzare per la medicina il concetto di “arte” contrapposta a “scienza”, anche se il termine è la traduzione latina del termine greco *τεχνη* (abilità pratica rispetto ad un sapere; operazione diretta da regole). Fin dall’antichità lo scontro tra la visione ippocratica e galeniana è stato il dibattito tra una logica induttiva e una deduttiva nell’indagine sulla malattia. In epoca moderna è Claude Bernard che codifica le regole del metodo scientifico in medicina sperimentale: osservazione; formulazione dell’ipotesi; deduzione e controllo. Da allora la strategia diagnostica viene di solito riferita come un procedimento ipotetico-deduttivo, ove il test di laboratorio interviene nella prima fase “semiologicamente” come segno per individuare l’ipotesi e nella seconda “nosologicamente” come elemento per provare deduttivamente l’ipotesi diagnostica¹. Tuttavia la “pseudodeduttività” della iniziale categorizzazione della malattia (etichettatura di un complesso sindromico, selezionato dal medico) e la limitata certezza ottenuta con la fase esplicitamente deduttiva, per l’assenza di leggi universali e per il raggiungimento solo di “gradi di conferma” o “probabilità induttiva” secondo Carnap, rilanciano un “neo-induttivismo” oscillante tra algoritmi diagnostici, *evidence-based medicine* e livelli probabilistici di evidenza. All’interno di questi percorsi si scontrano metodi di prova per verifiche corroborative o per confutazioni, modi di affrontare il problema diagnostico con un sapere di fondo (*tabula plena*) o senza pre-concetti (*tabula rasa*), un metodo scientifico unificato o la possibilità per le cosiddette scienze umane di un approccio ermeneutico, in grado di comprendere ed interpretare sinteticamente comportamenti ed azioni umane senza utilizzare lo schema esplicativo caratteristico delle scienze naturali.

Nell’esperienza del medico generale, tuttavia, “sono quasi esclusivamente i principianti ad usare un ortodosso quanto lento metodo ipotetico deduttivo, mentre i clinici navigati se ne servono solo nel caso si presentino problemi davvero inconsueti. Gli esperti decidono in condizioni di incertezza utilizzando per lo più un processo di riconoscimento di schemi globali...Anche se spesso non sa descrivere in cosa consista la propria competenza, l’esperto ha una conoscenza tacita del problema che va al di là delle regole codificate...”¹⁴. Nell’approccio con il paziente, l’intervallo medio di tempo trascorso dal momento in cui il paziente comincia a raccontare la sua storia al momento in cui il medico lo interrompe per la prima volta è di 18 secondi (28 se è lasciato libero di parlare) e l’82.5% delle diagnosi formulate dopo questa prima fase con-

cordano con quella finale individuata anche con accertamenti strumentali¹⁵. In uno studio su 20 Medici di Medicina Generale (MMG) veneti¹⁶, strutturato con un questionario indicante modi e pesi circa le strategie diagnostiche applicate, i risultati indicano una scarsa importanza data al “procedimento per esclusione”, tecnica vissuta come “rigida e forte” più tipica dello specialista, mentre predomina un approccio cognitivo, deduttivo-verificazionista e bayesiano, basato su anamnesi ed esame obiettivo brevi e mirati e sui criteri di semplicità e probabilità, ma anche un procedimento gestaltico (riconoscimento di quadri) del ricordo-confronto e l'importanza fondamentale della conoscenza del paziente. Più autorevolmente Gill et al¹⁷, universitari bostoniani, sostengono con aneddoti ed esempi che il processo mentale della diagnosi è inconsciamente probabilistico e bayesiano, non nel senso che i clinici calcolino esattamente il rapporto di verosimiglianza mentre esaminano un paziente ma nel senso che interpretano ciascun test positivo come “qualcosa di più suggestivo” e un test negativo come “qualcosa di meno suggestivo” per diagnosticare una certa malattia nel dato paziente, via via che i test vengono eseguiti. Nella categoria dei test diagnostici sono compresi tutti gli aspetti della storia clinica e dell'esame obiettivo e l'inconscia valutazione delle probabilità prima e dopo ogni passaggio diagnostico avviene sequenzialmente. Tuttavia il limite del processo, riconosciuto dagli Autori, sta nella relativa soggettività della definizione delle probabilità pre-test e cioè nella sintesi tra la prevalenza della malattia e il contesto specifico del paziente. All'incertezza il medico cerca di rispondere con l'uso di massime bilanciate, concorrenti e talvolta paradossali, distillate dalla saggezza della pratica clinica e dall'atteggiamento narrativo del medico: “*when you hear hoofbeats, don't think zebras*” rappresenta il prototipo di queste, l'epitome della saggezza pratica della medicina, la sua regola ermeneutica¹⁸.

Un'impostazione che tende a superare le aporie del processo diagnostico tradizionalmente descritto è quella cognitiva. Nel modello cognitivo o epistemologico il modello inferenziale del ragionamento diagnostico è descritto con i termini di astrazione dei dati in problemi, di abduzione dai problemi all'ipotesi (riconoscimento di una nosografia predeterminata, più o meno selettivo in rapporto alle informazioni disponibili), di deduzione dalle ipotesi ad informazioni non ancora controllate, di induzione come controllo eliminativo delle nuove informazioni per stringere il campo delle ipotesi precedentemente determinate. Si tratta di processo ciclico che conduce ad una sola o poche ipotesi classificate secondo plausibilità decrescente.

Nel caso della medicina l'abduzione può essere considerata un evento “creativo” quando l'ipotesi diagnostica viene istituita per la prima volta nella scoperta di una nuova malattia; nella pratica comune si tratta di un'abduzione più modesta: la selezione di una ipotesi in un'enciclopedia disponibile, la riscoperta di un sape-

re preesistente. Talora si sostanzia in una “abduzione automatica” secondo Peirce, una “percezione” come procedura di ri-produzione di conoscenza veloce ed incontrollata, recupero istantaneo di conoscenze già organizzate una prima volta per processi inferenziali creativi in strutture cognitive (schemi). Secondo la teoria cognitiva, gli schemi percettivi o cognitivi sono i modelli o piani attraverso i quali procediamo alla previsione, raccolta, selezione, organizzazione ed uso dei dati disponibili nella realtà ambientale che ci circonda. Poiché la percezione avviene processualmente e nel tempo, gli schemi rappresentano programmi fisico-mentali retroattivi ed automodificantesi che consentono anticipazioni di raccolta dei dati informativi e, nella rete derivante dalle integrazioni metaforiche e dalle specializzazioni percettive, l'intuizione creativa dipendente dalla capacità acquisita di integrare schemi e di separarli dal contesto reale per mezzo dell'immaginazione. Nella letteratura medica la conoscenza clinica è stata da Feltovich e Barrows in poi descritta dal concetto di *script* (copione), come una struttura di memoria organizzata a fini diagnostici. Gli script sono una variante del più generale concetto di schema che contiene, in più, l'associazione con sequenze di eventi che accadono frequentemente in quel dato ordine. La conoscenza delle malattie include informazioni sulla sequenza spazio-temporale dell'esordio e decorso di sintomi e segni. Lo script diagnostico è un insieme di attributi di varia probabilità, non esclusivi ma in relazione tra loro, che viene inconsciamente attivato, se rilevante, dalla situazione clinica di cui costituiscono un'interpretazione da verificare, un organizzatore del flusso informativo comprese le azioni da intraprendere e la sintesi delle verifiche per cui si raggiunge una sufficiente evidenza nel processo diagnostico per considerarlo concluso¹⁹.

La certezza della diagnosi non esiste; essa mantiene operativamente lo statuto di ipotesi di malattia.

L'incertezza della diagnosi dipende da molteplici fattori²⁰, in quanto è “una negoziazione che ha lo scopo di ricongiungere la visione delle componenti oggettive (patologia) e soggettive (infermità) in un processo non codificabile a priori”, per l'ambiguità di segni e sintomi, per la carenza o ignoranza di strumenti diagnostici e terapeutici legata anche al proliferare delle conoscenze, per la ridotta probabilità della consultazione medica in un certo numero di situazioni sintomatologiche, per la comune evenienza di sintomi senza diagnosi. Infatti un conto sono le prevalenze teoriche di malattia in alcuni ambiti geografici e in situazioni specifiche, un conto la prevalenza stimata nel singolo caso. In ambito anglosassone i MMG si attendono nella pratica clinica di incontrare una serie di patologie minori (24% infezioni del tratto respiratorio superiore, 13% disturbi emotivi, 8% disturbi gastro-intestinali, 8% tonsilliti, 3% otite media, 1% otite esterna, 2% infezioni delle vie urinarie, 2% lombosciatalgia acuta e 6% cronica) e di patologie maggiori (5% bronchite acuta e 1% polmonite, 0,4% trombosi coronarica e 0,2% accidente cerebro-vasco-

lare, 0,2% nuovo caso di cancro, 0,15% appendicite acuta). In ambito americano le prevalenze attese sono talora simili (4% faringite, 3% bronchite) ma talora diverse (6% ipertensione, 2,5% diabete, 2% obesità). Esse variano con la stagionalità, l'età, gli ambiti geografici interni e sono fortemente diverse da quelle attese in un pronto soccorso ospedaliero²¹. D'altra parte, nell'ambito della Medicina Generale, un dolore toracico ha 1 probabilità su 14 di essere riferito al medico, un dolore addominale 1 su 29, un disturbo gastrico 1 su 109, una cefalea 1 su 184 e una astenia 1 su 456 e la percentuale delle situazioni sintomatologiche che giungono a diagnosi definitiva è solo del 50% nel dolore toracico nel dolore addominale nei maschi e del 27% nella cefalea²⁰.

