

5 • EFFICACIA ED EFFICIENZA

MEDICINA PREVENTIVA

All'interno del NHS certi aspetti della medicina preventiva manifestano una tale efficacia, sia in assoluto sia in rapporto con il settore terapeutico, che mi sembra opportuno, per stabilire una base di paragone, incominciare da essi questa panoramica. Sembra inoltre ragionevole partire dal programma di immunizzazione, che costituisce certamente il settore più efficace della medicina preventiva.

È vero che alcuni elementi del programma, come la vaccinazione antidifterica, non sono sostenuti da RCT, poiché l'epoca dell'introduzione del vaccino ha preceduto quella della tecnica di valutazione; esiste tuttavia un numero sufficiente di osservazioni in appoggio alla loro efficacia.

A parte questa notevole eccezione, tutto il resto del programma è interamente basato sugli RCT: tanto da costituire, secondo me, un esempio per l'intera medicina.

Sono state avanzate tuttavia alcune valide critiche. Il dottor Springett ed io avevamo suggerito (6) l'abbandono della vaccinazione con il BCG qualora il trattamento individuale dei rari casi di tubercolosi fosse risultato meno costoso della loro prevenzione per mezzo di una sistematica vaccinazione. Abbiamo poi confrontato la vaccinazione antitubercolare con altre misure preventive sulla base di un'analisi costi/benefici: in confronto a vasti settori della medicina curativa, il BCG sembra un buon affare tanto che appare scorretto prospettare l'abbandono. Tuttavia, se ci si trova in presenza di due soluzioni ugualmente efficaci, in un certo campo, il NHS trae vantaggio dalla scelta di quella meno costosa, qualunque sia la situazione negli altri settori.

Dal punto di vista dell'efficienza, il programma per le vaccinazioni è ancora suscettibile di miglioramento. Per esempio, l'uso del calcolatore nel West Sussex sembra aver permesso una più completa copertura a costi inferiori (7, 8). Potrebbe essere un esempio da seguire.

I meriti del NHS in questo settore, a parte il programma di vaccinazione, sono miseri; gli si possono contestare omissioni e eccessi.

Tra questi ultimi il più deplorabile a mio parere è l'introduzione dello striscio vaginale per lo *screening* del cancro del collo dell'utero, in quanto chiara dimostrazione di quanto sia pericoloso introdurre nella pratica corrente un nuovo esame prima di averne verificato la efficacia.

Scientificamente la storia è relativamente semplice e le intenzioni originarie erano indubbiamente buone. Questa tecnica venne perciò adottata con entusiasmo, e l'esame divenne rapidamente una pratica di routine negli Stati Uniti da cui si diffuse un po' ovunque. Quando in seguito si suggerì di valutarne l'efficacia con un RCT, questo procedimento, in tutti i paesi interessati, fu considerato contrario all'etica per cui attualmente è difficile testare l'ipotesi con studi osservazionali.

Il tasso di mortalità per cancro del collo dell'utero cominciò a diminuire molto tempo prima dell'introduzione degli *screening* mediante striscio e prosegue la sua regressione ad un ritmo approssimativamente identico nella maggioranza dei paesi. Non abbiamo ancora nessuna prova che esistano differenze a favore delle zone in cui l'esame di *screening* copre un'importante frazione della popolazione femminile. Bisogna aggiungere che, anche se si manifestasse un'importante differenza nei tassi di mortalità specifica, essa sarebbe di difficile interpretazione. Infatti le regioni in cui questa pratica riveste una certa importanza sono quasi sempre le stesse in cui si osserva il maggior numero di donne sottoposte ad isterectomia¹ per molteplici affezioni non correlate al cancro. In questo caso diminuisce in proporzione il numero di donne esposte al rischio e, in un piccolo numero di donne operate, viene ad essere eliminato un tumore non diagnosticato (generalmente in stadio precoce). Se la storia naturale della malattia viene influenzata da quest'ultimo fattore, diventa quasi impossibile decidere quale elemento sia responsabile della differenza osservata.

A mio avviso ci sono ancora delle probabilità che la pratica sistematica dello *screening* mediante striscio vaginale abbia un certo effetto preventivo, ma può darsi che non se ne raggiungano mai le prove, e i servizi sanitari di tutto il mondo continueranno a spendere miliardi nella speranza di prevenire una malattia relativamente rara – anche se molto grave – e il cui tasso di mortalità specifica sta cadendo rapidamente per conto suo.

Quando si discusse questo problema in Gran Bretagna l'atteggiamento dei giornalisti – e anche di molti medici – fu veramente deplorabile. Non si era mai tenuto così poco conto della prova sperimentale e così

¹ Asportazione chirurgica dell'utero.

tanto dell'opinione. Del problema si era impadronita tutta la stampa. Mi ricordo in particolar modo una trasmissione televisiva in cui si esaltava lo striscio vaginale in modo spaventosamente efficace. Nel 1967 nel corso di una conferenza a Cardiff sul tema degli *screening* pronunciavi una frase secondo me del tutto innocua: «Al momento attuale non sono a conoscenza di fatti convincenti che mostrino l'efficacia preventiva dello striscio vaginale». Con mia grande sorpresa tutta la stampa gallese mi mise alla berlina trattandomi, per voce di alcuni colleghi, da «pericoloso eretico», e ricevetti un buon numero di lettere molto offensive. Un insigne collega mi scrisse per accusarmi di «seminare lo sconforto affermando che non esistono cure per il cancro del collo dell'utero». Gli risposi, chiedendogli di fornirmi una prova in favore, ma non ricevetti mai risposta. Peccato, perché mi sarebbe piaciuto saperlo...

Mi sono soffermato alquanto sugli eccessi del NHS; sarò più breve per quanto riguarda le omissioni. Le sue due più vistose carenze riguardano la prevenzione del consumo del tabacco e il controllo delle nascite. Il primo tema è stato trattato in modo adeguato dal *Royal College of Physicians* (10); io non ho nulla da aggiungere alla discussione, se non una precisazione che lo collega al secondo: penso, infatti, che sarebbe opportuno se, prima di consigliare a chiunque di smettere di fumare, ci preoccupassimo di controllare la nostra espansione demografica. Mi sembra del tutto chiaro che il mondo in generale, ed il Regno Unito in particolare, siano sovrappopolati, ma ammetto che si tratta di una valutazione soggettiva; non c'è modo di provarla, ma vorrei lo stesso far notare alcuni punti. Un giorno, se il tasso di natalità non diminuirà, la maggioranza della popolazione si accorgerà che siamo effettivamente troppo numerosi e chiederà che si faccia qualcosa per porvi rimedio. Sarà allora troppo tardi. Non sembra dunque irragionevole sollecitare fin d'ora l'adozione di misure come il controllo delle nascite e l'aborto.²

Una grande percentuale di gravidanze è oggi indesiderata e sarebbe irragionevole per una società che si dice liberale costringere una donna ad avere un figlio che non desidera. Mi rendo perfettamente conto che ai miei colleghi ginecologi questo discorso non è gradito: il raschiamento non è una operazione molto piacevole, e, spesso, le pazienti non provengono dall'alta società.

