
BOLLETTINO

UNIONE MATEMATICA ITALIANA

Sezione A – La Matematica nella Società e nella Cultura

MAURO GASPARINI

La statistica nelle prove cliniche

Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. 6-A—La Matematica nella Società e nella Cultura (2003), n.1, p. 119–140.

Unione Matematica Italiana

http://www.bdim.eu/item?id=BUMI_2003_8_6A_1_119_0

L'utilizzo e la stampa di questo documento digitale è consentito liberamente per motivi di ricerca e studio. Non è consentito l'utilizzo dello stesso per motivi commerciali. Tutte le copie di questo documento devono riportare questo avvertimento.

*Articolo digitalizzato nel quadro del programma
bdim (Biblioteca Digitale Italiana di Matematica)
SIMAI & UMI*

<http://www.bdim.eu/>

La statistica nelle prove cliniche.

MAURO GASPARINI

Riassunto. – *Si descrivono le attività e le responsabilità dello statistico professionale nella conduzione di una prova clinica in una industria farmaceutica o in un ente di ricerca.*

1. – Un felice momento.

È un momento particolarmente felice per la biostatistica in generale e per la statistica clinica in particolare. Gli sviluppi della biologia molecolare e della medicina stanno producendo enormi quantità di dati che devono essere ordinati e interpretati, creando così una domanda di competenze biostatistiche mai vista in precedenza. Gli statistici clinici, cioè gli statistici che lavorano nella ricerca medica su umani, partecipano al clima di euforia grazie anche alla crescente disponibilità di risorse della ricerca medica, alla sua crescente matematizzazione e, con riferimento all'industria farmaceutica, grazie ad un ventennio di impressionante sviluppo. Per doveri legali e per tradizione culturale, l'industria farmaceutica è uno dei pochi settori produttivi che offre agli statistici la possibilità di una carriera non accademica di alto profilo scientifico. Insieme ai centri di cura e ricerca medica pubblici e privati, l'industria farmaceutica partecipa così attivamente alla richiesta e alla produzione di metodologia statistica. Ne è esempio primario la conduzione di quella attività scientifica conosciuta come prova clinica.

Questo articolo è una rassegna discorsiva delle tecniche e delle questioni statistiche che caratterizzano oggi le prove cliniche.

2. – Che cos'è una prova clinica.

«Una prova clinica è uno studio prospettico volto a confrontare gli effetti e il valore di un insieme di interventi su esseri umani, rispetto a un controllo», definizione tratta da Friedman *et al.* (1998), che con Piantadosi (1997), Senn (1997) e con la raccolta Buncher Tsay (1994) costituiscono i primi riferimenti per eventuali approfondimenti. Questa definizione sottolinea che la prova clinica:

- deve avere necessariamente un gruppo di controllo;
- è fatta su umani, caratteristica cui si riferisce l'aggettivo «clinico»;
- è prospettica, a differenza, per esempio, della maggior parte delle indagini epidemiologiche.

Alla definizione si potrebbe aggiungere in maniera più esplicita che per «effetti e valore» di un intervento si intendono tutti gli aspetti legati in primo luogo alla sicurezza e in secondo luogo all'efficacia clinica di un intervento.

Il gruppo di controllo può essere trattato sia con un controllo attivo, nella maggior parte dei casi la migliore terapia esistente, sia con un controllo inattivo, cioè o un placebo oppure nessun trattamento. La presenza di un controllo è a tal punto caratterizzante di che cosa sia una prova clinica, che l'aggettivo «controllato» diventa superfluo e lo si può eliminare dalla pur comune espressione «prove cliniche controllate».

Studi controllati sono comuni ad altri campi, come per esempio la sperimentazione agricola e l'ottimizzazione di un processo industriale. L'applicazione agli esseri umani differenzia la prova clinica da queste altre attività in quanto impone una primaria considerazione dell'etica e della sicurezza. Per esempio, il ricercatore medico ha il dovere di informare i partecipanti sulle caratteristiche della prova clinica e i partecipanti sono liberi di ritirarsi dalla prova in qualsiasi momento (consenso informato).

Il carattere prospettico della prova clinica ne sancisce l'identità di vero e proprio esperimento in senso statistico. La prova clinica è dunque una serie di operazioni che si svolgono sequenzialmente nel

tempo secondo canoni stabiliti e includono in genere la stesura del protocollo, l'approvazione delle autorità sanitarie, l'allestimento del centro o dei centri, l'apertura del database, il reclutamento dei partecipanti, l'applicazione dei trattamenti, la raccolta dati, la chiusura del reclutamento, la chiusura del database, l'analisi e le conclusioni, l'eventuale continuazione dell'osservazione o del trattamento di alcuni pazienti.

3. – Lo statistico e il controllo esterno.

Il protocollo è il documento essenziale che deve preesistere ad una prova clinica e che ne contiene la descrizione dello scopo primario (ed eventualmente degli scopi secondari), delle motivazioni cliniche, del piano statistico e di tutti i dettagli che ne possano permettere la valutazione e l'ispezione da parte di un controllo esterno. Lo statistico partecipa attivamente alla stesura del protocollo e ne ha spesso la responsabilità di firma congiuntamente al ricercatore medico (o meglio, che lo statistico ponga o meno la sua firma sul protocollo è un indicatore di quanta considerazione la statistica goda in quell'istituzione e quindi, in definitiva, un indicatore della qualità di quella ricerca). Che il protocollo di una prova clinica ne dichiari esplicitamente lo scopo può sembrare banale, ma qualsiasi statistico applicato può testimoniare quanto poco espliciti siano gli scopi di tanti studi effettuati in ambienti non sottoposti a controllo esterno, non escluse le università.

