Pillole di EBM - Capitolo 8

Data 05 dicembre 2006
Categoria scienze_varie

La differenza è significativa?
Significato della P e dell'intervallo di confidenza

La "P"

Gli esami non finiscono mai, diceva qualcuno. Nel nostro caso potremmo dire che quelle che non finiscono mai sono le difficoltà. Abbiamo visto nei capitoli precedenti che i risultati di uno studio possono essere espressi in termini di rischio assoluto, di NNT e di rischio relativo. Una volta però che abbiamo espresso i risultati dello studio in termini matematici, dobbiamo affrontare una questione cruciale: **la differenza che abbiamo trovato tra il trattamento e il placebo (o il farmaco di confronto) è significativa dal punto di vista statistico oppure è dovuta semplicemente al caso?**
Si tratta di una domanda fondamentale, come si può capire, perché solo se la differenza è significativa possiamo attribuire un reale valore clinico al risultato che abbiamo trovato (non è sempre detto che ad un risultato statisticamente significativo corrisponda un reale beneficio clinico, come vedremo in seguito, ma per ora accantoniamo questo aspetto).
Ma come si fa a dire che il risultato è statisticamente significativo? Come possiamo essere sicuri che il dato non è dovuto al semplice gioco capriccioso della casualità ? Affermare che un risultato è statisticamente significativo mentre è puramente dovuto al caso viene definito dagli esperti in statistica **"errorealfa"** . Gli studiosi si sono messi d'accordo e dicono: accettiamo una probabilità di errore alfa inferiore al 5%; **se questa probabilità è inferiore al 5% possiamo dire che il risultato è statisticamente significativo** . A noi non interessa sapere come viene calcolata questa probabilità di errore, visto che il nostro mestiere è quello di fare i medici e non i matematici. Importa invece conoscere come si esprime. Ebbene si esprime con la **famosa"P"** . Perciò quando troviamo che un determinato farmaco ha ridotto il rischio di infarto di una certa percentuale con valore di "P" < 0,05 significa che il dato è significativo perché la "p", cioè la probabilità di errore alfa è inferiore al 5% (ricordo che dire 0,05 e dire 5% è la stessa cosa espressa in modo diverso). Vi sembra poco una probabilità di errore alfa del 5%? Se uno studio dimostra che un farmaco riduce l'ictus di un tot rispetto al placebo e che il risultato è statisticamente significativo perché la P è uguale a 0,049 vi sentireste tranquilli? Una P = 0,049 significa che c'è il 4,9 % di probabilità che quel risultato sia semplicemente casuale. Mettiamola in altro modo: voi state comodamente viaggiando con la vostra auto e ad un certo punto dovete attraversare un ponte e vi accorgete che all'imbocco qualche bontempone ha messo un bel cartello in cui sta scritto: **"Attenzione questo ponte crolla all'incirca ogni 20,4 passaggi"** (in altre parole c'è una probabilità del 4,9% che crolli quando state passando voi). Vi sentireste tranquilli? Forse si, ma neppure molto. Questo per dire che tanto più bassa è la "P" tanto più possiamo stare tranquilli che il caso non ci ha messo lo zampino. In effetti la significatività statistica non è altro che un'invenzione dell'uomo, una convenzione: che differenza può esserci per una "p" di 0,049 e una di 0,051? Eppure la prima permette di affermare la significatività statistica, la seconda no!
En passant dirò che vi è anche **l'errorebeta** , che è l'opposto dell'errore alfa: si considera statisticamente non significativo un risultato che invece lo è. Gli studiosi si sono messi d'accordo per accettare una probabilità di errore beta inferiore a 0,1 (cioè 10%).

L'intervallo di confidenza

E' arrivato adesso il momento di complicare ancor più la faccenda e di spiegare il famigerato **intervallo di confidenza** .
Per farlo ritorniamo un attimo indietro all'RR, cioè al rischio relativo. Ricorderete che esso può essere inferiore a 1 (l'intervento fa meglio del controllo), superiore a 1 (l'intervento fa peggio del controllo) oppure può essere uguale a 1 (intervento e controllo sono uguali).
Prendiamo per esempio uno studio in cui sia stato confrontato un nuovo farmaco con un farmaco di controllo e che abbia valutato tre end-points: mortalità totale, infarti non fatali, ictus.
Sotto sono riassunti i dati finali dello studio in cui RR esprime il rischio relativo del farmaco rispetto al controllo:

MORTALITA' TOTALE: RR = 1,20
INFARTI NON FATALI: RR = 0,80
ICTUS: RR = 1,0

Quindi:
- il farmaco testato ha fatto peggio del controllo sulla mortalità totale: un RR = 1,20 significa che con il farmaco la mortalità è aumentata del 20%
- il farmaco testato ha fatto meglio del controllo per quanto riguarda l'infarto non fatale, riducendo il rischio del 20%
- non c'è differenza tra farmaco e controllo per quanto riguarda l'ictus.