Questo non toglie che gli aborti pesino sul bilancio del NHS meno delle gravidanze condotte a termine, e che spetti agli specialisti renderli interventi brevi e senza pericoli, praticati ambulatorialmente. Pongo l'aborto al primo posto nelle misure destinate a controllare la espan-

² I dati demografici recenti (1976) riguardanti l'Inghilterra e il Galles sembrano indicare che questo obiettivo è in via di realizzazione.

sione demografica perché personalmente dubito che gli altri metodi proposti (pillola, vasectomia³) abbiano, nell'immediato, l'effetto desiderato.

Un'altra ragione per agire nel campo della limitazione delle nascite è che non esiste alcuna prova che la qualità della vita possa migliorare con l'incremento della popolazione. Si può infine aggiungere che il mondo è senza alcun dubbio sovrappopolato in rapporto alle sue risorse e che è compito dei paesi sviluppati far da esempio su questo problema.

TRATTAMENTO

Se qualcuno dubitasse ancora della necessità di valutare l'efficacia di ogni terapia per mezzo di RCT, gli basterebbe consultare le recenti pubblicazioni che hanno messo in evidenza quanto sia pericoloso supporre che una terapia ben condotta sia sempre efficace, quando non è stata ancora sottoposta a controlli.

A questo proposito uno degli studi più significativi è quello di Mather e collaboratori (11) che hanno realizzato a Bristol un RCT confrontando i risultati del trattamento della cardiopatia ischemica a domicilio e in ospedale (compreso un periodo variabile in unità coronarica).

I risultati, esposti dettagliatamente nel capitolo seguente, dimostrano che non esistono reali vantaggi clinici per coloro che sono ricoverati in unità coronariche in rapporto a coloro che sono trattati a domicilio.

Altrettanto sorprendente è uno studio multicentrico americano sul valore rispettivamente degli anti-diabetici orali, dell'insulina e della dieta nella terapia del diabete dell'adulto.⁴ I risultati – discussi anch'essi in dettaglio nel capitolo seguente – dimostrano che l'uso dei primi (tolbutamide e fenformina) è palesemente svantaggioso e che non esiste alcun vantaggio nella somministrazione di insulina rispetto alla dieta (12, 13).

Nella mia unità di ricerca, Elwood (14) ha dimostrato in modo molto semplice quanto sia priva di fondamento l'opinione generale sulla terapia marziale nel trattamento dei «sintomi» dell'anemia in donne non gravide con livelli di emoglobina tra 9 e 12 g per 100 ml. Sempre nella mia unità, Waters (15) ha smantellato la fiducia nel valore dell'ergotamina⁵ nel trattamento dei casi di emicrania recente.

³ Sterilizzazione maschile.

⁴ Cfr. nota 12, cap. 6.

⁵ Derivato dalla segala cornuta che entra nella composizione della maggior parte dei farmaci tradizionalmente prescritti ai pazienti con emicranie e cefalee (mal di testa), di qualsiasi origine.

Non ho né le capacità, né il tempo di proporre in questa sede una classificazione esauriente delle terapie rispetto alla loro più o meno corretta valutazione: mi limiterò perciò a proporre un'approssimativa classificazione:

1. Le terapie la cui efficacia non è provata da un RCT, ma giustificate dal loro effetto immediato, quali l'insulina nel diabete giovanile, la vitamina B₁₂ nell'anemia perniciosa o la penicillina in certe infezioni.
2. Le terapie la cui efficacia è garantita da RCT. L'esempio migliore è quello della chemioterapia antitubercolare, ma ne esistono evidentemente molti altri.
3. Le terapie per le quali esistono serie dimostrazioni sperimentali del loro effetto ematologico, biochimico, fisiologico, psicologico, ma il cui effetto benefico o dannoso sui pazienti, soprattutto a lungo termine, non è garantito da un RCT. Si tratta di un gruppo molto importante e molto interessante, perché il mondo medico è troppo facilmente impressionato da argomenti provenienti da scienze più fondamentali. L'esempio che colpisce di più a questo proposito è quello della terapia con il ferro citato precedentemente: l'elevazione del tasso di emoglobina è di facilissimo riscontro obiettivo e, prima che Elwood pubblicasse i suoi risultati, tutti credevano che elevando il tasso di emoglobina si guarissero i sintomi tradizionalmente associati alla anemia. Appartengono a questo gruppo anche tranquillanti, antidepressivi e altri psicofarmaci, che in gran parte hanno mostrato effetti psichici misurabili in condizioni sperimentali, ma di cui ancora non si sa se, a lungo termine, faranno ai pazienti più bene che male.
4. Le terapie già codificate prima dell'avvento degli RCT, la cui efficacia non ha potuto essere verificata per ragioni etiche, per cui ci sono buone ragioni per dubitarne. Posso citare due esempi: il trattamento del cancro del seno e di quello bronchiale (16,17).
5. Le terapie per le quali gli RCT hanno dato risultati equivoci. Il migliore esempio è quello della tonsillectomia, che sarà discusso più avanti.
6. Le terapie poco o nulla valutate con gli RCT, benché non vi si opponga alcuna controindicazione di tipo etico. La psicoterapia e la fisioterapia sono probabilmente le rappresentanti più importanti di questo gruppo.

Sarebbe evidentemente utile classificare in questo modo tutte le terapie oggi in uso, ma i dati per farlo sono ora del tutto insufficienti. Se l'efficacia delle terapie è stata valutata scarsamente, l'efficienza con cui esse sono applicate non è stata esaminata per nulla. Vista la diffi-

coltà di parlare di «efficienza» in termini analitici aggirerò l'ostacolo proponendo innanzi tutto una classificazione dei differenti tipi di «inefficienza».