Il compito del medico è scoprire il legame aleatorio che lega segni e sintomi alla categorizzazione nosologica con un livello di ridotta incertezza tale da consentire decisioni cliniche. Il paradigma probabilistico è il modello computazionale che meglio si presta alla descrizione della diagnosi e a vincolarne la pratica a criteri espliciti di razionalità. La probabilità è un concetto complesso che include principalmente due componenti: una probabilità "oggettiva", stima basata sull'osservazione degli avvenimenti passati, e una probabilità "soggettiva", misura indipendente da osservazioni ed esperimenti, che comprende sia la probabilità a priori (probabilità di malattia a prescindere da segni e sintomi: prevalenza nella popolazione) che la misura della credenza del soggetto, entrambe parte del sapere medico e della comunicazione verbali come riferimenti alla verosimiglianza degli eventi. Nel modello probabilistico della diagnosi, il teorema di Bayes è lo standard della razionalità. Esso descrive quantitativamente la pratica diagnostica, cioè usare induttivamente nuova informazione per rifinire il giudizio su quanto si conosce, e rende esplicita e matematicamente solida una modalità di inferenza che è l'equivalente della predizione e lega le nozioni di probabilità oggettiva, evidenza e probabilità soggettiva in un costrutto coerente ed interpretabile ($\text{odds a posteriori} = \text{odds a priori} * \text{rapporto di verosimiglianza}$). L'applicabilità pratica si riferisce prevalentemente al passaggio dallo studio dei sintomi in funzione della malattia all'individuazione delle malattie in funzione dei sintomi, all'esplicita importanza della probabilità a priori nella scelta di un test diagnostico e nella valutazione matematica della sequenza del processo diagnostico dove l'esito di un test è la base per interpretare il successivo.

Un altro modello computazionale, basato su termini linguistici piuttosto che quantitativi, è la logica fuzzy (sfumata) proposta da Zadeh²² nel 1965 per trattare al meglio l'incertezza del concetto di malattia a livello di paziente, medico e società. Anche i test diagnostici hanno livelli di incertezza che appartengono al paziente (qual è il valore "normale", abituale della variabile misurata?), al medico (qual è il livello decisionale adeguato?), al contesto diagnostico (qual è il significato di un iden-

tico dato analitico nel paziente ambulatoriale o critico?). La diagnosi fuzzy è una soluzione euristica (scorcio mentale per prendere decisioni), caratterizzata da insiemi fuzzy non mutuamente esclusivi, che accetta dimensioni diverse della conclusione diagnostica: tipicità del caso, gravità della malattia, sostegno all'ipotesi in oggetto, probabilità che la diagnosi formulata sia vera. Questo modello è meno generalizzato del precedente, che offre una logica di ragionamento descrittiva del processo diagnostico.

Nella ricerca psicologica sulle decisioni in medicina il modello normativo probabilistico della diagnosi può servire per valutare la performance diagnostica in termini di euristiche e *bias*. Gli errori di ragionamento più comuni in medicina, laddove si usa il paradigma della probabilità, sono il *bias* di rappresentatività per insensibilità alla prevalenza di malattia, il *bias* di disponibilità per sovrastima di eventi familiari o recenti, il *bias* di ancoraggio dove la probabilità viene corretta (aggiustamento) per confronto, il *bias* di similarità dove il *pattern recognition* oscura la probabilità di altre alternative, il fallimento di congiunzione ove il rasoio di Occam sottovaluta le etiologie multiple, il conservativismo che non rivede le probabilità sulla base delle probabilità a posteriori originate via via negli eventi sequenziali e il *bias* di conferma o "pseudodiagnosticità" dove si procede per confermare ipotesi già robuste piuttosto che richiedere nuove informazioni che escludano ipotesi competitive.

Nonostante le incertezze in tema di diagnosi, anzi per diminuirle nella pratica clinica, sono stati suggeriti molti metodi di supporto alla diagnosi clinica, sia come regole predittive o algoritmi, costituiti da un insieme di variabili combinate in base a un dispositivo inferenziale induttivo, sia come linee guida, raccomandazioni valide in generiche situazioni cliniche, sia come percorsi critici, piani di gestione del processo diagnostico e terapeutico che raccomandano la sequenza temporale e operativa e gli obiettivi utili al paziente e alla struttura, sia come strumenti di analisi decisionale, metodo esplicito quantitativo e prescrittivo per analizzare sistematicamente scelte complesse e produrre la maggior utilità attesa, sia come "sistemi basati sulla conoscenza" dotati di un motore inferenziale e di una base dati, talora implementabili automaticamente secondo le regole delle reti neurali.

La complessità del contesto clinico e sociale rendono talora ulteriormente difficile la formulazione del quesito clinico. Ancora nel 2001 Bauchner et al²³, esaminando l'attività diagnostica e terapeutica dei pediatri americani, notavano la discrepanza tra la letteratura scientifica e i libri di testo che indicavano il processo della decisione clinica come definito matematicamente dagli strumenti bayesiani (sensibilità, specificità, valore predittivo e rapporto di verosimiglianza) e la sensazione personale di una maggiore complessità in cui 3 *domain* si sovrappongono: l'esperienza e la conoscenza del medico, le caratteristiche e i valori del paziente, l'evi-

denza clinica esterna. Il modello ha proprietà dinamiche, poiché il peso dei diversi “domini” dipende dal tipo di decisioni mediche da intraprendere: nelle condizioni urgenti ed acute le caratteristiche del paziente hanno minor peso rispetto alle conoscenze del medico e all'evidenza esterna, mentre hanno un valore preminente nella cura delle condizioni croniche. Il modello coniuga i postulati dell'evidence-based medicine (conoscenze ed esperienze del medico e migliori evidenze esterne disponibile) con la visione del paziente come partner negoziale nelle scelte che lo riguardano ed inoltre prevede un elemento che contiene e condiziona le 3 “regioni”: quelle che vengono definite “norme sociali”, talmente interiorizzate nella coscienza da essere applicate nelle decisioni senza un'esplicita riflessione. Come esempio Bauchner et al citano la diversità di obbligo alla rianimazione di un neonato di 580 grammi alla nascita nelle diverse parti del mondo.

Anche per la diagnostica di laboratorio il modello può funzionare. Nei pazienti ambulatoriali i “motivi minori” nella richiesta di esami sono sempre più evidenti. La pressione del paziente, l'invasione dello specialista, il clima culturale caratterizzato da un eccesso di informazione e da un deficit di formazione erano ancora 10 anni fa percepiti come cause importanti di richiesta dai MMG in Italia (C. Scaramuzza, Comunicazione personale, 1994). Nel 2004 una ricerca di MMG bresciani (O. Brignoli, comunicazione personale, 2005) ha mostrato, su 638 richieste di test di laboratorio, che il 70% erano originate dal medico per propria decisione (43%) o per induzione da parte dello specialista (27%), ma il 30% derivavano dalla esplicita richiesta del paziente che in più della metà dei casi (16% del totale) è asintomatico e motiva la richiesta con la prevenzione (4.5%) o con la conferma di salute (11%). Il quadro sintetizza i mutamenti culturali della società del benessere, focalizzata non più al “silenzio degli organi”, come Bichat definiva la salute due secoli fa, né solo alla salute come “benessere fisico e psichico”, come OMS definiva la salute nel secolo scorso, ma alla cura del corpo, come oggetto preferenziale della propria attenzione. In più, l'accesso diretto dei cittadini alla diagnostica di laboratorio è diventato un'importante realtà commerciale sospinta dal diffondersi delle fonti di conoscenza sempre disponibili su Internet e dalla crescente volontà di contare anche nella gestione della salute (come negoziazione) e dalla mobilità nel rapporto medico-paziente. In particolare la diffusione delle tecnologie informatiche e della “rete” stanno modificando pesantemente il rapporto medico-paziente nei modi di approccio ai problemi e nei sistemi comunicativi relazionali, complicando il quadro generale dove sia tra i medici che tra i pazienti vi è chi ingloba freneticamente le novità e chi resiste al cambiamento²⁴.

E' difficile se non controproducente puntare sulla cooptazione dei cittadini alla gestione della propria salute e poi non assicurare loro adeguata possibilità di

scelta. Tuttavia, i medici di medicina notano come l'ipochondria collettiva e una diffusa paura di malattia e morte difficilmente si trasformano in comportamenti più sani della gente ma piuttosto in una richiesta eccessiva di test, al punto di sostenere che l'uso del laboratorio sembra il sostituto moderno dell'antica pratica di leggere il futuro nelle viscere degli animali²⁵. Vi è preoccupazione per l'assenza di adeguate informazioni su vantaggi e svantaggi dei test, sui limiti delle interpretazioni statistiche e del significato relativo di intervalli di riferimento e limiti decisionali, sulla equivocità di molti test nell'identificazione di malattie o nella definizione di prognosi, sulla fallibilità dei check-up. “Il compito dell'equipe di laboratorio verso i cittadini, dunque, è quello di fornire adeguate informazioni alla popolazione sull'uso e le modalità di accesso alla diagnostica, di formare attraverso interventi generali e specifici l'opinione pubblica rispetto al significato interpretativo e non assoluto dei risultati degli esami di laboratorio, di indirizzare correttamente ad un appropriato iter terapeutico chi si rivolga direttamente al laboratorio. Questo compito sociale enfatizza la nuova frontiera del laboratorio, fuori dalle mura”²⁶.