Però, per la serie le cose semplici non vanno bene, dobbiamo considerare che l'RR che abbiamo trovato è una **stima unitaria del rischio** , ma in realtà non è così, non può essere un singolo numero. Per farla breve dobbiamo immaginare di ripetere per 100 volte la stima dell'RR (non spaventatevi, non spetta a noi farlo). Così facendo troveremo 100 RR diversi, il 95% di questi RR viene a trovarsi entro un certo intervallo che viene detto **intervallo di confidenza al 95%**

(IC95%).
Lo so, così non è troppo chiaro, e allora facciamo un esempio. Supponiamo di trovare un RR = 0.60. Ripetiamo per 100 volte la stima dell'RR e troveremo 100 valori e poniamo che 95 di questi valori siano compresi tra 0.50 e 0.70, mentre 5 valori cadranno fuori da questo intervallo, cioè potranno essere più piccoli di 0.50 o più grandi di 0.70. Allora diremo che l'RR trovato è di 0.60 ma il suo IC95% va da 0.50 a 0.70. **Detto in altro modo: l'intervallo di confidenza al 95% esprime il range entro cui può trovarsi l'RR con una probabilità del 95%** .
Per capire ancora meglio l'IC95% possiamo pensare ad un treno che viaggi da Napoli a Milano. A una certa ora noi sappiamo che il treno dovrebbe trovarsi in una stazione intermedia tra Roma e Firenze, ma in realtà non sappiamo esattamente quale sia per cui diciamo che con una probabilità del 95% si trova comunque in un tratto compreso tra Firenze e Roma.
Facciamo un altro esempio. In un trial si trova che un farmaco riduce la probabilità di sviluppo di cancro mammario del 25%, quindi l'RR sarà 0.75, ma il suo IC95% va da 0.60 a 0.85. Ciò vuol dire che in realtà la riduzione del rischio (con una precisione del 95%) potrebbe, nella migliore delle ipotesi, essere del 40% (RR = 0.60) e nella peggiore del 15% (RR =0.85).
Chiaramente in questo caso sia l'RR che il suo IC95% sono sempre inferiori a 1: l'RR, pur spostandosi all'interno del suo intervallo, non potrà mai essere uguale o superiore a 1. Si può affermare pertanto che il **risultato così trovato è statisticamente significativo perché non potrà mai essere che il farmaco faccia venire più tumori mammari del controllo**. Tutto questo si trova scritto negli studi come segue: **RR 0.75; IC95% 0.60-0.85** .
Prendiamo un esempio opposto: un farmaco non solo non riesce a ridurre il rischio di infarto rispetto al controllo, ma addirittura lo fa aumentare del 20%. Pertanto l'RR sarà 1.20. Ma il suo IC95% trovato va da 1.10 a 1.40. Ciò vuol dire che nella migliore delle ipotesi l'aumento del rischio è del 10% (RR = 1.10), nella peggiore del 40% (RR = 1.40). In questo caso l'RR sarà sempre superiore a 1 perché il suo IC95% va da 1.10 a 1.40. Anche qui diremo che il **risultato così trovato è statisticamente significativo perché non potrà mai essere che il farmaco faccia venire meno infarti del controllo** . Tutto questo si trova scritto negli studi come segue: **RR 1.20; IC95% 1.10-1.40** .
Prendiamo infine il terzo e ultimo caso, quello in cui l'IC95% **comprenda il numero 1** , dove sappiamo essere la parità tra i due gruppi. Poniamo che uno studio abbia valutato l'efficacia di un farmaco nel ridurre il rischio di infarto e di ictus. I risultati trovati sono i seguenti:
INFARTO: RR = 1,20; IC95% = 0.90 - 1.40
ICTUS: RR = 0.80; IC95% 0.70 - 1.20
Come si può ormai facilmente arguire, per quanto riguarda l'infarto il farmaco ha provocato un aumento del rischio del 20% (RR = 1.20) ma il suo intervallo di confidenza COMPRENDE IL NUMERO 1, quindi nella migliore delle ipotesi il rischio potrebbe venire ridotto del 10%, nella peggiore potrebbe essere aumentato del 40%. Per quanto riguarda l'ictus il farmaco riduce il rischio del 20% (RR = 0.80) ma anche qui il suo intervallo di confidenza COMPRENDE IL NUMERO 1 e il rischio, per quanto ne sappiamo, potrebbe nella migliore delle ipotesi essere ridotto del 30% e nella peggiore essere aumentato del 20%. In entrambi i casi diremo che la differenza trovata per i due end-points **non è statisticamente significativa** e in effetti non sappiamo realmente se il rischio aumenti o si riduca, dato che l'intervallo di confidenza può essere sia maggiore che minore dell'unità.
Gli esempi dovrebbero essere chiari, ma se non avete compreso vi conviene rileggerli perché si tratta di un punto fondamentale per capire i risultati degli studi così come poi li troverete riportati.
In conclusione possiamo dire che l'IC95% ci permette di vedere se l'RR è statisticamente significativo: se comprende il numero 1 non lo è, mentre lo è se al contrario non lo comprende.