I. Il più evidente tipo di inefficienza è in realtà la combinazione di due separati gruppi: l'uso di terapie inefficaci e l'uso di terapie efficaci al momento sbagliato. Essi sono strettamente connessi; ad esempio si potrebbero classificare tutti i ricostituenti nella categoria dei trattamenti inefficaci, bisogna però riconoscere che la maggior parte di essi contiene composti efficaci in determinate circostanze. Il ferro e le vitamine, che entrano frequentemente nella composizione di ricostituenti, ne sono un esempio: inutili in questo genere di «cocktail», essi hanno delle indicazioni precise in alcuni tipi di anemie o in certi stati carenziali. Allo stesso modo, è importante mettere in luce l'uso cosciente e degno di rispetto di veri «placebo» la cui efficacia è ampiamente dimostrata dagli RCT; il loro uso corretto è dunque da incoraggiare. Ciò che invece va contro all'efficienza è l'uso come placebo di farmaci inefficaci e costosi e rincesce (ma lo si capisce bene!) che nessun laboratorio farmaceutico abbia avviato la fabbricazione di placebo a buon mercato, allettanti e non tossici.

Abbiamo citato or ora, nel gruppo dei placebo inutili e costosi, i ricostituenti; vi si potrebbero aggiungere gli sciroppi per la tosse e gli stimolanti dell'emopoiesi.⁶ Non sappiamo esattamente quale sia l'importanza quantitativa di questo genere di prescrizioni, perché non disponiamo a livello nazionale di dati sufficienti a stabilire un rapporto tra diagnosi e terapia. È inoltre molto difficile dichiarare dei farmaci inefficaci sulla sola base dell'elenco delle prescrizioni, la cui quantità è certamente molto considerevole e, dal punto di vista del NHS, molto onerosa. Ne sono una prova sufficiente le decine di milioni spesi soltanto per le vitamine. In questa stessa prospettiva merita menzionare un altro tipo di inefficienza: l'uso abituale di terapie di per sé efficaci in affezioni per cui la loro efficacia non è stata assolutamente provata, per esempio la vitamina B₁₂ nell'*herpes zoster* e nella sclerosi multipla (18).

All'altro capo della scala si trovano delle terapie la cui efficacia non è stata provata, ma che sono utilizzate «per fare qualcosa». In questi casi è di fondamentale importanza stabilire quale sarà il trattamento più accettabile da parte del paziente e che non compromette il risultato. Un buon esempio è la mastectomia semplice nel tumore della mammella, da non considerarsi inefficiente, mentre in base all'evidenza odierna la

⁶ Farmaci che stimolano la produzione di globuli rossi, chiamati anche: «rigeneratori del sangue».

mastectomia radicale⁷ non può essere classificata come efficiente (16). Un altro esempio è la somministrazione di antimetaboliti⁸ in pazienti con cancro polmonare inoperabile, che non sembra migliorare la qualità della sopravvivenza (17).

II. *Luogo del trattamento inadeguato.* È probabilmente il tipo di inefficienza meno sovente segnalato. È tuttavia probabile che il costo sempre crescente delle degenze ospedaliere vi concentrerà ben presto l'attenzione. Si possono distinguere cinque luoghi in cui il trattamento può essere eseguito:

- a domicilio;
- nell'ambulatorio del medico generico;
- nel corso di una consultazione ospedaliera;
- nell'ospedale regionale;⁹
- e, più recentemente, negli «ospedali di comunità».¹⁰

Solo la tradizione ha finora determinato in quale di questi luoghi dovevano essere trattate le diverse malattie. Fino ad un'epoca molto recente nessuno aveva messo in dubbio questo stato di cose considerandolo come un'ipotesi da verificare. Ho già citato il lavoro di Mather (11) a proposito del trattamento dell'infarto del miocardio a domicilio o in un centro ospedaliero specializzato. Farquharson ha dimostrato (19) che era possibile operare le ernie ambulatorialmente, ma sfortunatamente non ha utilizzato un gruppo di controllo adeguato. Chant et al. (20) servendosi della tecnica degli RCT hanno confrontato i risultati della terapia chirurgica delle varici, per cui è necessaria l'ospedalizzazione, con quelli del trattamento con iniezioni sclerosanti, effettuabile ambulatorialmente. Dopo cinque anni di osservazione, non si è riscontrato alcun importante vantaggio a favore del trattamento ospedaliero, tranne in un piccolo gruppo di pazienti con importanti segni di insufficienza venosa cronica (21).

Studi ora in corso paragonano l'evoluzione di certi malati cronici seguiti nell'ambulatorio dell'ospedale o nello studio del medico generico. Questi lavori dovrebbero non solo fornire gli argomenti necessari per giudicare l'efficienza delle terapie praticate, ma anche la prova che è possibile usare gli RCT e che essi non sono contrari all'etica, il che potrebbe incoraggiare altri ricercatori ad avventurarsi in questo nuovo campo.

⁷ La mastectomia (asportazione chirurgica della mammella in casi di tumore) può limitarsi alla ghiandola mammaria (mastectomia semplice) o estendersi ad una parte della muscolatura toracica ed ai linfonodi e vasi linfatici dell'ascella (mastectomia radicale). Quest'ultima, più mutilante, ha degli importanti inconvenienti: edema del braccio e diminuzione della forza muscolare.

⁸ Farmaci antitumorali.

⁹ District General Hospital.

¹⁰ Community Hospitals.

III. *Inadeguata durata dell'ospedalizzazione.* Non sorprende, conoscendo le caratteristiche economiche e psicologiche del NHS, che la durata media dei ricoveri sia più lunga in Gran Bretagna che altrove. Locan et al. (22) hanno attirato utilmente l'attenzione su questo problema. Si è anche potuto dimostrare che esistevano importanti differenze, per il trattamento di una stessa malattia, tra le diverse regioni e i diversi specialisti. La dimostrazione migliore (e più accurata) è stata fornita da Heasman e Carstairs (23) dal cui lavoro è tratta la figura 5.1.

L'ampiezza delle differenze osservate è veramente allarmante, se si pensa che la degenza in un centro ospedaliero regionale è uno dei trattamenti più costosi che possano essere prescritti, e che la maggioranza dei pazienti desidera lasciare l'ospedale il più presto possibile.