La selezione dei test

Il secondo passo della fase pre-preanalitica è la selezione dei test. Secondo il teorema di Bayes l'utilità del test diagnostico sta nel mutamento della probabilità post test rispetto a quella pre-test di una certa diagnosi, nel senso lato descritto da Sackett. Secondo una vulgata diffusa se un test non modifica il giudizio ed il comportamento del medico non deve essere richiesto. Il rapporto di verosimiglianza ne condensa quantitativamente l'utilità.

Nella selezione dei test la virtù cardinale è (o dovrebbe essere) l'appropriatezza. Tuttavia definire l'uso appropriato degli esami di laboratorio resta un obiettivo largamente eluso, per le difficoltà di individuarne l'essenza, circoscriverne il campo, dividerne i contorni e fondarlo su elementi provati.

Nel Glossario del Programma Nazionale Linee Guida dell'Istituto Superiore di Sanità²⁶ si distingue una appropriatezza clinica o specifica (diagnostica, terapeutica) ed una appropriatezza organizzativa o generica. La definizione puntualizza i possibili diversi piani anatomici del problema. Qualche anno fa Michael Wilson²⁷, insigne microbiologo americano, ha descritto efficacemente i diversi punti di vista dai quali il problema può essere indagato. Per il medico curante l'uso appropriato del laboratorio è definito da ciò che crede essere necessario per la cura di uno specifico paziente in uno specifico momento. Per i medici di medicina generale l'uso appropriato può essere definito dal *benchmarking*, da obblighi contrattuali verso il (o da incentivi pagati dal) “terzo pagante” e da standard professionali condivisi. Per il laboratorio l'uso appropriato può essere definito dalle caratteristiche operative del test, dal costo o dalla disponibilità di metodi alternati-

vi. Per il terzo pagante l'uso appropriato è probabilmente definibile come quello che limita la richiesta di esami per ridurre i costi della sanità. Per gli incaricati della sorveglianza della salute pubblica l'uso appropriato è quello richiesto per individuare nella popolazione le malattie di interesse pubblico. Per le istituzioni e gli organismi normativi l'uso appropriato è definito dal principio di necessità medica. È evidente che le definizioni sono largamente incompatibili e che questa è la principale ragione per la quale gli studi sull'uso appropriato dei test di laboratorio o i tentativi di protocollarlo hanno spesso prodotto dati inconsistenti o conflittuali.

Nel Glossario dell'Organizzazione Mondiale della Sanità²⁸ un intervento (ovviamente anche richiedere un esame di laboratorio è un intervento) viene considerato appropriato se il beneficio atteso in termini di salute è superiore ai possibili effetti negativi previsti, con un margine sufficiente a giustificarne la scelta. Questa definizione esprime in termini moderni l'antico precetto ippocratico *Primum non nocere* e risente degli studi sull'appropriatezza dell'uso di tecnologie ed interventi invasivi in campo cardiologico negli anni '80 in USA. Tuttavia il concetto di appropriatezza si è evoluto dalla definizione della RAND Corporation, stilata da Robert H. Brook²⁹ - "il beneficio atteso di salute supera le conseguenze negative attese di un margine sufficientemente ampio che la procedura valga la pena di essere praticata, eccetto il costo" - in una serie di definizioni degli anni '90 che cercano di puntualizzare l'appropriatezza come "assenza di errori medici e spreco di risorse". Nel 1996 il College of American Pathologists³⁰ definisce l'appropriatezza come "il grado in cui una procedura, trattamento, test o servizio è efficace, chiaramente indicato, non eccessivo, qualitativamente adeguato e fornito a ricoverati, ambulatoriali, a domicilio o in qualunque altra situazione si trovino, per rispondere al meglio ai bisogni del paziente". Gli studi degli anni '90 sulle dimensioni dell'inappropriatezza e sui mezzi efficaci per contenerla, in effetti, presuppongono il concetto condensato nella definizione "inappropriato è il test che può essere evitato senza detrimento per le cure del paziente"³¹. Il miglioramento dell'appropriatezza dovrebbe condurre al miglioramento dell'efficacia e dell'efficienza clinica dei servizi di laboratorio che sostiene il ruolo degli specialisti di laboratorio come consulenti e collega la diagnostica di laboratorio con gli *outcome* clinici³². Ma nella visione sopra descritta, l'appropriatezza è un prerequisito, focalizzato alla selezione dei test, dell'efficacia clinica ma non la garantisce di per sé. Addirittura, secondo Coast³³, una cura "inappropriata", secondo le formali definizioni anglosassoni che riecheggiano quella in questione, potrebbe essere potenzialmente più *cost-effective* di quella "appropriata", poiché gli strumenti di revisione dell'utilizzazione del test diagnostico o della cura sono di solito rivolti all'efficienza più che all'efficacia clinica e sono spesso strumenti teorici e burocratici che possono condurre a risultati ambigui. Secondo Wilson²⁷ gli

elementi che compongono l'appropriatezza, dal punto di vista del laboratorio, sono (o dovrebbero essere) la rilevanza clinica dell'esame, la sua *cost-effectiveness*, le caratteristiche operative del test e il suo *turnaround time* (TAT). Identificare la rilevanza clinica non è così semplice come potrebbe sembrare, perché i medici ordinano gli esami per molte e svariate ragioni e in situazioni cliniche molto diverse, dove la rilevanza clinica muta; perché gli esami di laboratorio sono interpretati alla luce di scenari clinici complessi e non come risultati isolati e o indipendenti; perché le diagnosi sono basate su dati obiettivi e soggettivi; perché varia la percezione della rilevanza clinica tra i medici posti dinanzi alle medesime informazioni. Nella realtà italiana dei MMG il 12% delle visite è corredato dalla richiesta di esami di laboratorio, con un ambito di variazione tra 8,4 e 17,6% e cioè superiore al 100% (O Brignoli, comunicazione personale, 2005). In ambito anglosassone, dove complessivamente il numero di test annui pro capite è decisamente inferiore a quello italiano (2,4 vs 17,7), ricerche³⁴ dei primi anni 2000 hanno mostrato la grande variabilità (700%) nella richiesta di esami diagnostici in 22 *general practice*, non motivata dalla diversità della popolazione servita per età, sesso e stato generale di salute (*Townsend score*) né dalla composizione del servizio sanitario inteso come numero di *general practitioner* nella singola *practice* e come presenza di minicliniche specialistiche. La conclusione degli autori è che la variabilità nella richiesta di esami di laboratorio è principalmente determinata dal diverso atteggiamento individuale nella pratica clinica. Inoltre, la maggior parte dei risultati di laboratorio modificano la loro rilevanza clinica in rapporto alla storia clinica, anamnesi familiare, segni sintomi ed esame obiettivo, risultati di altre indagini di laboratorio e non; cioè, la probabilità pre-test teorica, conosciuta sulla base delle prevalenze, ma calata nel singolo caso inteso come nosologia presente situazione clinica e conoscenza del medico, influenza potentemente l'interpretazione clinica sulla rilevanza dell'esame, che non può essere considerata al di fuori del contesto clinico. Stuart Smellie³¹, noto biochimico inglese, scrive che l'appropriatezza di un esame di laboratorio dipende dal contesto clinico del singolo paziente, dalla gravità della malattia, da necessità amministrative e da altri molteplici fattori e che quindi l'interpretazione dell'appropriatezza è qualcosa di soggettivo. Solo il dialogo tra laboratorio ed utilizzatori, in un clima di mutuo aiuto e non di critica o di pratica prescrittiva, può indirizzare alla miglior pratica nell'uso degli esami di laboratorio, superando il limitato concetto di appropriatezza come linee guida, protocolli e standard per disegnare un sistema idoneo alla sua applicazione individuale. Le definizioni di appropriatezza degli organismi internazionali per l'accreditamento e la qualità focalizzano la questione: "appropriatezza come il grado nel quale il servizio si adegua alle richieste del cliente, fornito in accordo con i principi della miglior pratica clinica" (ISQUA)³⁵, "appropriatezza come il grado

nel quale le cure e il servizio offerti sono rilevanti nel soddisfare i bisogni clinici di un individuo, alla luce delle conoscenze attuali" (JCAHO)³⁶.