A seconda degli scopi dello studio, lo statistico deve specificare in protocollo quale piano degli esperimenti, quali test e in generale quale metodologia statistica intende usare per fini probatori a conclusione della prova clinica. Ciò vale sia che le tecniche statistiche siano di tipo frequentista, nel qual caso la dichiarazione di intenti ben si accorda con l'impostazione oggettivista della metodologia, sia che esse siano di tipo bayesiano, quando informazioni a priori possono essere immesse nella prova clinica d'intesa con il controllo esterno. Per controllo esterno si intende un osservatore imparziale o, più spesso, l'autorità sanitaria (si noti l'uso della parola controllo, che qui è completamente diverso da quello fatto nella Sezione 2). Una

prova clinica è infatti iniziata da uno sponsor, solitamente un'azienda farmaceutica oppure un centro di cura e ricerca pubblico o privato, e soggiace ad una precisa legislazione protettiva dell'interesse pubblico. Il momento più significativo del processo di creazione di questa protezione fu il *Kefauver-Harris Drug Amendments*, approvato dal Congresso statunitense nel 1962 in seguito ai danni causati dal talidomide in diversi paesi, ma la legislazione è in continua evoluzione, come dimostra l'iniziativa planetaria di armonizzazione nota come International Conference on Harmonization (ICH) in corso in questi anni (<http://www.ifpma.org/ich1.html>). Un documento ICH di immediato interesse statistico è la direttiva ICH-E9 *Statistical Principles for Clinical Trials*, International Conference on Harmonization (1998).

Autorità sanitarie sono la *Food and Drug Administration* americana (FDA, <http://www.fda.gov>), la European Agency for the Evaluation of Medicinal Products (EMA, <http://www.emea.eu.int>) e i diversi enti nazionali di controllo della sanità pubblica. Esse sono percepite come una presenza e un riferimento costante da chi lavora nelle prove cliniche, specialmente in quelle condotte da aziende farmaceutiche.

Gli strumenti del controllo non sono solo il protocollo ma anche, per esempio, il diritto/dovere dell'autorità sanitaria di monitorare, anche con ispezioni *in loco*, che lo svolgimento di una prova clinica si svolga secondo i canoni stabiliti nel protocollo e che i dati vengano immagazzinati in maniera totalmente fedele e validata. La qualità dei dati è una questione di immediato interesse per lo statistico, ma non si pensi che lo statistico sia il responsabile della qualità di un database clinico. Infatti la figura del programmatore clinico, sostanzialmente un informatico, responsabile del database è ben distinta da quella dello statistico, responsabile dell'analisi dei dati, almeno nelle grosse industrie farmaceutiche che investono risorse considerevoli nella costruzione e nel mantenimento dei database clinici. Una delle sorprese più piacevoli per uno statistico neoassunto in una grossa industria farmaceutica è l'accorgersi che non si ha la responsabilità del database, ma che si attinge ad esso tramite i servizi informatici dell'azienda.

La qualità dei dati migliora, in ultima analisi, la qualità e la riproducibilità della ricerca. A sua volta, la qualità dei dati è un effetto del ruolo di controllo giocato dall'autorità sanitaria, che dona quindi alla statistica medico-farmaceutica sfide e miglioramento continuo. Da ultimo, non si dimentichi che dalle autorità sanitarie stesse viene una forte richiesta di competenze statistiche altamente specializzate.

4. – Una prova clinica standard di fase III.

Per illustrare, in termini un po' stilizzati, il ruolo dello statistico nello svolgimento di una prova clinica si fa riferimento, in questa sezione, ad una prova clinica altrettanto stilizzata. La prova clinica per eccellenza, a cui si pensa immediatamente a proposito di sperimentazione medica, è la prova clinica di superiorità di fase III (per una discussione delle diverse fasi si veda la Sezione 6). Per superiorità si intende la migliore efficacia di un nuovo trattamento rispetto a un comparatore, quando si sia già dimostrata la sicurezza del trattamento stesso.

Il nuovo trattamento può essere una nuova entità chimica (NCE) di cui si vogliono provare le capacità terapeutiche; il comparatore è spesso la migliore terapia correntemente usata. I risultati positivi di fase III sono pivotali per una azienda farmaceutica, cioè sono quelli che in genere si presentano alle autorità sanitarie per ottenere l'autorizzazione a immettere sul mercato la NCE in una specifica formulazione galenica.

Uno degli ingredienti più importanti del protocollo della prova clinica è la specificazione della variabile principale (*primary endpoint*), che in molti casi sarà una delle tre tipologie seguenti:

- una variabile dicotomica indicante se al paziente giova o meno il trattamento;
- una misurazione clinica di tipo continuo come pressione arteriosa sistolica;
- una durata, per esempio un tempo di sopravvivenza in senso stretto oppure un tempo di assenza di certi sintomi.

Nelle prove cliniche concrete, una descrizione oltremodo precisa delle variabili sarà contenuta nel protocollo. Per esempio, se pressione arteriosa sistolica è la variabile principale, allora ora del giorno, strumento di misurazione, stato di digiuno del paziente e sua posizione saranno specificati in dettaglio.

Una volta identificata la variabile principale, una prova di superiorità viene spesso formulata come problema di verifica di ipotesi relativo ad un parametro di interesse della distribuzione di quella variabile. I parametri di interesse variano dalla semplice media, se la variabile principale è quantitativa, a quantità più specifiche come percentuale di sopravvissuti a due mesi o come l'altezza del picco della curva di concentrazione di una certa sostanza nel plasma sanguigno (C_{\max} , vedi Sezione 7). La scelta dei parametri varia a seconda dell'area terapeutica, delle richieste delle autorità sanitarie e dei vantaggi che si vogliono rivendicare per il nuovo trattamento.

Siano μ_T e μ_C i parametri di interesse della distribuzione della variabile principale sotto il nuovo trattamento e sotto il comparatore. Si è interessati a rifiutare l'ipotesi nulla (anglicismo ormai entrato nell'uso per indicare una ipotesi conservativa sul fenomeno che si sta studiando)

$$(1) \quad H_0: \mu_T = \mu_C$$

a favore di un'ipotesi alternativa

$$(2) \quad H_A: \mu_T > \mu_C,$$

quando si supponga, a scopo esemplificativo, che un aumento del parametro rappresenti una maggiore efficacia del trattamento. Si è usata qui la lettera H per indicare una ipotesi (dall'inglese *hypothesis*), seguita da due punti e dalla sua descrizione formale. Sia inteso infine che in questo articolo «rifiutare un'ipotesi» è un'espressione statistica tecnica usata per indicare che si è raccolta sufficiente evidenza empirica contro l'ipotesi in esame.