Il solo caso in cui la durata della degenza è stata oggetto di precisa indagine è nuovamente la terapia chirurgica dell'ernia. Blodgett e Beattie (24) per primi pubblicarono una serie controllata, in cui uno dei gruppi veniva dimesso il primo giorno dopo l'operazione. Morris et al. (25) hanno pubblicato su questo soggetto uno studio analogo; più recentemente, un interessante studio di Adler et al. (26) ha dimostrato, così come i precedenti, che una dimissione precoce dall'ospedale non causa nessun serio inconveniente. Tuttavia una simile prassi si impone con difficoltà. La durata media delle degenze per l'Inghilterra e il Galles era ancora, nel 1972, di 7-8 giorni, per gli uomini.

Purtroppo questo tipo di evidenza non ci porta lontano. Non è possibile che tutti i medici abbiano contemporaneamente ragione, ma questo non ci aiuta a determinare la durata ottimale dell'ospedalizzazione. Una volta di più, vi si potrebbe arrivare per mezzo di RCT, ma i criteri di valutazione sarebbero di difficile definizione. Infatti nel caso particolare dell'ernia, il criterio fondamentale sarebbe l'incidenza delle complicazioni che, com'è noto, sono rare; bisognerebbe perciò lavorare su gruppi molto numerosi, con tutte le difficoltà che questo comporta.

Vorrei inserire a questo punto una piccola digressione a proposito dei cosiddetti «ospedali di comunità».¹¹ Il loro obiettivo generale è «fornire cure appropriate in un ambiente adatto, il più vicino possibile alla abitazione del paziente» (27). In questo senso essi hanno un ruolo di rilievo a proposito degli ultimi due punti sull'inefficienza. Il loro diretto antenato è il «Cottage Hospital», che ho avuto modo di visitare nella cittadina di Galashiels. Mi è rimasto un vivido ricordo di quell'ospedale, che corrisponde appieno con i due principali argomenti oggi portati a favore degli ospedali di comunità:

¹¹ Cfr. nota precedente.

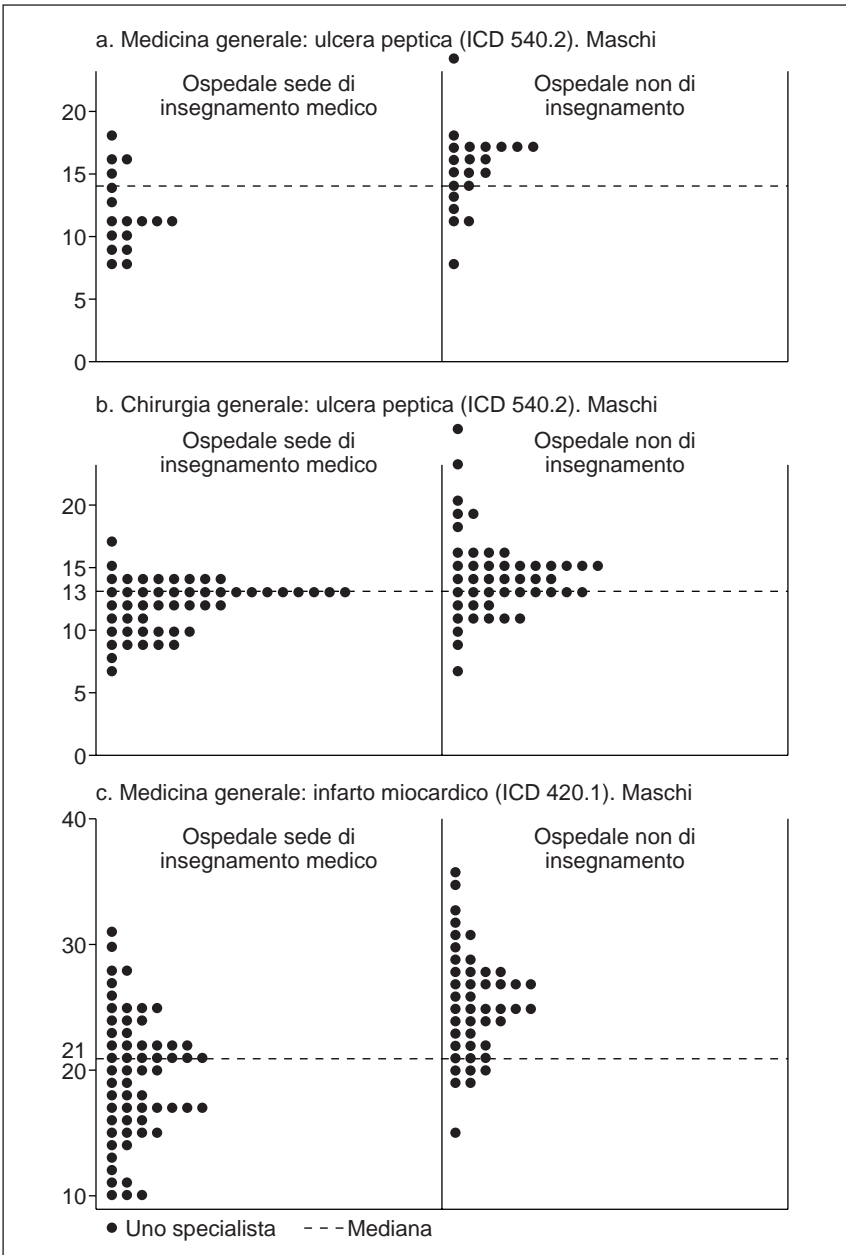


Figura 5.1 • Durata media (in giorni) del ricovero ospedaliero per numero di specialisti presenti, per due diagnosi (Scozia, 1967)

- una possibilità di accesso facile e piacevole. Nella cittadina di Galashiels, dove sono cresciuto, visitare gli amici all'ospedale era come una passeggiata serale o domenicale;
- il fatto di essere centri in grado di catalizzare il sostegno di volontariato e finanziario di tutta una collettività locale, intesa in senso sociologico, in una misura che l'ospedale regionale non può sperare di raggiungere.

I difetti del vecchio Cottage Hospital sono ben noti. Le sue prerogative sono state aumentate al punto da renderlo «pericoloso». Se, invece, si ricoverano negli «ospedali di comunità» solo pazienti per i quali esiste una buona evidenza che sia minimo il rischio nell'essere curati in questa struttura piuttosto che in un centro ospedaliero regionale, si potranno ridurre considerevolmente i due ultimi tipi di inefficienza prima elencati. Per esempio, se un RCT dimostra che una certa malattia può benissimo essere trattata a domicilio, ci sarà un certo numero di pazienti che, per ragioni sociali, non potrà beneficiare di questa possibilità: per essi un ospedale «di quartiere» sarebbe la soluzione ideale. Inoltre, se tutti i pazienti fossero dimessi dai centri ospedalieri regionali dopo una degenza di durata ottimale, potrebbero essere trasferiti negli ospedali di comunità:

- i pazienti che, per ragioni sociali, non potessero ritornare subito a casa loro;
- coloro che devono restare in osservazione, qualora sia possibile proseguirla senza rischio in una struttura di questo tipo.