Questa crescita culturale conduce alla definizione di Christopher P. Price³⁷, per la quale un test appropriato è "quello in cui il risultato fornisce una risposta alla questione che mette in grado di prendere una decisione ed intraprendere un'azione che porta ad un beneficio di salute per il paziente", concezione che modifica radicalmente la visione dell'appropriatezza da "assenza di errori medici e spreco di risorse" a "decisione di efficacia clinica e di risorse nella situazione pratica specifica". Price è cosciente che la definizione può apparire un punto di vista idealistico, ma essa serve a comprendere le situazioni in cui anche un test che fornisce benefici meno ovvi o misurabili vale la pena di essere eseguito, se migliora l'esito complessivo. Questo concetto tiene al proprio interno sia la variabilità individuale del medico e del paziente sia la questione del fondamento dell'appropriatezza, cioè il suo legame con l'*outcome*. Oggi per appropriatezza va intesa "la componente della qualità dell'assistenza che fa riferimento a qualità tecnico-scientifica, accettabilità e pertinenza (rispetto a persone, circostanza e luogo, stato corrente delle conoscenze) delle prestazioni sanitarie, raggiungibile attraverso il corretto impiego di strumenti analitici ed in cui confluiscono e si fondono efficienza, efficacia e convenienza". Nella Medicina di Laboratorio ciò significa che l'appropriatezza non è un predicato della sola fase pre-preanalitica (la selezione dei test) ma dell'intero processo diagnostico fino a farsi carico che la risposta del laboratorio sia correttamente utilizzata per lo specifico paziente³⁸.

Le difficoltà della definizione teorica di appropriatezza e il lento evolvere del concetto attraverso una visione tendente ad escludere test inappropriati ad una visione che include le problematiche idiografiche del singolo caso sono strettamente connessi ai modi e agli strumenti utilizzati nel tempo per determinare e pesare l'inappropriatezza nella complessa realtà clinica. In origine il sistema di definizione dell'appropriatezza poggia su metodi di consenso informale quali le opinioni di esperti e singoli studi dedicati a situazioni specifiche, ma negli anni '80 si avvale di metodi di consenso formale, costruito con procedure strutturate secondo il metodo del gruppo nominale e *round robin*, o della *consensus conference*, incentrati sullo scambio verbale regolato delle opinioni in assenza o presenza di pubblico, oppure secondo il metodo *Delphi*, basato sulla raccolta di questionari sequenziali ben strutturati, via posta o e-mail. Negli anni '80 la RAND Corporation ha introdotto un metodo *Delphi* modificato con una prima fase di raccolta e valutazione a punti della letteratura ed una seconda con un faccia a faccia tra componenti per la votazione finale. È evidente il tentativo di superare il livello delle impressioni aneddotiche e tuttavia, come scriveva David Witte³⁹ ancora nel 1995, "*results based on expert panels often only solidify old beliefs rather than lead to*

new information". D'altra parte la riproducibilità dell'approccio RAND è stato considerato eccellente ma non perfetto. Shekelle et al⁴⁰ hanno mostrato come 3 gruppi di valutatori hanno ritenuto appropriata la rivascolarizzazione coronarica, sulla base dei dati retrospettivi di 1294 pazienti cardiologici, in proporzioni considerevolmente differenti, cioè in 498, 464 e 402 pazienti rispettivamente.

Alla fine degli anni '90 nel fondamentale articolo "*Do we know what inappropriate Laboratory utilization is?*", Carl van Walraven e David Naylor⁴¹ esaminano sistematicamente la letteratura per raccogliere gli studi di inappropriatezza, riferiti ai soli test di laboratorio esclusi quelli di urgenza e quelli a fini di *screening*, provvisti di criteri specificati, utilizzati realmente negli *audit* e applicati ad un gruppo di pazienti diverso da quello usato per definirli. Dei 4039 articoli estratti da MEDLINE tra il 1966 e il 1997, i criteri impostati hanno selezionato 45 articoli, 11 che utilizzano criteri impliciti di appropriatezza e valutano l'insieme dell'utilizzo del laboratorio in istituzioni accademiche e 34 che hanno criteri espliciti e valutano singoli test biochimici o ematologici, microbiologici o standard su urine, di enzimi cardiaci, per la funzione tiroidea, di monitoraggio terapeutico. L'inappropriatezza totale rilevata è tra il 5 e il 95%, ma se si escludono i lavori con criteri opinabili (come il risultato negativo di un test o la mancata modifica della terapia e timing troppo stringenti nel monitoraggio di specifici parametri), gli studi con criteri impliciti mostrano un range di inappropriatezza tra 10 e 50% e quelli con criteri espliciti un range tra 5 e 50%, con una media intorno al 33%, dati che confermano, secondo un'analisi sistematica, quel 30-50% tramandato dalla letteratura precedente⁴².

Un esame attento dei risultati mostra come la media dell'inappropriatezza della richiesta salga da valori piuttosto bassi per i test di base ematologici e di chimica clinica (15%) a valori più elevati per esami specialistici (30-40% per esami tiroidei e cardiaci; 45-50% per esami microbiologici e di monitoraggio farmacologico). Questi dati sono in accordo con altre segnalazioni di letteratura che indicano nei test più rari o specialistici le aree di maggior inappropriatezza^{43,44}.

Oltre ai risultati quantitativi ottenuti con metodologia sistematica, gli autori suggeriscono un paradigma di classificazione dei criteri di appropriatezza, basati sul confronto diretto tra test secondo le loro caratteristiche operative (sensibilità, specificità, rapporto di probabilità) oppure, per un singolo test isolato, basati sulla fisiologia, la farmacologia e la probabilità clinica, quest'ultima valutata sia per i valori predittivi conosciuti del test nella specifica situazione (caratteristiche operative) sia per la prevedibilità del risultato (futilità). Molti sono gli spunti di riflessione: la differenza tra uso di criteri impliciti rivolti all'uso complessivo della diagnostica e uso di criteri espliciti rivolto prevalentemente a test singoli; il limitato uso di valutatori diversi e l'episodica valutazione formale (come k) della concordanza

interosservatore (un lavoro che usa criteri impliciti e uno che usa criteri espliciti); la larga assenza di *compliance* con le linee guida per stabilire la validità dei criteri e quindi la scarsità di evidenze di elevato livello. Gli Autori esplicitamente evidenziano alcune corpose criticità: definire inappropriato un test che non produce un cambiamento terapeutico è irragionevole, così come è contestabile utilizzare il criterio della “normalità” del risultato; la totale assenza del rapporto tra la *compliance* ai criteri e gli *outcome* dei pazienti, sconosciuti e solo in pochissimi casi sostituiti da *surrogate outcome* quali i costi o la prevalenza dei risultati positivi è un limite molto forte; infine, praticamente tutti gli studi sono stati effettuati in centri universitari o d’insegnamento, su test a pazienti ospedalizzati, analizzando la pratica della richiesta dei test di un piccolo numero di medici in specializzazione!

Le difficoltà della valutazione dell’inappropriatezza erano state molto ben descritte da Edward Wong e Sunita Saxena⁴⁵ ancora nel 1992 in un editoriale dell’*American Journal of Clinical Pathology* che commentava un articolo di Bunton e Gade sull’utilizzo dei profili d’ingresso in un ospedale americano valutati secondo le linee guida del CAP. Notavano gli Autori alcune criticità: l’erronea assunzione che un test negativo o normale sia di per sé inutile, dato che può avere l’importante valore di *rule out*; l’erronea convinzione che sia non necessario il test che non conduce ad azioni cliniche evidenti: talora si tratta di un elemento della costellazione clinica rispetto al quale non è necessario un intervento specifico, come nel caso di una ipoalbuminemia moderata in soggetto noto per cirrosi alcolica, talaltra gli interventi sono considerati talmente ovvi e routinari da non condurre ad una chiara annotazione sulla cartella clinica, come nel caso di alterazioni degli elettroliti; la necessaria consapevolezza che difficilmente i test più usati di chimica clinica sono utilizzati come singoli e non piuttosto in batteria e che interventi di eliminazione di *little ticket test* sono economicamente poco produttivi; la necessità di abbandonare standard immobili di appropriatezza per determinarla piuttosto sulla base degli *outcome* clinici.

Carl van Walraven e David Naylor⁴¹, tuttavia, sono consci che l’appropriata richiesta di diagnostica non si trasferisce immediatamente in un’appropriata utilizzazione dei risultati della stessa, ma sono convinti che comunque lo sforzo per ottenerla sia motivato da una migliore qualità delle cure come inappropriati interventi conseguenti a risultati falsi positivi e per la provata associazione tra inappropriatezza nell’utilizzo del laboratorio ed altre inefficienze dell’assistenza sanitaria.

Durante gli anni ’90, i medici di laboratorio nord-americani hanno messo a punto metodologie di studio dell’inappropriatezza di cui conoscono potenzialità e limiti e le usano per misurare il sovra-utilizzo della diagnostica di laboratorio. Si rende evidente la centralità del concetto di *outcome* e si appronta una metodologia di studio, basata su un coerente *background* cultu-

rale e teoretico che collega indissolubilmente appropriatezza ed *evidence-based medicine*.

Ma, una volta in possesso di strumenti di misura, emergono prepotentemente i sotto-utilizzi della diagnostica, come dimostrato da alcuni, pochi, studi specifici, dalle indagini delle grandi organizzazioni sanitarie americane pubbliche e private, dall’indagine nazionale USA sullo stato della salute nel 2003. In ogni caso si studiano le strategie di intervento, considerate efficaci solo se combinazioni di interventi educativi, amministrativi e di feed-back, e se ne pesa l’effetto in termini di volume di attività, di costi e di efficacia clinica.