All'interno di tale programma, il ruolo dello statistico si realizza almeno nei seguenti momenti:

- scelta delle variabili;

- pianificazione dell'esperimento e della numerosità campionaria;
- casualizzazione e assegnazione dei trattamenti;
- analisi prevista;
- eventuali analisi esplorative non previste.

In base a quanto già detto sull'importanza delle variabili principali, ben si comprende come lo statistico debba collaborare fin dall'inizio con il ricercatore medico nella definizione di quelle variabili, che saranno accompagnate da altre variabili secondarie con ruolo satellitare.

La casualizzazione (*randomization*) è uno dei momenti più squisitamente statistici della prova clinica. Si assegnano i pazienti ai trattamenti secondo un algoritmo fondato sulla generazione di numeri casuali al fine di evitare distorsioni sistematiche dovute a eventuali fattori noti e non noti allo sperimentatore che possono falsare i risultati della prova confondendosi con i trattamenti studiati. L'operazione è tutt'altro che intuitiva e rappresenta una conquista intellettuale della scienza moderna cui hanno contribuito non poco le applicazioni cliniche.

Allo stesso fine di evitare distorsioni sistematiche dei dati si tengono, per quanto possibile, all'oscuro dei trattamenti applicati sia i pazienti che gli sperimentatori (doppio cieco). Si noti a questo punto che il doppio cieco non deve essere in contrasto con il consenso informato; il paziente deve comunque essere informato che sarà randomizzato e che, con certe probabilità, riceverà l'uno o l'altro trattamento, senza sapere quale fino allo scioglimento del cieco.

Per quanto riguarda la pianificazione della numerosità campionaria, si applica alla prova clinica l'idea di scegliere una dimensione campionaria tale da potere raggiungere una potenza adeguata, cioè una probabilità sufficientemente grande di rifiutare l'ipotesi nulla nel caso sia vera una specifica ipotesi alternativa di particolare interesse clinico. In alternativa, si può scegliere una dimensione campionaria tale che l'intervallo di confidenza sul valore vero del vantaggio clinico che si sta stimando sia sufficientemente stretto.

A tale proposito, non mancano nella letteratura clinica le raccomandazioni a preferire un approccio all'analisi delle prove di supe-

riorità basato sugli intervalli di confidenza. Infatti, l'impostazione tradizionale della verifica di ipotesi è basata sul livello di significatività osservato (*p-value*), cioè sulla probabilità presperimentale di osservare un valore così estremo o più estremo di quello osservato, sotto l'ipotesi nulla. Ebbene, si raccomanda sempre più spesso di riportare come risultato dell'esperimento non il *p-value*, ma un intervallo di confidenza per l'effetto di interesse. La verifica del sistema di ipotesi (1) e (2) può del resto essere riformulata come problema di intervalli di confidenza. Si consideri una misura del vantaggio clinico del trattamento rispetto al controllo. Tale misura può essere la differenza dei parametri nelle due popolazioni

$$(3) \quad \delta = \mu_T - \mu_C$$

o il rapporto o altra trasformazione, secondo il modello statistico che si assume. Senza perdita di generalità, si consideri la differenza (3) una misura appropriata di vantaggio clinico e si costruisca un intervallo di confidenza per tale differenza

$$(4) \quad (\widehat{\delta}_I, \widehat{\delta}_S),$$

cioè un intervallo (aleatorio) di numeri reali compresi tra l'estremo inferiore $\widehat{\delta}_I$ e l'estremo superiore $\widehat{\delta}_S$, che abbia una alta specificata probabilità presperimentale di contenere il valore incognito (3). Sotto opportune assunzioni, nella verifica del sistema (1) e (2) si rifiuta l'ipotesi nulla se e solo se l'intervallo di confidenza (4) è un sottoinsieme dei numeri reali positivi. Un intervallo del tipo (4) ha il vantaggio, sia rispetto alla semplice decisione binaria di rifiutare o meno l'ipotesi nulla che rispetto al livello di significatività osservato, di dare allo sperimentatore una stima concreta del vantaggio clinico. Non è difficile fare accettare questa impostazione al ricercatore medico, anche se l'usanza di riportare solo *p-values* come risultato di una prova clinica è ancora abbastanza diffusa.

La critica ai *p-values* raggiunge piena forza all'interno di una visione bayesiana, che nel campo delle prove cliniche viene esemplificata nei lavori di Spiegelhalter *et al.* (1994), Goodman (1999a) e Goodman (1999b). In questi ultimi due articoli si sostiene in particolare come la teoria fisheriana dei test di significatività sia andata fondendosi, nella pratica clinica, con la teoria dei test di ipotesi di

Neyman-Person, con il risultato che l'interpretazione del *p-value* è andata offuscandosi. Negli stessi articoli si raccomanda inoltre il fattore di Bayes come più appropriata alternativa al *p-value*.

A differenza della statistica classica, nell'impostazione bayesiana è possibile parlare di probabilità dell'ipotesi nulla e distinguere tra una probabilità presperimentale $P(H_0)$ e una probabilità postsperimentale, o *a posteriori*, $P(H_0 | \text{osservazioni})$. Grazie al teorema di Bayes, si perviene ad una semplice espressione degli *odds* a posteriori (quote di scommessa, un modo equivalente di esprimere una probabilità)

$$\frac{P(H_0 | \text{osservazioni})}{P(\overline{H}_0 | \text{osservazioni})} = \frac{P(H_0)}{P(\overline{H}_0)} \times \frac{P(\text{osservazioni} | H_0)}{P(\text{osservazioni} | \overline{H}_0)}$$

dove \overline{H}_0 è la negazione di H_0 . Gli *odds* a posteriori coinvolgono gli *odds* a priori $P(H_0)/P(\overline{H}_0)$ e un fattore moltiplicativo che è proprio il fattore di Bayes in favore dell'ipotesi nulla

$$\text{Fattore di Bayes} = \frac{\text{Odds a posteriori}}{\text{Odds a priori}}.$$

Nel caso di ipotesi semplici il fattore di Bayes è un rapporto di verosimiglianza di semplice interpretazione, mentre nel caso di ipotesi complesse è una quantità meno facile da calcolare e da impiegare. In ogni caso, il fattore di Bayes è una misura di supporto empirico dell'ipotesi nulla che Goodman, come altri bayesiani, rivendicano come migliore misura del peso dell'evidenza nei confronti di un'ipotesi clinica. A mo' di riferimento, la Tavola 1 contiene le equivalenze tra *p-value* e fattore di Bayes e proposte per la loro interpretazione, sotto ipotesi semplificatrici di normalità.