Riporto ora i risultati del Women's Health Study, uno studio randomizzato e controllato in cui erano state arruolate quasi 40.000 donne di almeno 45 anni e senza storia di malattie cardiovascolari o neoplastiche. Le pazienti vennero trattate con aspirina oppure placebo e seguite per oltre 10 anni. Lo scopo dello studio era di stabilire se l'aspirina fosse in grado di ridurre l'incidenza di neoplasie. I risultati del farmaco rispetto al placebo sono stati:
- sviluppo totale di neoplasie: RR 1.01; IC95% 0.94-1.08 (vi è stato un aumento del numero totale di neoplasie dell'1% nel gruppo aspirina, ma il dato non è significativo perchè l'intervallo di confidenza al 95% va da 0.94 a 1.08, quindi i due trattamenti hanno pareggiato)
Lascio ora a voi l'interpretazione degli altri dati:
- cancro mammario: RR 0.98; IC95% 0.87-1.09
- cancro del colon: RR 0.97; IC95% 0.77-1.24
- cancro polmonare: RR 0.78; IC95% 0.59-1.03
- mortalità totale: RR 0.95; IC95% 0.81-1.11
- mortalità da cancro polmonare: RR 0.70; IC95% 0.50-0.99

Siccome le cose appena dette sono troppo semplici qualcuno si diverte a complicarle e invece di riportare i risultati di uno studio con la formula classica li riporta in questo modo:
RIDUZIONE DEL RISCHIO RELATIVO DEL 25% (DAL 10% AL 45%)
Conviene non spaventarsi: la riduzione del rischio relativo del 25% corrisponde ovviamente ad un RR = 0,75 mentre i due numeri tra parentesi corripondono ai due estremi dell'IC95%, dove 10% corrisponde a 0,90 e 45% corrispoponde a 0,55. Così se volessimo scrivere il risultato secondo la formula classica avremmo:
RR = 0,75; IC95% = 0,55-0,90
Nello stesso modo se troviamo scritto:
RIDUZIONE DEL RISCHIO RELATIVO DEL 15% (DA - 10 A 25)
sappiamo che RR = 0,85 e l'IC95% va da 0,75 a 1,10.

Ovviamente così come è possibile calcolare l'intervallo di confedenza per il rischio relativo, lo si può fare anche per la riduzione del rischio assoluto e per l'NNT. Prendiamo questo esempio: in uno studio l'intervento ha ridotto l'outcome considerato in termini assoluti del 2,5% (ARR = 0,025 = 2,5%) e quindi si ha un NNT di 40; tuttavia l'intervallo di

confidenza dell'ARR varia da 1% a 5%(o se si preferisce da 0,01 a 0,05), perciò l'NNT varierà da 20 (ARR = 5%) a 100 (ARR = 1%) . Questo vuol dire che nella migliore delle ipotesi basta trattare 20 soggetti per evitare l'end-point e nella peggiore occorre trattarne 100.

Interpretazione clinica dell'intervallo di confidenza

Come si può immaginare l'IC95% può essere largo, stretto, vicino o lontano dalla linea di parità. Questo può dipendere da vari fattori come per esempio la numerosità della popolazione arruolata nello studio, la sua potenza statistica, ecc. Però quello che interessa a noi è un'altra cosa, cioè la possibilità di trarre delle conclusioni di tipo clinico dalla morfologia e dalla posizione dell'IC95% rispetto alla linea di parità.
Se l'intervallo di confidenza è largo e una delle estremità si avvicina alla linea di parità (per esempio 0.40- 0.99 oppure 1.01-2.04) significa che il risultato è poco riproducibile e il suo significato clinico probabilmente poco importante.
Invece se l'IC95% è stretto e lontano dalla linea di parità (per esempio 0.35-0.55 oppure 1.60-1.80) significa che il risultato è riproducibile e probabilmente anche importante dal punto di vista clinico.
Infine se è stretto ma vicino alla linea di parità (per esempio 0.88-0.99 oppure 1.01-1.12) significa che il risultato è riproducibile ma clinicamente forse poco importante.
Insomma possiamo dire che tanto più l'IC95% è stretto tanto più elevata è la riproducibilità del risultato mentre tanto più è lontano dalla linea di parità tanto più cresce l'importanza clinica del dato. Al contrario tanto più una delle estremità dell'intervallo di confidenza è prossima alla linea di parità probabilmente tanto meno importante è l'intervento dal punto di vista clinico. Infatti se noi troviamo che con un certo farmaco si è avuto una riduzione degli infarti del 30% con RR 0,70 ma con un IC95% di 0,50 - 0,99 potremo certamente dire che il risultato è statisticamente significativo ma in realtà non sappiamo bene se la riduzione degli eventi sia del 50% (nella migliore delle ipotesi) o solo dell'1% (nella peggiore delle ipotesi). Per motivi precauzionali conviene sempre tener per buona l'ipotesi peggiore per cui si può ragionevolmente concludere che l'intervento ha un impatto clinico probabilmente modesto.

RenatoRossi