L’intrinseca connessione tra appropriatezza ed “evidence-based medicine” trova fondamento nell’efficacia ed efficienza clinica dalla cui misura vengono definite⁴⁶. La “evidence-based laboratory medicine” (EBLM) è definita come “informazioni raccolte in modo sistematico e valutate in modo critico, provenienti di preferenza da studi di ricerca primaria ben progettati, per rispondere a quesiti specifici su diagnosi, diagnosi differenziale, screening, monitoraggio e prognosi, che forniscono un contesto esplicito per prendere decisioni mediche informate”. “EBLM integra nel processo decisionale clinico la migliore evidenza della ricerca per l’uso degli esami di laboratorio con l’esperienza clinica del medico e le necessità, le attese e le preoccupazioni del paziente”. Ma quale tipo di evidenza abbiamo in Medicina di Laboratorio? Andrea Horvath⁴⁷ sintetizza i molti motivi della limitata disponibilità: il problema dello standard di riferimento, progetti di studi primari inadeguati che determinano *bias* consistenti, difficoltà negli studi di *outcome*, i problemi e i *bias* delle revisioni sistematiche in Medicina di Laboratorio.

Le evidenze in Medicina di Laboratorio sono meno forti che in campo terapeutico, per la carenza di studi primari e la debolezza delle ricerche sistematiche e delle meta-analisi. Gli studi di *outcome* correlati alla diagnostica sono difficili innanzitutto per la stessa definizione di *outcome*, “risultato sanitario od economico di un intervento”, che viene percepito non dipendente dal test diagnostico, per la complessa identificazione della necessità del test e per la necessità di utilizzare *outcome* surrogati. Le difficoltà degli studi di *outcome* sono state ben elencate da David Bruns⁴⁸: costo dello studio *vs* potenziali profitti, dimensioni del campione, eticità della privazione di test notoriamente utili, approfondimenti per definire incertezze, lontananza temporale e concettuale dell’*outcome* dal test, risposta medica al risultato del test non conseguente, impossibilità di esperimenti in doppio cieco, gruppi di arruolamento noti. Nel documento del gruppo Cochrane sulle revisioni sistematiche dei test di screening e diagnostici⁴⁶ si sottolinea che nella valutazione di un test diagnostico occorre maggiore attenzione all’esito finale nello specifico paziente, ma si riconosce che, pur essendo teoricamente possibile valutare gli effetti di un test sugli esiti mediante trial randomizzati, nella pratica non è possibile ap-

plicare questa metodologia ad ogni nuovo test e che sembra più fattibile valutare l'accuratezza di test indirizzati alla diagnosi di condizioni per le quali esistono già evidenze dell'efficacia delle cure. Ma anche gli studi di accuratezza diagnostica sono inficiati da numerosi handicap come la mancanza di *gold standard* o la possibile presenza di *bias* importanti. Nonostante l'introduzione di regole condivise per il disegno degli studi di accuratezza diagnostica come lo STARD (*Standard for Reporting of Diagnostic Accuracy*)⁴⁹ o il QUADAS (*Quality Assessment of Diagnostic Accuracy Studies*)⁵⁰, recentemente sono state pubblicate segnalazioni assai sconcertanti da questo punto di vista. Nel febbraio del 2006, Anne Rutjes et al⁵¹ hanno focalizzato la evidenza di *bias* e variabilità negli studi di accuratezza diagnostica, rafforzando le conclusioni empiriche di Lijmer et al⁵² che nel 1999 avevano suggerito che imperfezioni nel disegno degli studi di accuratezza diagnostica potevano condurre a sovrastimare il potere diagnostico soprattutto se includevano pazienti non rappresentativi o utilizzavano standard di riferimento differenti. Gli Autori⁵¹, attraverso l'analisi di 31 revisioni sistematiche e 487 studi primari, identificano 15 potenziali sorgenti di *bias*, tra cui i principali sono il *bias* di spettro, il *bias* di verifica e la mancanza di doppio cieco (RDOR 1.5-4.9; 1.4-1.6; 1.3-1.6 rispettivamente). Un solo studio è stato giudicato privo di difetti, tuttavia il problema fondamentale è quello dello standard di riferimento che spesso è più *silver* che *gold standard*. Ancor più drasticamente Ioannidis⁵³ sostiene che la maggior parte delle ricerche pubblicate sono false, in rapporto ai *bias* e al numero di *team* coinvolti. I fattori condizionanti sarebbero: la dimensione dello studio, la grandezza dell'effetto misurato, il numero e la preselezione delle relazioni testate, la flessibilità del disegno dello studio e degli strumenti dello stesso, gli interessi e pregiudizi esistenti, la attualità del tema trattato. Gli *Evidence-based Practice Centers* (EPCs) della *Agency for Healthcare Research and Quality* utilizzano per il *Technology Assessment Program* un metodo gerarchico su 6 livelli. La cornice concettuale per la valutazione dell'efficacia dei test diagnostici è sostanzialmente quella proposta⁵⁴, ancora 15 anni fa, da Dennis Fryback, un ricercatore con esperienza nell'ingegneria industriale e nella scienza delle decisioni, e Jack Thornbury, radiologo all'Università del Wisconsin: efficacia tecnica, accuratezza diagnostica, efficacia del ragionamento diagnostico, efficacia gestionale, *outcome* del paziente, efficacia sociale. Un recente lavoro⁵⁵ presenta alcuni report di EPCs per evidenziare le sfide nelle revisioni sistematiche delle tecnologie diagnostiche: identificare gli studi rilevanti pubblicati e non; valutare la qualità metodologica dello studio; valutarne l'applicabilità dello studio di popolazione; sintetizzare le misure di accuratezza del test; definire l'impatto sociale della tecnologia diagnostica.

Non sempre le evidenze di EBLM sono poi trasferite in adeguati metodi di supporto alla diagnosi come linee guida o come percorsi critici. La presenza della

medicina di laboratorio nel disegno ed esecuzione di studi di accuratezza ed efficienza, nella interpretazione della letteratura in tema di strumenti diagnostici ed, infine, nella scelta e diffusione locale di linee guida e protocolli è essenziale per la creazione di evidenze e per la loro applicazione. La collaborazione multidisciplinare nell'implementazione di nuovi test è, d'altra parte, sempre più esplicitamente raccomandata dagli organismi professionali che definiscono linee guida diagnostiche e terapeutiche. Tuttavia, quando esistono, la scelta di quella più adatta localmente e la sua implementazione sono attività estremamente impegnative. Diverse sono le ragioni della difficoltà alla disseminazione delle linee guida: la carenza di metodi adatti al trasferimento delle evidenze della ricerca nella conoscenza in pratica, il timore di una interpretazione della EBM come un manuale di ricette o un mezzo di controllo amministrativo, la sovrastante influenza del contesto locale sui tentativi di modificare la pratica clinica ed infine l'evidenza del ruolo degli *script*, euristiche e *rules of thumb* (regole del pollice) nel *clinical decision making*. Nel 2004 Gabbay e le May⁵⁶ hanno pubblicato uno studio etnografico usando metodi standardizzati su 2 *general practices* del sud e del nord dell'Inghilterra per analizzare in profondità come i gli MMG e gli infermieri professionali derivino le loro decisioni assistenziali individuali e collettive. I MMG utilizzano raramente l'evidenza proveniente dalla ricerca o da altre fonti dirette, come le linee guida formalizzate, piuttosto si affidano a *mindlines*, cioè linee guida tacite, interiorizzate, rinforzate collettivamente. Esse derivano da brevi letture ma soprattutto dall'esperienza personale e dei colleghi, dall'interazione con i colleghi e *opinion leader*, pazienti, informatori farmaceutici ed altre fonti di conoscenza tacita. Mediate dalle richieste e dai vincoli organizzativi, le *mindlines* sono iterativamente negoziate con una varietà di attori chiave, spesso attraverso un insieme di interazioni informali in *communities in practice* fluide ed esitano in una *knowledge in practice* socialmente determinata. Secondo gli Autori, le *mindlines* sono qualcosa di più delle *rules of thumb* o delle euristiche ed anche più flessibili degli *script* ed implicano un diverso modo di approccio ai MMG da parte della EBM. Commenta Paul Glasziou⁵⁷ del *Center for Evidence-Based Medicine* di Oxford: è venuto il momento di trasformare la *evidence based medicine* in *evidence informed practice*, ma mette in guardia dalla nascita e moltiplicazione di memi mutanti, non correttamente informati, e suggerisce di partire dall'evidenza aggiungendovi un poco di saggezza nata dall'esperienza. Le segnalazioni sul basso effetto delle linee guida e dei mezzi educativi da soli vanno di pari passo con quelle sull'efficacia di strategie multifattoriali e di quelle basate sulla verifica dell'applicazione e dei suoi effetti. I tradizionali metodi dei lavori scientifici e della educazione medica continua hanno mostrato di essere deludenti. Il tema del trasferimento dalla conoscenza scientifica alla pratica clinica del potenziale diagnostico della disciplina è oggi il tema più avvincente e difficile della Medicina di

Laboratorio.