Che il piano di proporre il fattore di Bayes come nuovo standard nella pratica clinica abbia successo o meno, la discussione dimostra comunque la vivacità del dibattito statistico circostante le prove cliniche, dibattito che non risparmia di affrontare argomenti anche fondazionali come quello della verifica di ipotesi.

Per quanto riguarda le eventuali analisi esplorative non previste, è opportuno sottolineare che, qualora un risultato di interesse non previsto in protocollo venga identificato durante una prova clinica,

Tavola 1. – Tavola di equivalenze tra *p-value* e fattore di Bayes e peso dell'evidenza suggerito. Tratto da Goodman (1999b).

P-value	Bayes factor (Minimum)	Strength of Evidence against H_0
0.10	1/3.8	Weak
0.05	1/6.8	Moderate
0.03	1/11	Moderate
0.01	1/28	Moderate to Strong
0.001	1/216	Strong to Very Strong

esso può e deve comunque essere segnalato in sede di analisi finale. Si parlerà allora di analisi di tipo esplorativo e non probatorio e tale risultato potrà in certi casi generare nuove ipotesi di lavoro per prove cliniche future. L'espressione usata per indicare questa attività è quella di *subset analysis*, in quanto le analisi esplorative non previste riguardano in genere un particolare sottoinsieme di pazienti, emerso durante la prova, che reagisce al trattamento in maniera imprevista o comunque di potenziale interesse.

5. – Non-inferiorità, variabili *baseline*, prove sequenziali di gruppo.

Nella sezione precedente si è illustrato un test di superiorità di un trattamento, secondo la trattazione classica della verifica di ipotesi su un parametro incognito che si studia in un tipico primo corso di inferenza statistica. Pur in questo quadro standardizzato sorgono varie questioni che rendono interessante anche la statistica per le prove cliniche di tale tipo istituzionale e offrono nuovi spunti alla ricerca. In questa sezione si considerano le variazioni sul tema della prova standard che portano alle prove di non-inferiorità, alla discussione delle variabili di partenza e alle prove sequenziali di gruppo, tre problemi al tempo stesso tipicamente clinici e tipicamente statistici.

Si consideri di nuovo il confronto di due terapie. Qualora lo sponsor intraprenda una prova clinica per dimostrare la superiorità di un trattamento, ovvero per verificare l'ipotesi alternativa (2), si parla di

test di superiorità. Una strada alternativa che in certi casi lo sponsor preferisce percorrere è la prova di non-inferiorità, dove invece si vuole provare l'unione delle ipotesi nulla (1) e (2) quando il comparatore sia la migliore terapia esistente. Per esempio, un'industria farmaceutica potrebbe ottenere una licenza per un farmaco dimostrandolo noninferiore a un altro farmaco già sul mercato.

Non si pensi che ciò possa essere fatto liberamente: non sarebbe etico sperimentare per la pura curiosità di sapere se due farmaci sono equivalenti. Tocca all'autorità sanitaria valutare se fornire o meno l'autorizzazione a procedere alla sperimentazione su umani sulla base di almeno due presupposti:

- una forte incertezza della comunità scientifica nel suo complesso su quale dei due trattamenti sia effettivamente il migliore, in modo da non procurare un danno sicuro a quei pazienti che ricevano il trattamento inferiore (tale principio si applica a tutte le prove cliniche ed è chiamato a volte principio di *equipoise*);

- l'aspettativa che la dimostrazione di equivalenza del nuovo trattamento arrechi un vantaggio ai pazienti o alla società nel suo complesso, per esempio in termini di minori costi, come nel caso dei farmaci generici recentemente introdotti anche nel sistema sanitario nazionale italiano.

Si è ricordato che la verifica del sistema di ipotesi (1) e (2) può anche essere riformulato come problema di intervalli di confidenza. Tale considerazione fornisce una possibile soluzione al problema della non-inferiorità. Si osservi innanzitutto che per «dimostrare» H_0 non è sufficiente non rifiutarla, in quanto assenza di evidenza (contro H_0) non è evidenza di assenza (di prove contro H_0). Se così fosse, basterebbe un esperimento il più possibile inquinato da disturbi casuali e con pochissime osservazioni per ottenere il riconoscimento dell'equivalenza di due farmaci!

La regola di decisione corrente per dimostrare la non-inferiorità è una procedura a due passi:

1. Concordare con l'autorità sanitaria un limite di non-inferiorità $L < 0$ per il vantaggio clinico (3) che abbia un senso dal punto di

vista clinico e pragmatico. Per esempio, L deve essere un numero ben più grande della differenza media tra un eventuale placebo e il comparatore oppure deve rappresentare un ordine di grandezza inferiore rispetto ad una soglia terapeutica quantitativa.

2. Decidere per l'equivalenza se l'intervallo di confidenza (4) giace interamente alla destra di L . Per ottenere una dimostrazione di non-inferiorità occorre quindi non solo che la differenza empirica media sia intorno allo zero, ma anche che ci sia abbastanza informazione campionaria da restringere a sufficienza l'intervallo di confidenza.

La Figura 1 riassume visivamente varie possibilità, inclusi alcuni casi relativi al concetto di bioequivalenza che sarà discusso nella Sezione 7.