Detto questo sono consapevole di aver soltanto sfiorato il tema della inefficienza nel settore terapeutico. Ho per esempio tralasciato di parlare della proporzione sempre crescente dei ricoveri per malattie iatrogene¹² ed ho evitato accuratamente di avventurarmi nelle acque torbide dell'amministrazione sanitaria; tuttavia, considerati i miei limitati propositi, penso di aver detto abbastanza.

DIAGNOSI

Contrariamente alla logica ho lasciato per ultimo il problema della diagnosi. Ritengo che in confronto al campo terapeutico e preventivo le

¹² *Malattie conseguenti all'intervento medico: reazioni a farmaci, infezioni contratte in ospedale, «incidenti» in corso di interventi chirurgici o di particolari esami ecc.*

considerazioni teoriche che sottendono al problema della diagnosi abbiano ricevuto molto meno attenzioni. In altri termini sono state spese minori energie per costruire un modello teorico corretto del procedimento diagnostico. Nei casi precedenti, infatti, è generalmente ammesso il concetto di «modificare la storia naturale della malattia» e la tecnica degli RCT è stata messa a punto per verificare le ipotesi ad esso legate. Per la diagnosi, il problema è diverso. In un certo senso si è più avanti che in campo terapeutico. Per lungo tempo i clinici più preparati non avevano quasi altra occupazione che quella di affinare l'arte della diagnosi, la quale restò in tal modo praticamente dissociata dalla terapia, e venne, più o meno coscientemente, considerata fine a se stessa. È ancora possibile trovare dei clinici intelligenti che a voce – ma non per iscritto – sostengono di poter identificare l'obiettivo generale della diagnosi nella descrizione il più completa possibile del paziente in termini medico-scientifici. Se l'obiettivo fosse veramente questo, diventerebbe praticamente impossibile applicare al problema il concetto di efficienza, e le basi economiche della diagnosi equivarrebbero ad un «assegno in bianco» rilasciato al laboratorio o al medico.

È stata data una buona descrizione del processo diagnostico da Medawar (28). Egli considera la diagnosi come un esempio di utilizzazione rapida e praticamente inconscia del metodo «ipotetico-deduttivo», generalmente considerato il metodo scientifico per eccellenza. Il medico, nel processo di ricostruzione della storia clinica del paziente e del suo esame, costruirebbe sulla base dei primi segni e sintomi un insieme di ipotesi che sarebbero in seguito confermate o confutate da altri segni o sintomi. Medawar non tratta dell'uso di esami complementari, ma penso che la sua descrizione, anche se un po' troppo ottimista, sia fondamentalmente corretta: immagino però che accetterebbe l'idea che alcuni clinici siano più «ipotetico-deduttivi» di altri! Sfortunatamente tutto ciò non ci aiuta a definire gli obiettivi preliminari ad una discussione sull'efficacia.

A tal fine, consideriamo la situazione diagnostica più semplice: lo *screening*, nel quale un solo esame è applicato ad una vasta popolazione. Gli *screening* sono divenuti sempre più popolari dopo la fine degli anni Cinquanta, questo nella convinzione che la scoperta di qualsiasi anomalia fosse in prospettiva vantaggiosa per la salute. Per fortuna questa teoria è stata oggetto di numerose discussioni e, poco a poco, si è giunti ad una concezione più razionale. Attualmente si considera utile un esame come test di *screening* «quando esistono prove sufficienti – possibilmente basate su RCT – che la sua applicazione ad una comunità, seguita se necessario da provvedimenti terapeutici, è in grado di cam-

biare il corso naturale di una malattia in una proporzione apprezzabile di casi e ad un costo ragionevole» (29). Questo approccio segna una svolta nella pratica degli *screening*. Sfortunatamente, pochi di essi rispondono attualmente a questi requisiti. Si potrebbe applicare una simile concezione alla diagnosi, nonostante le differenze esistenti tra le condizioni in cui si svolge uno *screening* e quelle della clinica? Chi conduce uno *screening* si trova in una situazione simile a quella evangelica: «Venite a me e vi guarirò dai vostri mali». Egli si colloca chiaramente su di un piano propagandistico e promette dei risultati. Nella situazione clinica è invece il paziente a consultare il medico, che ha il dovere di fare tutto il possibile. Penso tuttavia che, mediante alcune modificazioni rese necessarie dalla situazione clinica, si possa applicare lo stesso approccio dello *screening* alla diagnosi clinica. Per esempio se si applica ad una determinata popolazione un test di *screening* per una certa patologia, corrispondente ai criteri sopra stabiliti, ci sono due possibilità: in una ristretta minoranza dei casi viene riscontrata la malattia – per cui esiste un trattamento efficace – e, nella grande maggioranza dei soggetti sottoposti ad esame, essa è esclusa con un ragionevole grado di probabilità. Lo stesso esame permette di confermare la diagnosi di malattia o di escluderla. Nella pratica clinica, il quadro è più complesso. L'obiettivo principale, come per gli esami di *screening*, è la diagnosi di una malattia curabile; ma, contemporaneamente, si cerca di diagnosticare o di escludere altre malattie, curabili o meno. Per questo sarebbe necessario ponderare il valore degli esami utilizzati nel complesso processo diagnostico, in modo da dare la priorità agli esami destinati a diagnosticare una malattia curabile, ed una importanza minore agli esami capaci di escludere altre malattie curabili, o di diagnosticare malattie incurabili. Ovviamente si provocherebbe in tal modo una riellaborazione radicale dei metodi di valutazione di un esame diagnostico. Fino ad oggi, i principali argomenti a favore di un esame erano – e sono tuttora – la frequenza di risultati anormali, la correlazione tra il risultato dell'esame e la conferma operatoria o post-mortem, e la migliore comprensione del caso (per il medico) grazie ai risultati del test. È importante constatare come fra i criteri di valutazione il benessere del paziente occupi un posto trascurabile; bisogna tuttavia ammettere che spesso ne riceve qualche secondario beneficio!