Infine, la metodologia della verifica del risultato finale (*outcome*) in rapporto all'appropriatezza del test diagnostico costituisce il metro di misura di più alto valore, quello della valutazione economica e sociale, in cui al sovrautilizzo si associa pesantemente il sottoutilizzo della diagnostica. Infatti, se qualche autore⁵⁸ sostiene che il crescente uso della diagnostica di laboratorio non corrisponde ad un misurabile miglioramento sociale, i dati sulle conseguenze sociali dell'inappropriatezza per difetto (*underuse*) della diagnostica negli Stati Uniti d'America sono drammatici³⁸. Ad esemplificazione si riporta la valutazione dei processi medici e dell'*outcome* dei pazienti per la cura del diabete effettuata da 109 *Veteran Affaire Medical Center* (VAMCs) nel 2001 e pubblicata nel 2004⁵⁹. I processi clinici sono apparsi nei VAMCs decisamente migliori rispetto ai dati di letteratura: il 94% dei pazienti fa una HbA1c annuale *vs* il 69-89% di altri campionamenti; l'87% ha una visita annuale ai piedi *vs* il 27-65%; l'81% fa 2 controlli all'anno dei lipidi *vs* il 50-68%; il 69% ha una visita annuale della retina *vs* il 26-78%. La misura dell'*outcome* è stata effettuata tramite esiti surrogati (HbA1c < 9,4%; LDL d'' 130 mg/dL; pressione arteriosa < 140/90 mm Hg; non fumo) ed è comparabile o migliore di altre valutazioni: 82,2% in controllo glicemico; 78,9% non fumatori; 70,5% in controllo lipidico; 59,3% normotesi. Questi dati e quelli sui positivi effetti economici (come costo complessivo dei ricoveri) e sociali (come numero di nuove diagnosi e complicanze evitate a 6 mesi), indotti per esempio dalla diffusione dei nuovi marcatori cardiaci e dall'introduzione delle linee guida relative, fanno ben capire quanto un aspetto della sanità così poco costoso come il laboratorio (dall'1 al 4% della spesa annuale complessiva in Europa) incida nella gestione della salute pubblica². La valutazione dell'impatto sociale ed economico della diagnostica può apparire un obiettivo difficile ed apparentemente irraggiungibile ma è ineludibile. Ad esso dovremo dedicarci più alacremente, in collaborazione con clinici, epidemiologi ed economisti.

La richiesta dell'esame di laboratorio

Ancora nel 1983 George Lundberg in un Editoriale su JAMA⁵⁹ indicava nella preparazione di opportuni formati di richiesta la via per guidare ed insegnare al clinico come richiedere i test più efficaci e più appropriati. Il ridisegno della richiesta è stato da allora uno dei cavalli di battaglia nella lotta per l'appropriatezza.

Nel 1998 Daniel Solomon e collaboratori⁶⁰ pubblicano una revisione sistematica della letteratura intorno alla modificazione delle abitudini dei clinici nel richiedere i test diagnostici, secondo tre fattori di cambiamento: predisponenti (attitudini, conoscenze, percezioni), permissivi (abilità, risorse, barriere strutturali), rinforzanti (*feedback*). Nel caso dei test diagnostici i fattori predisponenti sono gli interventi educativi, quelli permissivi i cambiamenti della struttura amministrativa

della richiesta di test e quelli rinforzanti gli *audit*. Secondo gli studi identificati (n. 102) e selezionati dai criteri prestabiliti (n. 49; 48%), il 76% degli interventi complessivi ha avuto successo: nel 62% dei casi con un solo tipo di intervento (60% esclusivamente educazionali; 36% esclusivamente di *audit*; 60% esclusivamente amministrativi); nel 86% combinando più tipi di intervento (88% con interventi educazionali associati ad *audit*; 75% con interventi amministrativi associati ad *audit*; 100% con interventi educazionali associati ad amministrativi - si tratta però di un solo lavoro!); nel 100% dei casi quando sono stati associati interventi educazionali, amministrativi e *audit*. Lo studio dimostra che nel controllo della proliferazione della richiesta una unica strategia non funziona ma, poiché i fattori comportamentali sono molteplici, anche gli interventi devono essere multifattoriali e sorretti da una precisa e conosciuta struttura cognitiva e comportamentale. L'intervento educativo tradizionale ha effetti deboli ma è condizione necessaria per l'efficacia degli interventi di rinforzo quali feedback informativi ed economici ed *audit*, i quali, da soli, hanno incerte probabilità di successo. Gli interventi amministrativi ed ambientali sono incredibilmente efficienti a basso costo, ma vanno scelti con cautela perché non conosciamo bene i loro effetti sull'efficacia delle cure.

Le conclusioni di Solomon et al⁶⁰ al sono ribadite dall'articolo immediatamente successivo del numero di JAMA del 16 dicembre 1998, in cui Carl van Walraven e colleghi⁶¹ riportano l'effetto dell'applicazione di diverse strategie di contenimento della diagnostica, applicate in Ontario tra il 1991 e 1997. È il primo lavoro veramente popolazionistico in un quadro sanitario comune e stabile, standardizzato per età e sesso e basato su misure cadenzate in un tempo sufficientemente ampio e in grado di valutare accuratamente il rapporto tra interventi e risultati anche a distanza. I test diagnostici affrontati e le misure adottate, nel tempo, sono: VES con interventi amministrativi (eliminazione della richiesta) ed educazionali (linee guida) ripetuti; esame delle urine con interventi amministrativi (separazione richiesta chimico-fisico e microscopico) ed educazionali ripetuti; creatinina e urea con interventi educazionali e amministrativi (eliminazione richiesta urea); metabolismo del ferro con interventi amministrativi (richiesta ferritina esclude esecuzione ferro e TIBC) ed educazionali ripetuti; funzione tiroidea con interventi educazionali ripetuti, amministrativi (su uptake e tiroxina totale), di nuovo educazionali ed amministrativi (eliminazione richiesta TSH). Tutti i test diminuiscono in volume (VES -58%; urea -57%; urine microscopico -14%; sideremia -80%; tiroxina totale -96%; TSH -12%), eccetto la richiesta di urine chimico-fisico (+1700%) e di ferritina (+34%). In tre casi (urea, esame urine microscopico, sideremia) la sola disseminazione di linee guida ha avuto effetti significativi. Le modifiche del formato di richiesta sono più efficaci perché raggiungono tutti i medici, ma la combinazio-

ne di linee guida, modifiche del formato di richiesta dei test e cambiamenti di politica economica sono efficaci in tutti i casi. Lo studio dimostra nella pratica gli effetti positivi in termini di riduzione dell'approccio multidimensionale degli interventi, ma non consente peraltro di capire se la diminuzione si coniuga ad una diminuzione dell'inappropriatezza, come sovra- e sotto-utilizzo, o si risolve semplicemente in un razionamento.

Commenta Bandolier⁶² che, presi insieme i due lavori su JAMA mostrano convincentemente che il comportamento dei medici nel richiedere la diagnostica di laboratorio può essere cambiato in termini quantitativi importanti, purché l'approccio sia multifattoriale e continuo, che il risparmio economico è peraltro assai contenuto e che la diminuzione del carico di lavoro non è solo un dato finanziario o manageriale ma l'occasione e la risorsa per pensare a ciò che rende i laboratori più efficaci e più importanti. Nell'Editoriale⁶³ di JAMA che accompagna i due lavori, George Lundberg indica la via dell'appropriatezza nella richiesta di diagnostica di laboratorio.

Nel 2003 Gopal Rao et al⁶⁴ si chiedono se è finalmente maturo il tempo per una gestione della domanda e si danno una risposta fiduciosa, alla luce degli sviluppi del NHS britannico: vi è una maggiore disponibilità di collegamenti bidirezionale tra laboratorio e clinici che consente di richiedere esami e ricevere risultati tramite computer e le possibilità di regole e ricordi automatici, nonché di cartelle cliniche informatizzate; l'enfasi sulla EBLM favorisce la revisione delle pratiche diagnostiche; il lavoro multidisciplinare all'interfaccia consente di elucidare le caratteristiche operative dei test e dei metodi e il loro ruolo negli algoritmi diagnostici o decisionali.