Un secondo problema statistico che ricorre frequentemente nelle

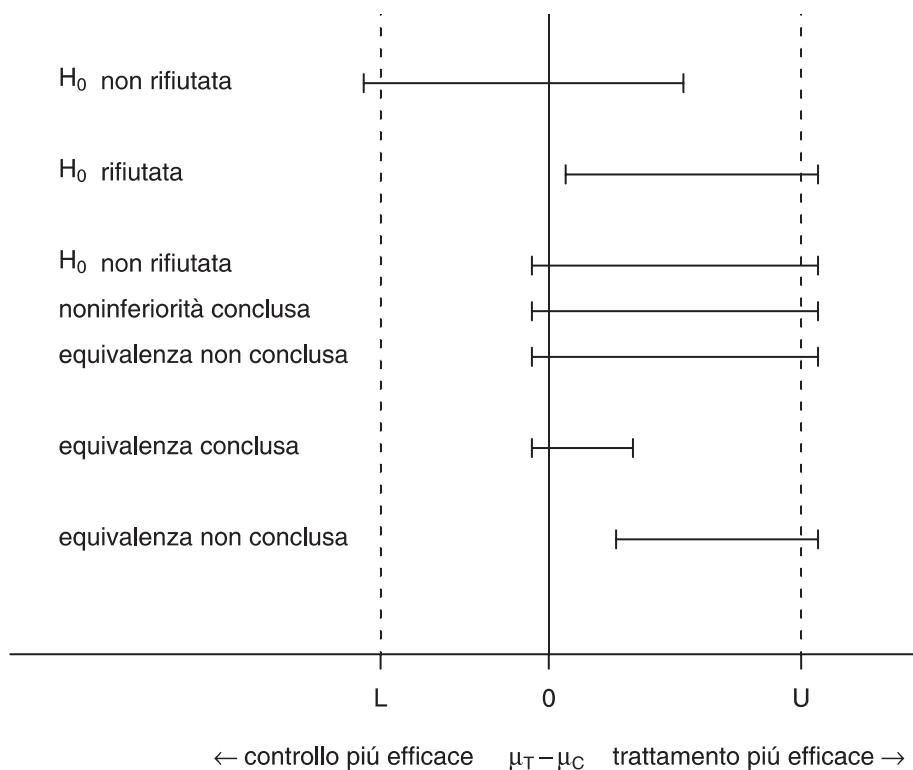


Figura 1. – Intervalli di confidenza per il vantaggio clinico espresso come differenza e varie regole di decisione.

prove cliniche è come trattare i valori di partenza (*baseline values*). Sia Y una variabile dipendente continua misurata in due gruppi paralleli e randomizzati, uno al trattamento (osservazioni Y_T) e uno al controllo (osservazioni Y_C). Siano $Y_{b,T}$ e $Y_{b,C}$ la stessa variabile misurata all'inizio della prova clinica, immediatamente prima dell'applicazione dei trattamenti, rispettivamente nel gruppo di trattamento e nel gruppo di controllo. Tale valore di partenza si chiama *baseline*. Il semplice modello lineare per il confronto tra i due gruppi illustrato nella sezione precedente è

$$Y_T = \mu_T + \text{errore}$$

$$Y_C = \mu_C + \text{errore}$$

cosicché trattamento e controllo si possono confrontare con un test dell'ipotesi (1) contro l'ipotesi (2). Ora, è chiaro che il valore di partenza sarà correlato alla variabile principale sia nel gruppo di trattamento che nel gruppo di controllo. È quindi una buona idea tenerne conto in sede di analisi per diminuire la variabilità delle stime del modello precedente. Se ne può tener conto in due modi:

1. rifacendo l'analisi semplice illustrata precedentemente in termini dei guadagni (differenze dalla *baseline*)

$$G_T = Y_T - Y_{b,T}$$

$$G_C = Y_C - Y_{b,C}$$

2. includendo il valore di baseline come covariata esplicativa.

Il modello risultante dalla seconda impostazione, da preferire perché più «statistica» e meno «algoritmica», è

$$Y_T = \mu_T + \gamma Y_{b,T} + \text{errore}$$

$$Y_C = \mu_C + \gamma Y_{b,C} + \text{errore}$$

chiamato tradizionalmente modello ANCOVA (analisi della covarianza). L'ipotesi nulla è ancora la stessa ipotesi (1) da valutare questa volta in presenza del parametro di disturbo γ .

Le variabili *baseline* sono un esempio dell'importanza di tener conto di fattori predittivi noti nella pianificazione delle prove clini-

che tramite operazioni di stratificazione, un problema di ricerca nel più vasto orizzonte della pianificazione degli esperimenti.

Un terzo importantissimo esempio di questione clinica che si è trasformata in campo di ricerca statistico è rappresentato dalle cosiddette prove cliniche sequenziali a gruppi. In esse si concretizza l'esigenza di condurre delle analisi intermedie (*ad interim*) nelle prove cliniche di una certa durata. Per ragioni etiche infatti, si vorrebbe controllare che una prova clinica già iniziata non abbia già dato indicazioni tali da ritenere il trattamento decisamente inferiore (o anche decisamente superiore, per opposte ragioni) al controllo. L'ispezione dei risultati intermedi di una prova clinica entra in conflitto con l'esigenza di mantenere il cieco — problema parzialmente risolvibile con la recente tendenza a costituire delle *Data and Safety Monitoring Committes* (DSMB) indipendenti dallo sponsor — e soprattutto con il mantenimento di un tasso di errore prestabilito. Ispezioni ripetute infatti sono come test ripetuti e creano una inflazione degli errori reali di prima e seconda specie rispetto a quelli nominali. Si è dunque creato un filone di ricerca che ha portato a delle barriere di rifiuto sequenziali di gruppo (per esempio le barriere cosiddette di Pocock o quelle di O'Brien e Fleming) simili alle barriere del classico test del rapporto sequenziale delle probabilità di Wald. Per la questione si rimanda a Jennison e Turnbull (1999).

6. – Le quattro fasi dello sviluppo clinico.

Nella Sezione 4 si è accennato allo standard di fase III, e la Sezione 5 ha mostrato come anche sullo standard esistano interessanti variazioni. Quando ci si discosta dalla fase III, più rigida e regolamentata, aumentano le possibilità di impostare un lavoro di ricerca statistica originale a partire dalla motivazione clinica. In questa sezione si fa un accenno ai problemi statistici più interessanti nello sviluppo clinico.