Penso che la sola speranza di giungere in avvenire ad efficaci procedure diagnostiche sia quella di sottometterle, con le necessarie modificazioni, agli stessi criteri proposti per gli *screening*. Sarà certo difficile procedere alle necessarie estrapolazioni, ma c'è chi, come il professor Card a Glasgow, già vi lavora. Nel frattempo, credo che potrebbero

essere ottenuti dei grandi progressi utilizzando gli RCT. Posso facilmente immaginare gli affascinanti risultati derivanti da esperimenti controllati sulle prove di funzionalità polmonare nei pazienti che si sottopongono ad una visita pneumologica, sulle tomografie¹³ in proiezione laterale ed obliqua in pazienti tubercolotici con escreato positivo, sulle radiografie del tubo digerente nei pazienti di un ambulatorio di gastroenterologia e sulle radiografie della colonna lombare nei pazienti sofferenti di lombalgie. Sarei molto sorpreso se si evidenziasse in tal modo, nell'una o nell'altra situazione, una differenza significativa nella storia naturale della malattia tra i pazienti, a seconda che essi abbiano o no «beneficiario» casualmente dell'esame in questione.

Ritornando ora all'«efficienza» vorrei affrontare prima di tutto il problema della «riproducibilità». I concetti di segno e di sintomo ci sono familiari da molto tempo, ma relativamente poco è stato fatto per standardizzare e rendere riproducibili questi elementi di informazione. La misura dell'errore legato all'osservatore è diventata di moda in medicina agli inizi degli anni Cinquanta, e si è fatto qualche progresso nella standardizzazione della raccolta dei segni e dei sintomi e nella misura della loro riproducibilità. Sono usciti molti eccellenti lavori che tendevano a standardizzare dei questionari riproducibili, la cui validità è stata confermata in differenti occasioni, soprattutto nel campo delle malattie cardiovascolari e respiratorie (30, 31). Anche i sintomi delle malattie mentali sono stati indagati in questo modo, ma persistono vasti settori totalmente inesplorati. I segni obiettivi, dopo un inizio promettente, sono stati oggetto di scarsa attenzione da parte dei ricercatori (32). Moltissimo resta perciò da fare per rendere gli elementi di informazione quanto più utili è possibile.

Per quanto riguarda gli esami di laboratorio mi ha sempre sorpreso lo scarso interesse dei clinici per le variazioni dei risultati in uno stesso laboratorio e fra laboratori diversi. Gli epidemiologi, che fanno misure in popolazioni definite, si servono spesso degli stessi laboratori dei clinici. La loro prima preoccupazione consiste nel valutare la riproducibilità dei risultati ottenuti e, molto spesso, questi sono così allarmanti che essi si rivolgono altrove o fanno per conto proprio il lavoro. Un aspetto particolare delle analisi biochimiche, che personalmente mi infastidisce, è la completa mancanza di interesse per il momento in cui è stato effettuato il prelievo, il tempo trascorso tra prelievo e analisi e la temperatura di conservazione del sangue. È normale, ritengo, che i clinici siano meno assillati degli epidemiologi da questo tipo di problemi, e che non

¹³ Si tratta di «sezioni» radiologiche attraverso un organo o una parte dell'organismo. In confronto ad una radiografia standard, in questo caso, la quantità di radiazioni somministrata è maggiore.

siano perciò interessati alle variazioni dei risultati da laboratorio a laboratorio, ma mi sorprende che non siano più esigenti riguardo alle variazioni all'interno di uno stesso laboratorio. Se ne deve dedurre che il risultato di un esame non è in fondo così importante nella decisione clinica. Si sono riposte molte speranze, per risolvere il problema della riproducibilità, nella automatizzazione degli esami, ma i primi risultati pubblicati in campo ematologico sono piuttosto deludenti (33), e richiedono ulteriori verifiche.

La maggior parte delle osservazioni fatte finora si applica agli esami ematologici e biochimici più che a quelli radiologici, i quali occupano una posizione speciale a causa dei rischi ad essi associati. Sono stati introdotti indubbiamente dei considerevoli miglioramenti per ridurre le dosi di raggi X ricevute dai pazienti per un dato esame, ma il numero degli esami prescritti è in continuo aumento. Probabilmente vi è stato un generale progresso delle tecniche radiologiche, indispensabile per migliorare sia la sicurezza sia la riproducibilità. Lavori originali sulla interpretazione delle radiografie del torace hanno suscitato un vasto interesse per la valutazione dell'errore dovuto all'osservatore in medicina. Questi lavori hanno senza dubbio avuto un certo effetto sulla radiologia in generale, ma permane una deprecabile tendenza a mettere in circolazione le nuove tecniche prima di averle correttamente sperimentate. Per esempio Elwood e Pitman (34), studiando la differenza tra otto radiologi nell'individuazione delle immagini di Paterson-Kelly¹⁴ in 132 soggetti colpiti da disfagia, hanno osservato che il rilevamento di queste immagini variava dal 6 al 59% dei soggetti.

Il prossimo punto che vorrei trattare è l'interpretazione dei risultati degli esami di laboratorio. Alcuni risultati sono facili da interpretare, soprattutto quando non esistono che due alternative, come «positivo» e «negativo» o «presente» e «assente». Ma la maggior parte degli esami, in particolare in campo ematologico e biochimico, sono espressi in termini quantitativi. L'evoluzione di questo problema è riassunta nella figura 5.2.

Ho avuto dapprima l'impressione, da studente, che le caratteristiche diagnostiche dei soggetti malati e di quelli sani fossero nettamente diverse. Penso che questa idea mi abbia poi inconsapevolmente accompagnato fino a quando, con i miei collaboratori, ho incominciato ad esaminare in modo sistematico la popolazione delle vallate del Rhondda e di Glamorgan. Mi fece molto riflettere la distribuzione di frequenze della pressione arteriosa osservata da Miall e Oldham (35) in un cam-

¹⁴ Immagini radiologiche caratteristiche nel caso di turbe della deglutizione (disfagia) dovute ad una anemia cronica da carenza di ferro.