In alcuni casi⁶⁵ si tratta molto semplicemente di distinguere i test richiedibili dal contesto ospedaliero e da quello territoriale - agire cioè sul menù disponibile - per ottenere risultati significativi sul contenimento di esami comuni come il calcio sierico, la lattato deidrogenasi, la proteina C reattiva e il fattore reumatoide. L'intervento limitativo della richiesta funziona anche nella definizione dei profili d'ingresso e di emergenza, meno per i profili pre-operatori, nonostante le linee guida ormai condivise come quelle NICE. Sistemi esperti⁶⁶ sulla base di 54 linee guida accettate dal Collegio Olandese dei Medici di Base, hanno fornito ottimi risultati sul volume di test ordinati da 22 medici di base (-20% complessivo), seppur con importanti differenze tra test: emoglobina -18%, VES -28%, ALT -38%, fT4 -43%, potassio -50%. Il sistema LAS (*Laboratory Advisory System*) è in grado di offrire la scelta tra test singolo e percorso diagnostico per problema. In una prova⁶⁷ relativa a casistica tiroidea, epatitica ed autoimmune in confronto con la pratica convenzionale, ha mostrato una diminuzione del numero totale dei test (media 17,8 vs 32,7), un minor numero di campioni per completare l'iter diagnostico (5,8 vs 7,5), minor tem-

po per giungere alla diagnosi (1 giorno vs 3,2), minori costi (\$ 194 vs 232), maggiore aderenza alle linee guida e una ricerca clinica più omogenea e diagnostica. Anche nella casistica in aree diagnostiche prive di linee guida riconosciute, con LAS, i test richiesti sono la metà, i prelievi il 27% in meno e decisamente minori le richieste di visita specialistica. Sistemi esperti sono capaci di avvertire della duplicazione della richiesta inferiore a tempi stabiliti. Nelle ricerche⁶⁸ in questo campo si è evidenziato che i dosaggi ripetuti inappropriatamente, definiti come richieste anteriori a criteri prestabiliti, rappresentano il 28% dei test di un grande ospedale americano, con un ambito di variazione da 2 a 62% in base alla tipologia dei test. L'analisi delle cartelle cliniche ha consentito di definire come "ridondanti" fino al 40% dei test richiesti, con un ambito di variazione tra il 26 e il 100%. In questo caso, diversamente dalla situazione di diagnosi in paziente non noto, la negatività del test replicato può essere considerata una seria misura di appropriatezza, quando effettuato in tempi troppo brevi per una modifica clinicamente significativa. In realtà asiatiche⁴⁴ tali sistemi hanno consentito di contenere del 17% la richiesta di esami immunologici. Tuttavia, molto più stimolanti per la Medicina di Laboratorio sono, da un lato, la condivisione con i clinici di algoritmi diagnostici che vengono automatizzati in cascate analitiche secondo le regole *if...then (reflex test)* e sono stati applicati con successo sia alla biochimica clinica (diagnostica tiroidea) che all'ematologia (diagnostica delle anemie), dall'altro la definizione di richieste per problemi invece che per test. Esempi molto interessanti sono quelli relativi alla diagnostica lipidica, a quella autoimmune, a quella coagulativa e a quella delle gammopatie monoclonali.

La potenza di una richiesta orientata per problemi è dimostrata da Smellie⁶⁹: con gli obiettivi di ridurre le disuguaglianze nella diagnostica dell'assetto lipidico in pazienti di base e di impostare un *audit* per i gruppi di pazienti in terapia, al posto del tradizionale formato di richiesta è stata offerta ai medici di base, secondo le linee guida britanniche, la scelta tra 4 problemi diagnostici (prevenzione primaria, non in terapia; prevenzione primaria in terapia; prevenzione secondaria non in terapia; prevenzione secondaria in terapia). I risultati sulla richiesta di test sono stati una diminuzione del rapporto HDLC/TCOL dal 24 al 14% e una marcata riduzione della variabilità tra i prescrittori. I risultati dell'*audit* sono stati la riduzione dell'HDLC dal 44 al 35% senza effetti su TCOL e TG e il raggiungimento del target per la prevenzione secondaria mediamente nel 42% dei pazienti e per la prevenzione primaria nel 66%. L'effetto di costo-utilità in quantità e qualità degli *outcome* per i pazienti sottoposti ad adeguata prevenzione e terapia non è calcolato ma è ovvio e rappresenta un esempio evidente dell'effetto anche economico di un diverso approccio alla diagnostica di laboratorio.

Politiche di *gating* sono efficaci per specifiche richie-

ste. L'esecuzione di ANCA⁷⁰ esclusivamente in pazienti con sospetto diagnostico prestabilito con i clinici (insufficienza renale, poliartrite nodosa, granulomatosi di Wegener, proteinuria/ematuria in nefrologia, sindrome di Churg-Strauss, sindrome nefrotica, anomalie polmonari radiografiche o bioetiche, emoftisi, perforazione del setto nasale, porpora di Henoch-Schonlein) conduce ad una appropriatezza di richiesta del 72.5%, rispetto alla politica concordata, e ad una elevata predittività del test: 42/57 pazienti ANCA positivi hanno una vasculite sistemica.

Il gruppo di Michael Laposata ha più volte illustrato i modi di richiesta per problemi e la conseguente interpretazione clinica con referto interpretativo in ambito coagulativo. Per esempio, un PTT prolungato origina la ripetizione del test dopo degradazione dell'eparina; se non si normalizza uno studio di prova con plasma normale (rapporto 1:1); se si normalizza si misurano i fattori singoli (VIII, IX, XI e XII); se non si normalizza si valutano gli inibitori a partire dall'inibitore del fattore VIII. Il referto interpretativo fornisce una descrizione della fisiopatologia ed eventualmente della fisiologia del caso in esame, nel 20-30% dei casi con commenti specifici per il paziente e non predefiniti (*canned comments*). Nell'*audit* presso i propri clinici, Kratz e Laposata⁷¹ hanno ottenuto le seguenti risposte: il 98% dei clinici valuta le interpretazioni utili o informative; il 59% indica che ha accorciato i tempi di diagnosi; il 71% valuta che abbiano diminuito il numero di test di laboratorio richiesti; il 71% è convinto che il commento interpretativo originato dall'algoritmo abbia evitato errori diagnostici.

Esistono algoritmi definiti per la ricerca delle monoclonalità proteiche, spesso trasferiti in algoritmi interni al laboratorio. In area anglosassone⁷² il punto di partenza suggerito dalla *best practice* è la richiesta di dosaggio delle immunoglobuline e di elettroforesi sierica ed urinaria per la definizione di un'elevazione generalizzata delle proteine (malattia cronica), di una generalizzata diminuzione (immunodeficienza congenita?), di una banda monoclonale da tipizzare e misurare. Su questa base si monitora ogni 4 mesi e poi ogni anno oppure, se > a 15 g/L se IgG o IgA o con sintomi se IgM all'esordio o se vi è un incremento del 20% del picco, si invia all'ematologo con il sospetto di monoclonalità neoplastica.

È importante accompagnare la diffusione dei formati di richiesta con adeguata condivisione degli aspetti educazionali e di ricerca e adeguati feed-back e *audit*. Si tratta di una incomparabile occasione di colloquio con i clinici e di condivisioni degli obiettivi generali di salute. D'altra parte se non si conosce il motivo della richiesta, la logica diagnostica, le problematiche della decisione clinica susseguente alle indagini non si è in grado di rispondere adeguatamente al quesito clinico sottostante. Joseph Alpert ha intitolato un recente Editoriale⁷³ "*The answer you get depends on the question you ask*", esemplificando con una storia clinica in cui la giusta

domanda del medico al paziente estrae la corretta ipotesi diagnostica e con una ricerca sull'efficacia di un farmaco, valutata contro placebo invece che contro una molecola già nota per la sua attività e quindi inevitabilmente positiva. Il commento è che sia nella clinica che nella ricerca si deve mantenere il senso critico e chiedersi qual è la domanda. Tuttavia anche la risposta può non essere adeguatamente interpretata, ma comprendere il quesito è l'unico modo, necessario anche se non sufficiente, perché la risposta sia interpretabile ed efficace.

Si è cercato di mostrare come questo passaggio, apparentemente ovvio, sia delicato e complesso nella medicina di oggi e come debba essere richiesta al clinico il massimo della chiarezza nel porre la domanda. È stato visto in ambito di medicina generale USA⁷⁴ che solo il 38% delle richieste di consulenza conteneva le componenti che garantivano un'efficace risposta al quesito clinico, cioè l'identificazione dell'intervento proposto e dell'esito desiderato. Possiamo solo stimare quanta parte dei quesiti clinici in Medicina di Laboratorio siano sufficientemente chiari e posti in modo adeguato. Ma non possiamo attendere il mutamento dei clinici senza un opportuno intervento. È compito dei professionisti della Medicina di Laboratorio offrire occasioni di studio e chiarimento dei passaggi essenziali nell'emergere del problema diagnostico e strumenti adeguati per la sua esplicazione appropriata, nella convinzione dell'intimo legame tra questa fase e quella del referto e della sua applicazione alla cura del paziente, momenti che rappresentano l'essenza e la nuova frontiera della professione.

Bibliografia

1. Büttner J. Good laboratory practice: the medical aspects. *Eur J Clin Chem Clin Biochem* 1997; 35:251-6.
2. Cappelletti P. La risposta della Medicina di Laboratorio al quesito clinico. *RIMeL - IJLaM* 2005; 3 Suppl:15-26.
3. Goldschmidt HMJ. The NEXUS vision: an alternative to the reference value concept. *Clin Chem Lab Med* 2004; 42:868-73.
4. Lundberg GD. Acting on significant laboratory results. *JAMA* 1981; 245:1763.
5. Lundberg GD. Critical (panic) value notification: an established laboratory practice policy (parameter). *JAMA* 1990; 263:709.
6. Burlina AB. La tollerabilità dei farmaci, il laboratorio ed il "brain-to-brain loop". *Progr Med Lab* 1990; 3:137-8.
7. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. Boston: Little, Brown; 1991.
8. Burlina AB. *Medicina di laboratorio. Fondamenti di diagnostica*. Torino: CG Edizioni MedicoScientifiche; 1992.
9. Price CP. Application of the principles of evidence-based medicine to laboratory medicine. *Clin Chim Acta* 2003; 333:147-54.
10. Goldschmidt HMJ. Postanalytical factors and their influence on analytical quality specifications. *Scand J Clin Lab Invest* 1999; 59:551-4.