Una classificazione tradizionale delle prova cliniche, tipica delle industrie farmaceutiche ma usata anche dai centri di ricerca, le struttura secondo le fasi I, II, III e IV.

In fase I, che rappresenta il passaggio dagli studi preclinici agli

studi su esseri umani, si pone innanzitutto il problema di capire come una sostanza terapeutica venga assorbita, distribuita, metabolizzata ed espulsa dal corpo umano. Tale campo di ricerca viene genericamente indicato come farmacocinetica, una parte della più ampia farmacologia clinica, e può concretizzarsi in una modellizzazione matematica anche molto avanzata. Qualche esempio viene fornito nella Sezione 7. Dalla farmacocinetica (PK), cioè lo studio di ciò che il corpo umano fa al farmaco, si passa immediatamente alla farmacodinamica (PD), lo studio di ciò che il farmaco fa al corpo umano. La comprensione della farmacocinetica di una sostanza in fase I è infatti immediatamente usata per scopi clinici, uno dei quali è tipicamente la stima della dose massima tollerata (MTD). Dal punto di vista statistico, la stima della MTD si traduce nella stima di un quantile di una distribuzione di tolleranza, problema che può essere affrontato secondo diversi schemi sequenziali e adattativi.

Gli studi di farmacologia clinica vengono spesso fatti su volontari sani, a meno che una previsione di tossicità non favorevole costringa per ragioni etiche il ricercatore a sperimentare su pazienti, in modo che la valutazione del beneficio clinico superi i danni rappresentati dalla tossicità. Un tipico esempio sono pazienti oncologici in stadio avanzato, che possono accettare di entrare in una prova di stima della MTD di una sostanza chemoterapica anche in presenza di probabili tossicità.

Le prove cliniche di fase II sono invece fatte già su pazienti selezionati. Il fine è quello di avere delle prime indicazioni di efficacia e di tollerabilità di un nuovo farmaco, di scegliere le più promettenti tra diverse possibili formulazioni o combinazioni ed eventualmente di ottenere una stima della dose minima efficace. Ci si può esprimere dicendo che in fase II si cerca una indicazione dell'attività del farmaco, cioè della sua capacità di indurre quelle modificazioni nel soggetto o nella malattia attraverso le quali si presume di potere poi dimostrare, in fase III, l'efficacia. Mentre l'efficacia ha quindi un significato diretto di beneficio clinico, l'attività è solo la base scientifica sulla quale costruire aspettative di efficacia. Si possono distinguere attività di tipo anti-ipertensivo, antiaggregante, ansiolitico, ecc... Lo stesso tipo di modellistica statistica usata in farmacologia clinica

può essere usata in fase II per capire il comportamento e gli effetti della medicina in popolazioni ristrette di pazienti e per stabilirne meglio la posologia. A tali tecniche si aggiungono, in fase II, i metodi statistici di selezione e ordinamento per scegliere la formulazione più promettente.

Del ruolo pivotale degli studi di fase III, chiamati spesso studi confermativi, si è già parlato nella sezione precedente. Vale la pena sottolineare ora lo sforzo di compendio di tutti i risultati di fase I, II e III che deve essere fatto da uno sponsor per ottenere l'approvazione di una nuova terapia a conclusione degli studi di fase III. Sempre più frequentemente, tale sforzo si concretizza nella redazione di documenti di riassunto chiamati ISS (*integrated summary of safety*) e ISE (*integrated summary of efficacy*), per i quali occorrono competenze statistiche di tipo meta-analitico (la meta-analisi è proprio quel campo della statistica che si occupa di ricondurre ad un'unica conclusione risultati di esperimenti statistici condotti in situazioni sperimentali diverse).

La fase IV riguarda la sorveglianza e l'ulteriore sperimentazione di una sostanza già immessa sul mercato. La popolazione esaminata in fase III non è mai la stessa che le popolazioni reali che useranno il farmaco, c'è quindi sempre la possibilità di effetti tossici collaterali imprevisti. Questo genera un forte interesse della società per il monitoraggio della sicurezza dei farmaci sul mercato (farmacovigilanza e farmacoepidemiologia). Le strutture per tale monitoraggio sono ancora molto carenti, come ancora carente è il coinvolgimento degli statistici. Questo è un campo nel quale c'è molto spazio per un intervento pubblico. Dal canto suo l'industria farmaceutica ha ovvii interessi commerciali nella identificazione di nuove indicazioni terapeutiche per il farmaco e nell'applicazione a popolazioni speciali (anziani, bambini, insufficienti renali) non considerate in precedenza.

7. – Modellistica farmacocinetica in fase I, bioequivalenza e prove *cross-over*.

Un farmaco viene introdotto nell'organismo e assorbito. Per un certo periodo di tempo il farmaco si trova nel plasma sanguigno, do-

ve è possibile misurarne la concentrazione prelevando campioni di sangue a certi intervalli temporali. Si ottiene così una immagine parziale del profilo di concentrazione del farmaco nel tempo per ogni paziente, immagine che serve come punto di partenza per diversi problemi di farmacologia clinica a forte carattere statistico.

In primo luogo occorre stimare, per ogni paziente, il profilo di concentrazione vero a partire da un insieme finito di punti osservati. A tal fine si può ricorrere a modelli compartimentali, del tipo di quelli usati anche in ingegneria, a più complicati modelli fisiologici oppure a modelli nonparametrici. In Figura 2 vengono schematizzati un profilo di concentrazione e i risultati di due prelievi a tempi diversi.