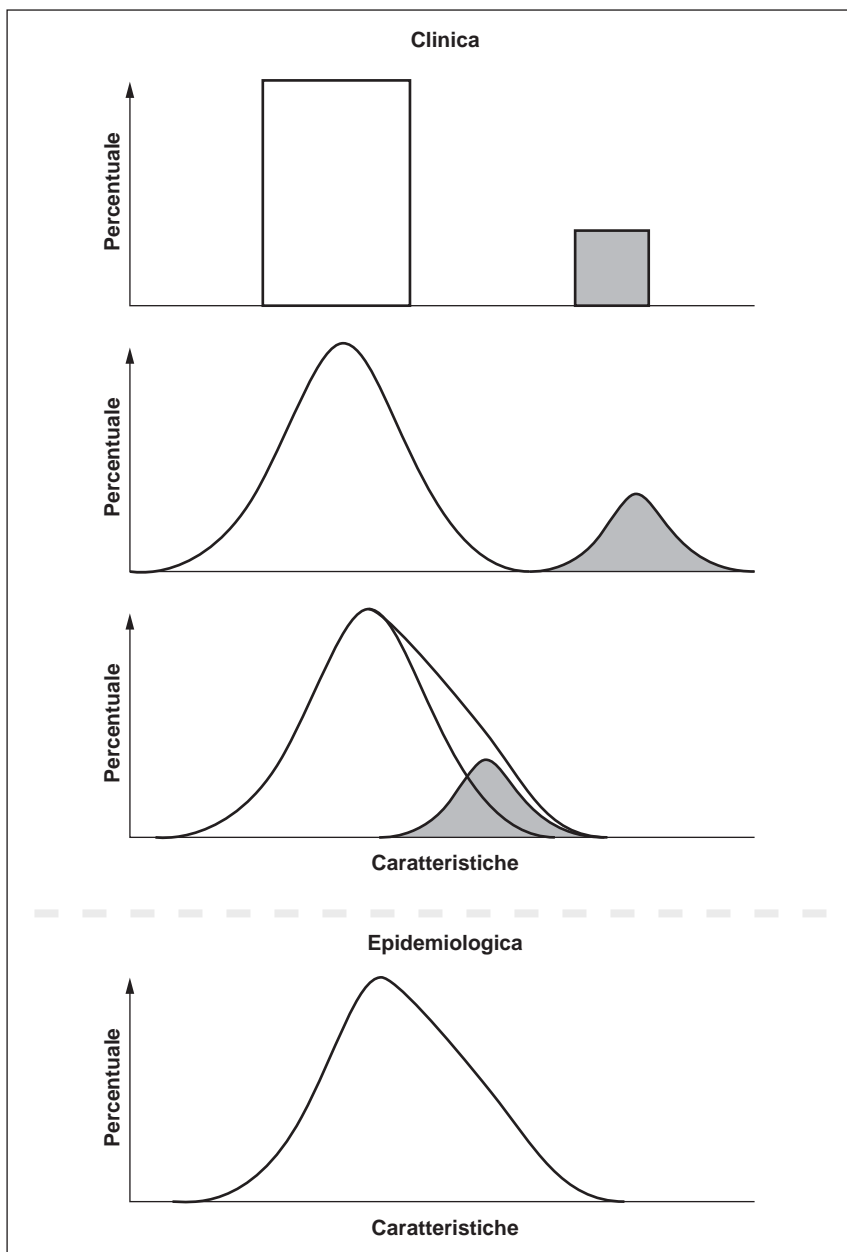


Figura 5.2 • Interpretazione clinica ed epidemiologica delle distribuzioni di caratteristiche biologiche quantitative (aree bianche: soggetti «normali»; aree grige: soggetti «malati»)

pione casuale della popolazione. Un grande specialista dell'ipertensione, Thomas Lewis, mi aveva assicurato che i soggetti con una pressione arteriosa diastolica¹⁵ superiore a 100 mmHg dovevano essere considerati ipertesi: mi sarei dunque atteso una distribuzione «bimodale» simile a quella del secondo diagramma della figura 5.2, mentre abbiamo invece trovato la distribuzione leggermente «asimmetrica» del diagramma in basso.

Da allora abbiamo studiato la distribuzione di frequenza di parecchie variabili fisiologiche e biochimiche. Non abbiamo mai trovato una sola distribuzione «bimodale» anche se, naturalmente, il grado di asimmetria della distribuzione varia considerevolmente. Gli unici resoconti riferiti a distribuzioni bi- o multimodali sono sempre stati spiegati dalla presenza di scelte preferenziali sulle scale di misura,¹⁶ di *bias* di osservazione o di popolazioni selezionate. Il caso limite, nella mia esperienza, è quello dell'esame di 152 dei 153 discendenti viventi del mio bisnonno che era affetto, come molti elementi lasciano pensare, da *porfiria variegata* (36), malattia ereditaria trasmessa in via dominante.¹⁷ C'erano dunque sufficienti elementi per aspettarsi una distribuzione bimodale nel dosaggio delle coproporfirine,¹⁸ ma la distribuzione osservata è solo fortemente spostata a destra ed è quasi impossibile stabilire il limite tra soggetti «normali» e «malati» (figura 5.3). (Un'altra alternativa sarebbe quella di pensare che ho esaminato i discendenti del bisnonno sbagliato!)

Una volta ammessa in parte l'idea di una distribuzione di frequenza non bimodale ma semplicemente asimmetrica, vi furono forti pressioni – più o meno coscienti – per fornire al medico pratico una regola semplice che spiegasse il tutto. Qualcuno (di cui non sono riuscito a rintracciare il nome nella letteratura) introdusse allora il concetto di «valori normali» e propose di definirli come quelle grandezze che non si discostavano dalla media più di due deviazioni standard.¹⁹ Questa idea non ha alcun fondamento teorico: si sa semplicemente che in questo modo si elimina il 5% dei valori osservati in una popolazione per un certo esame. Il presupposto è che la distribuzione di frequenza sia

¹⁵ Pressione arteriosa minima o «residua» durante le brevi pause di rilasciamento del muscolo cardiaco. La pressione sistolica è quella massima provocata dalla contrazione cardiaca.

¹⁶ L'osservatore sceglie o evita involontariamente certe cifre delle scale di lettura (per esempio certi osservatori evitano sistematicamente i numeri dispari o il numero 100).

¹⁷ Cfr. nota 7, cap. 4.

¹⁸ Cfr. nota 7, cap. 4.

¹⁹ La deviazione standard (standard deviation: SD) è, in statistica, una misura della «dispersione» dei valori osservati intorno alla «media» (\bar{x}). Il 95% dei valori è compreso tra due punti corrispondenti il primo a $\bar{x}-2SD$ e l'altro a $\bar{x}+2SD$. Sono perciò esclusi con questo calcolo il 2,5% dei valori a ciascuna estremità della curva nel caso di una distribuzione gaussiana.