11. Wertman BG, Sostrin SV, Pavlova Z, Lundberg GD. Why do physicians order laboratory tests? A study of laboratory test request and use patterns. *JAMA* 1980; 243: 2080-2.
12. Lundberg GD. Perseverance of laboratory test ordering: a syndrome affecting clinicians. *JAMA* 1983; 249:639.
13. McNeely MD. Computerized interpretation of laboratory tests: an overview of systems, basic principles, and logic techniques. *Clin Biochem* 1983; 16:141-6.
14. Benincasa F, Castaldi G. La decisione nel corso dell'azione. In Caimi V, Tombesi M ed. *Medicina Generale*. Torino: UTET; 2003. p.56-62.
15. Bernabè S, Benincasa F, Danti G. Il processo diagnostico. In Caimi V, Tombesi M ed. *Medicina Generale*. Torino: UTET; 2003. p.45-55.
16. Stefan E, Fassina R. Strategie diagnostiche nel setting della Medicina Generale. *Rivista di politica professionale della Medicina Generale-SIMG* 2004; 3:41-4.
17. Gill CJ, Sabin L, Schmid CH. Why clinicians are natural bayesians. *BMJ* 2005; 330:1080-3.
18. Hunter K. "Don't think zebras": uncertainty, interpretation, and the place of paradox in clinical education. *Theor Med* 1996; 17:225-41.
19. Charlin B, Tardiff J. Scripts and medical diagnostic knowledge. *Acad Med* 2000; 75:182-90.
20. Tombesi M. Decidere in condizioni di incertezza. In Caimi V, Tombesi M ed. *Medicina Generale*. Torino: UTET; 2003. p.78-9.
21. Dorizzi RM. Approccio agli esami di laboratorio in medicina generale. In: Pigoli G, ed. *Gli esami di laboratorio: prescrizione e interpretazione*. Torino: UTET; 2002. p 3-20.
22. Sadeh-Zadeh K. Fundamentals of clinical methodology. 4. Diagnosis, *Artif Intell Med* 2000; 20:227-41.
23. Bauchner H, Simpson L, Chessare J. Changing physician behaviour. *Arch Dis Child* 2001; 84:459-62.
24. Kassirer JP. Patients, Physicians, and the Internet. *Health Aff* 2000; 6:115-23.
25. Pagni A, Plebani M. The laboratory and the general practitioner. *Clin Chim Acta* 1999; 280:13-24.
26. http://www.pnlg.it/glossario/gloss_a.htm (data di consultazione: 28.3.2006).
27. Wilson ML. Appropriate use of clinical microbiology tests. *Clin Lab Med* 2002; 22:491-503.
28. <http://www.euro.who.int/observatory/Glossary/TopPage?phrase=A> (data di consultazione: 28.3.2006).
29. Brook RH. Appropriateness: the next frontier. *BMJ* 1994; 308:218-9.
30. College of American Pathologists. Standards for laboratory accreditation, 1996 edition. College of American Pathologists, Northfield (IL).
31. Smellie WSA. Appropriateness of test use in pathology: a new era or reinventing the wheel? *Ann Clin Biochem* 2003; 40:585-92.
32. Plebani M. Appropriateness in programs for continuous quality improvement in clinical laboratories. *Clin Chim Acta* 2003; 333:131-9.
33. Coast J. Appropriateness versus efficiency: the economics of utilization review. *J Health Policy* 1996; 36:69-83.
34. Smellie WSA, Galloway MJ, Chinn D, Gelding P. Is clinical practice variability the major reason for the differences in pathology requesting patterns in general practice? *J Clin Pathol* 2002; 55:312-4.
35. <http://www.isqua.org.au> (data di consultazione: 27.3.2006).
36. <http://www.jcaho.org> (data di consultazione: 27.3.2006).
37. Price CP. Application of the principles of evidence-based medicine to laboratory medicine. *Clin Chim Acta* 2003; 333:147-54.
38. Cappelletti P. La modernizzazione dei laboratori orientata all'appropriatezza diagnostica e all'efficacia dei trattamenti. *Riv Med Lab - JLM* 2004; 2:147-63.
39. Witte DL. Measuring outcomes: why now? *Clin Chem* 1995; 41:775-80.
40. Shekelle PG, Kahan JP, Bernstein SJ, Leape LL, Kamberg CJ, Park RE. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med* 1998; 338:1888-95.
41. van Walraven C, Naylor CD. Do we know what inappropriate Laboratory utilization is? *JAMA* 1998; 280:550-8.
42. Hindmarsh JT, Lyon AW. Strategies to promote rational clinical chemistry test utilization. *Clin Biochem* 1996; 29:291-9.
43. Sinclair D, Saas M, Stevens JM. The effect of a symptom related "gating policy" on ANCA requests in routine clinical practice. *J Clin Pathol* 2004; 57:131-4.
44. Kwolek J, Jones B. Unnecessary repeat requesting of tests: an audit in a government hospital immunology laboratory. *J Clin Pathol* 2005; 58:457-62.
45. Wong ET, Saxena S. Medical Appropriateness of Laboratory Tests. *Am J Clin Pathol* 1992; 97:748-50.
46. Giavarina D. L'appropriatezza e l'EBM. *Riv Med Lab - JLM* 2004; 3 Suppl:64-7.
47. Horvath AR. Quale evidenza abbiamo degli esami di laboratorio? *Riv Med Lab - JLM* 2004; 4:274-82.
48. Bruns DE. Laboratory-related outcomes in Healthcare. *Clin Chem* 2001; 47:1547-52.
49. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwing LM et al: The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003; 49:7-18.
50. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen R. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3:3-25.
51. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of Bias and variation in diagnostic accuracy studies. *CMAJ* 2006; 174(4). DOI:10.1503/cmaj.050090.
52. Lijmer JG, Willem Mol B, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282:1061-6.
53. Ioannidis JP. Why most published research findings are false. *PloS Med* 2005; e124. <http://medicine.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pmed.0020124> (data di consultazione: 02.05.2006).
54. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11:88-94.
55. Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, Schmid C, et al. Challenges in Systematic Review of Diagnostic Technologies. *Ann Intern Med* 2005; 142:1048-

- 55.
56. Gabbay J, le May A. Evidence based guidelines or collectively constructed “mindlines”? Ethnographic study of knowledge management in primary care. *BMJ* 2004; 329:1013-7.
57. Glasziou P. Make it evidence informed practice with a little wisdom. *BMJ* 2005; 330:92.
58. Winkens R, Dinant GJ. Rational, cost effective use of investigations in clinical practice. *BMJ* 2002; 324:783-5.
59. Lundberg GD. Laboratory request forms (menus) that guide and teach. *JAMA* 1983; 249:3075.
60. Solomon DH, Hashimoto H, Daltroy L, Liang MH. Techniques to improve physicians’ use of diagnostic tests. A new conceptual framework. *JAMA* 1998; 280:2020-7.
61. van Walraven C, Goel V, Chan B. Effect of population-based interventions on Laboratory utilization. A time-series analysis. *JAMA* 1998; 280:2028-33.
62. Editorial. Using labs best. *Bandolier* 1999; 61:4.
63. Lundberg GD. How clinicians should use the diagnostic laboratory in a changing medical world. *Clin Chim Acta* 1999; 280:3-11.
64. Gopal Rao G, Crook M, Tillyer ML. Pathology tests: is the time for demand management ripe at last? *J Clin Patol* 2003; 56:243-8.
65. Bailey J, Jennings A, Parapia L. Change of pathology request forms can reduce unwanted requests and tests. *J Clin Pathol* 2005; 58:853-5.
66. van Wijk MA, van der Lei J, Mossevald M, Bohmen AM, van Bommel JH. Assessment of decision support for blood test ordering in primary care. A randomized trial. *Ann Int Med* 2001; 134:274-81.
67. Smith BJ, McNeely MDD. The influence of an expert system for test ordering and interpretation on laboratory investigations. *Clin Chem* 1999; 45:1168-75.
68. Bates DW, Boyle DL, Rittenberg E, Kuperman GJ, Ma’Luf N, Menkin V et al. What proportion of common diagnostic tests appears redundant? *Am J Med* 1998; 104:361-8.
69. Smellie WSA, Lowrie R, Wilkinson E. A laboratory based intervention to improve appropriateness of lipid tests and audit cholesterol lowering in primary care. *BMJ* 2001; 323:1224-7.
70. Sinclair D, Saas M, Stevens JM. The effect of a symptom related “gating policy” on ANCA requests in routine clinical practice. *J Clin Pathol* 2004; 57:131-4.
71. Kratz A, Laposata M. Enhanced clinical consulting – moving toward the core competencies of laboratory professionals. *Clin Chim Acta* 2002; 319:117-25.
72. Smellie WSA, Wilson D, McNulty CAM, Galloway MJ, Spickett GA et al. Best practice in primary care pathology: review 1. *J Clin Pathol* 2005; 58:1016-24.
73. Alpert J. The answer you get depends on the question you ask. *Am J Med* 2005; 118:693.
74. Bergus GR, Randall CS, Sinift SD, Rosenthal DM. Does the structure of clinical questions affect the outcome of curbside consultations with specialty colleagues? *Arch Fam Med* 2000; 9:541-7.