Può darsi che si abbiano a disposizione pochi prelievi per indivi-

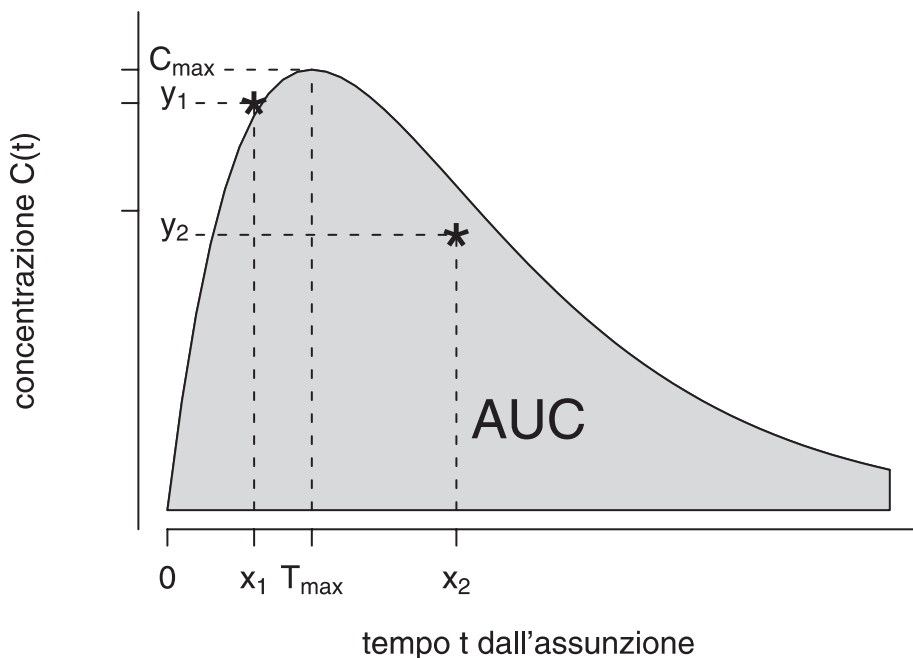


Figura 2. – Schema di un grafico della concentrazione nel plasma sanguigno di una dose singola in funzione del tempo dall'assunzione del farmaco (linea solida), con rispettivi T_{\max} , C_{\max} , $AUC = \int_0^{\infty} C(t) dt$ e due esempi di concentrazioni osservate (con errore) y_1 e y_2 rispettivamente ai tempi x_1 e x_2 .

duo, come per esempio i due indicati in figura, che non permettono di ricostruire con precisione il profilo individuale. La situazione è tipica per esempio di uno studio di fase III dal quale si vogliono ancora ricavare informazioni di farmacologia clinica, ma dove, dato il gran numero di pazienti in centri e condizioni sperimentali diverse, non si dispone di più che qualche dato sparso per paziente. Si può allora adottare un'impostazione chiamata farmacocinetica di popolazione, che consiste nell'adottare per il profilo di concentrazione un semplice modello (per esempio, compartimentale con uno o due compartimenti) con pochi parametri, e nell'ammettere che questi parametri si comportino nella popolazione dei pazienti come effetti casuali. Si può immaginare la complessità e l'interesse statistico dei problemi che ne derivano e per i quali esiste un'ampia letteratura sia bayesiana che non. Si veda per esempio la rassegna Racine-Poon e Wakefield (1998).

Un altro problema fondamentale è quale riassunto unidimensionale del profilo di concentrazione si possa usare per riassumerne le caratteristiche salienti. I riassunti più usati sono l'area sotto la curva (AUC), cioè l'integrale del profilo di concentrazione, il massimo del profilo, chiamato anche C_{\max} , il punto di massimo T_{\max} e altri parametri con diversi significati farmacologici. Si sceglie il riassunto da usare in base a criteri sia statistici che clinici; per esempio, se interessa di più l'assorbimento totale della medicina ci si può orientare sulla AUC, mentre se esistono livelli soglia di concentrazione al di sopra dei quali possono sorgere problemi di sicurezza si porrà attenzione su C_{\max} .

Una volta stabilito il riassunto da usare, si pone il problema di stimarlo a partire dall'insieme finito di punti osservati. Per esempio, la AUC può essere stimata o come funzione dei parametri di un modello compartimentale ipotizzato oppure in modo nonparametrico secondo la regola dei trapezi dell'analisi numerica, usando come nodi i punti campionati. È singolare come in uno stesso problema statistico due capisaldi dell'analisi (l'integrale di Riemann e la sua approssimazione per trapezi) trovino un nuovo significato clinico e statistico.

Il profilo di concentrazione e i riassunti di interesse si complicano

nel caso di dosi multiple, per la cui discussione si rimanda a testi specializzati, per esempio Gibaldi e Perrier (1982).

Il problema della non-inferiorità illustrato nella Sezione 5 ha come equivalente, in farmacologia clinica, il problema della bioequivalenza, cioè della uguale biodisponibilità relativa. Una prova clinica di bioequivalenza rappresenta spesso un canale abbreviato per ottenere la licenza a produrre un farmaco generico. Le industrie farmaceutiche investono cifre enormi nella ricerca di nuovi farmaci (si parla di una cifra media intorno ai 300 milioni di euro per lo sviluppo di un singolo nuovo farmaco) e vengono ricompensate con un brevetto, limitato a un periodo tra dieci e venti anni, secondo i casi. Scaduto il brevetto, qualsiasi competitore è libero di produrre lo stesso farmaco in versione generica. Se si tratta di una molecola sufficientemente vicina alla molecola di brevetto, allora si può ottenere la licenza a produrla come generico con prove che dimostrino non la non-inferiorità dell'efficacia ma la uguale biodisponibilità. Si tratta cioè di dimostrare che i due prodotti, già simili dal punto di vista chimico, vengano anche assorbiti dal corpo umano in maniera equivalente nelle due formulazioni galeniche.

Nei casi più semplici, per dimostrare la uguale biodisponibilità di un farmaco ci si può ridurre alla considerazione di un riassunto del profilo di concentrazione, come AUC o C_{\max} . Supponiamo dunque che si voglia dimostrare l'ipotesi nulla (1) quando μ_T e μ_C siano la AUC media, per esempio, di un candidato generico e di un farmaco brevettato. Poiché ci si vuole proteggere sia dal caso di ipodosaggio che dal caso di iperdosaggio del farmaco, si considera dimostrata la bioequivalenza delle due formulazioni quando l'intervallo di confidenza per la differenza delle AUC medie sia interamente contenuto in un intervallo $[L, U]$ ritenuto clinicamente appropriato. Di nuovo, la Figura 1 illustra la situazione.