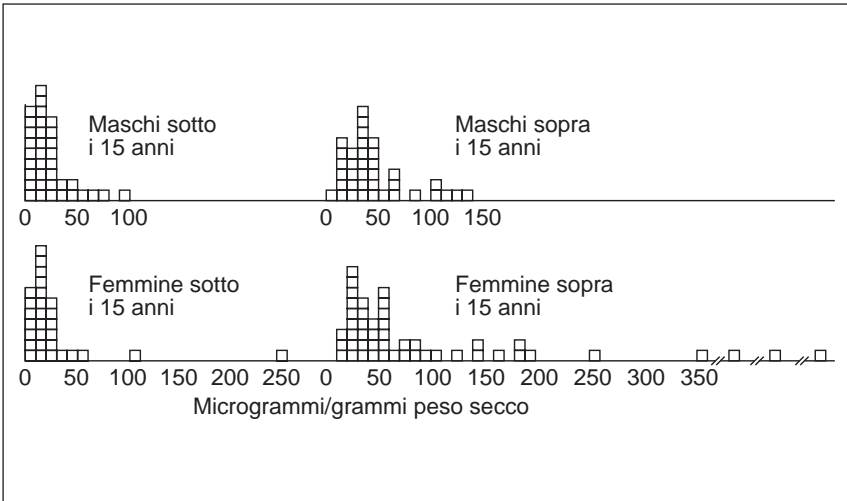


Figura 5.3 • Distribuzione delle porfirine fecali totali in una famiglia inglese affetta da porfiria ereditaria (solo primo campione)

«gaussiana»,²⁰ cioè che gli scarti dai due lati della media siano di uguale importanza, e soprattutto, a prescindere dal problema di sapere se il calcolo della deviazione standard ha un qualche significato nel caso di una distribuzione molto asimmetrica e/o di una popolazione altamente selezionata, ciò presuppone che i medici debbano intervenire quando un valore cade al di fuori di questi limiti.

La sola alternativa che conosco a questo approccio poco soddisfacente è una proposta del dottor Elwood e mia: nel caso di una semplice analisi univariata, si tratta di determinare il punto o i punti della curva di distribuzione di frequenza a partire dai quali un atto terapeutico produce più effetti positivi che negativi. Elwood e Wood (37) ne hanno mostrato una applicazione nella loro ricerca del tasso di emoglobina al di sotto del quale sembra indicata una terapia marziale; altre applicazioni sono state proposte da Keen et al. (38, 39) per la glicemia, da Graham (40) per la pressione intraoculare e da altri autori per la pressione arteriosa (41, 42, 43), benché le indagini sulla pressione arteriosa siano state realizzate con popolazioni in parte selezionate. Va notato a questo proposito come sia limitato il numero delle distribuzioni che sono state analizzate con questo metodo.

²⁰ La «curva di Gauss» è una delle rappresentazioni grafiche della frequenza con cui un carattere biologico quantitativo (per esempio la taglia dei soggetti) «si distribuisce» in una popolazione (ciascuna delle due curve del secondo grafico della figura 5.2 è una curva di Gauss).

Un problema meno importante, ma non meno interessante, è l'ordine in cui gli esami dovrebbero essere effettuati. Sono legato ad una vecchia regola che afferma che, prima di far praticare un esame, bisogna chiedersi quale sarà l'atteggiamento del medico: *a)* se il risultato è positivo; *b)* se il risultato è negativo. Se nei due casi l'atteggiamento previsto è identico, l'esame non va richiesto. All'altro estremo si trova la più recente situazione in cui il medico riceve nello stesso tempo, talvolta persino prima dell'esame clinico, tutta una serie di risultati di esami di laboratorio. Dal punto di vista logico la prima soluzione è irreprensibile, ma le si può contestare il ritardo che può derivarne per la diagnosi. Questa critica non può evidentemente essere mossa alla seconda soluzione, ma il solo studio disponibile sul valore di questo «pacchetto» di esami non è molto rassicurante; si tratta di un RCT tendente a determinare se questo procedimento può ridurre il tempo di degenza. I risultati molto variabili così ottenuti possono naturalmente essere dovuti «al difetto di comprensione da parte dei medici del significato di tutti i risultati» (44).

Un approccio un po' sofisticato degli esami multipli è la cosiddetta scala di Wayne (o di Newcastle) per la diagnosi della tireotossicosi²¹ (45). In essa si è cercato di classificare il valore relativo di una serie di segni e sintomi, al fine di aiutare il clinico a decidere, con il minor margine di errore possibile, l'opportunità di far praticare degli esami più complessi o più costosi.

Questo tentativo potrebbe condurci ad una soluzione più generale del problema. I computer sarebbero di incalcolabile aiuto in questo tipo di lavoro, ma attualmente li si utilizza più per le necessità della ricerca che per quelle dei pazienti.

Non intendo trattare in modo dettagliato il problema dei rischi associati agli esami. Le poche pubblicazioni che hanno posto l'accento su questi rischi hanno un valore limitato a serie selezionate e ad esami molto particolari; sarebbe perciò più che augurabile uno studio «definitivo» sui rischi associati agli esami in generale, in relazione con la mortalità e la morbosità. Lo stesso si può dire per il problema dei costi: non esiste ancora un sistema che permetta di valutare il costo degli esami e soprattutto di confrontare il loro costo nei diversi ospedali. In ogni caso sarebbe opportuno che il loro prezzo fosse affisso a grandi lettere in tutti gli ambulatori.

Si potrebbe aggiungere ancora una parola riguardo all'efficienza degli esami partendo da un altro punto di vista: sui laboratori pesa una

²¹ *Iperfunzione della tiroide che porta ad una «intossicazione» dell'organismo ad opera dell'ormone tiroideo.*

notevole domanda di esami motivati da una «pseudoricerca» che è endemica in numerosi ospedali e particolarmente in quelli di insegnamento. Non aggiungo altro su questo punto, che riguarda soprattutto l'organizzazione interna dei laboratori, perché non ho alcuna competenza in questo campo, ma è chiaro che riveste una grande importanza nell'efficienza globale del servizio.

In breve, il costante incremento delle richieste di esami di ogni tipo (radiologici, anatomo-patologici, ematologici ecc.) rende necessaria l'adozione di misure per una valutazione della loro efficacia e per il controllo della loro efficienza. Spero che gli spunti qui proposti, benché molto incompleti, vadano nella giusta direzione.