Recentemente si sta discutendo di sostituire a questi criteri di equivalenza media altri criteri più attenti alla sostituibilità del farmaco dal punto di vista del paziente individuale.

Le prove cliniche di bioequivalenza sono in genere condotte seguendo un piano degli esperimenti incrociato (*cross over*), piuttosto

che a gruppi paralleli. Ciò significa che ogni paziente riceve più di un trattamento in periodi diversi e viene casualizzato per ricevere una particolare sequenza di trattamenti. La Tabella 2 illustra il piano per due trattamenti e due periodi.

Un piano incrociato semplice è possibile solo se l'applicazione del trattamento non lascia tracce rilevanti dopo un opportuno periodo di *wash out*. Se persistono per esempio tracce della sostanza nel sangue anche dopo un lungo periodo di tempo e se ne vuole stimare l'entità (*carry over effect*), si deve utilizzare un piano incrociato più complicato, come per esempio quello illustrato in Tabella 3.

D'altra parte, quando è praticabile, un piano incrociato permette che ogni paziente sia il controllo di se stesso, in quanto i trattamenti vengono confrontati entro ogni paziente, con il risultato di eliminare la variabilità interindividuale.

Il piano incrociato è un altro concetto statistico sviluppatosi soprattutto grazie alla sperimentazione clinica. Un filone di ricerca ben sviluppato all'interno del disegno degli esperimenti (DOE) studia i piani incrociati incompleti con più di due trattamenti e le analisi possibili per i piani incrociati tramite modelli lineari con o senza effetti casuali.

Tabella 2. – Piano degli esperimenti incrociato per due trattamenti A e B in due periodi.

sequenza	periodo	
	1	2
1	A	B
2	B	A

Tabella 3. – Piano degli esperimenti incrociato per due trattamenti A e B in tre periodi.

sequenza	periodo		
	1	2	3
1	A	B	A
2	B	A	B

8. – Due parole sul software.

Le parole esprimono idee e, più modernamente, lo stesso fa il software. Se linguaggi *procedure-oriented* come SAS dominano ancora la ricerca clinica per motivi quasi di inerzia burocratica, ciò non impedisce a linguaggi più moderni come S-plus di essere sempre più presenti anche in questo campo. Si veda per esempio Millard e Krause (2001).

9. – Farmacogenomica.

I progressi della biologia molecolare e della genetica sono destinati a cambiare la ricerca farmaceutica sotto forma di quella disciplina conosciuta come farmacogenomica. La scoperta di nuovi agenti terapeutici viene facilitata dalla possibilità di ottenere *in vitro* cloni di specifici geni umani e di studiarne l'attività, dalla possibilità di manipolare con sicurezza geni di agenti patogeni pericolosi e dalla capacità di indurre mutazioni molto selettive contro obiettivi selezionati, in modo da aumentare l'efficacia e la sicurezza delle terapie.

L'ultima decade ha visto la decodifica del genoma umano, cosicché ora conosciamo l'anatomia dei geni, ma non conosciamo la loro fisiologia o patologia. Recenti tecniche note con il nome di genomica funzionale stanno dando alla ricerca una svolta fondamentale in questa direzione, e al momento attuale la tecnica più promettente è quella dei *microarray*. Grazie a un singolo *microarray*, il biologo è in grado di ottenere simultaneamente l'espressione genica di migliaia di geni in due situazioni sperimentali a confronto, permettendo il monitoraggio della risposta genica ad una terapia.

I *microarray* possono anche essere usati per investigare la variabilità genetica tra gli individui (*genotyping*), variabilità responsabile della predisposizione individuale alle patologie e della diversa risposta ai farmaci.

Senza entrare nei dettagli della tecnologia, si sottolinea qui che l'analisi dei risultati da *microarray* comporta una fortissima domanda di competenze statistiche per disegnare l'esperimento, per tenere conto delle diverse fonti di variabilità e per costruire dei modelli

statistici adeguati alla complessità e multidimensionalità dei dati.

Grazie alla genetica, ci si attende che le prove cliniche possano aggiungere alle due importanti dimensioni della sicurezza e dell'efficacia quella dell'individualizzazione, ultimo importante tassello di un'opera di enorme impatto scientifico e sociale.

RIFERIMENTI BIBLIOGRAFICI

- C. BUNCHEER - J. Y. TSAY (Eds.), *Statistics in the pharmaceutical industry*, Dekker, second edition, 1994.
- L. FRIEDMAN - C. FURBERG - D. DEMETS, *Fundamentals of Clinical trials*, Springer, third edition, 1998.
- M. GIBALDI - D. PERRIER, *Pharmacokinetics*, Dekker, 1982.
- S. GOODMAN, *Toward evidence-based medical statistics. 1: The P-value fallacy*, *Annals of Internal Medicine*, **130** (1999a), 995-1004.
- S. GOODMAN, *Toward evidence-based medical statistics. 2: The Bayes factor*, *Annals of Internal Medicine*, **130** (1999b), 1004-1013.
- International Conference on Harmonization, *Statistical principles for clinical trials*, <http://www.ich5e.html> (1998).
- C. JENNISON - B. TURNBULL (Eds.), *Group Sequential Methods with Applications to Clinical Trials*, CRC Press, 1999.
- S. MILLARD - A. KRAUSE, *Applied Statistics in the Pharmaceutical Industry*, Springer, 2001.
- S. PIANTADOSI, *Clinical Trials. A methodological perspective*, Wiley, 1997.
- A. RACINE-POON - J. WAKEFIELD, *Statistica methods for population pharmacokinetic modelling*, *Statistica Methods in Medical Research*, **7** (1998), 63-84.
- S. SENN, *Statistical Issues in Drug Development*, Wiley, 1997.
- D. SPIEGELHALTER - L. FREEDMAN - M. PARMAR, *Bayesian approaches to randomized trials*, *Journal of the Royal Statistical Society series A*, **157** (1994), 357-416.

Mauro Gasparini, Dipartimento di Matematica, Politecnico di Torino
gasparin@calvino.polito.it