

Rilevanza dei risultati di uno studio clinico

P. Ravani¹, S. Andrulli²

Divisione di Nefrologia e Dialisi, Azienda Istituti Ospitalieri di Cremona

Divisione di Nefrologia e Dialisi, Azienda Istituti Ospitalieri di Lecco

Clinical relevance of study results

The critical appraisal process of available evidence includes the evaluation of clinical importance of the study findings. This should coincide with the minimum worthwhile effect expected by the investigators in the study design and planning. In hard outcome studies it can be quantified by the absolute risk difference between groups and its reciprocal, known as number needed to treat to avoid one adverse event, or benefit (NNTB). The number needed to treat to produce harmful consequences of treatment (NNTH) should also be taken into account. Finally these effect measures, like risks and incidence proportions, are usually rough estimates of the true effects, due to non-complete follow-up of the observations under study. Underlying assumptions and design issues are especially important to assess the clinical relevance of any results. (G Ital Nefrol 2006; 23: 163-72)

KEY WORDS: Clinical epidemiology, Measures of disease occurrence, Effect measures, Number needed to treat

PAROLE CHIAVE: Epidemiologia clinica, Misure di malattia, Misure di effetto, Numero di pazienti da trattare

Introduzione

La rilevanza clinica dell'effetto favorevole di un trattamento (o sfavorevole di un fattore di rischio) riguarda l'entità della modificazione del rischio (probabilità di un evento) e non coincide con la significatività statistica del risultato osservato in uno studio clinico. Discuteremo in una successiva rassegna come la significatività statistica ci aiuti a misurare l'imprecisione di una stima (ruolo del caso). L'importanza di un effetto riguarda la natura della stima stessa e il suo significato clinico. Per agevolare la comprensione del problema, dobbiamo chiarire cos'è e come si misura il rischio (misure di malattia) e come si misurano gli effetti di un'esposizione (come un fattore di rischio o un intervento) sul rischio di evento (misure di effetto).

Tipo di evento: hard outcomes e surrogati

In molti lavori clinici l'effetto di un'esposizione è misurato confrontando l'occorrenza di malattia in base ai diversi livelli dell'esposizione stessa. Se, ad esempio, volessimo conoscere i benefici di un nuovo trattamento anti-ipertensivo (esposizione A vs. B) potremmo rilevare l'occorrenza di eventi cardiovascolari non fatali come angina, infarto, TIA

o ictus, oppure fatali (mortalità) in un gruppo sperimentale (A) e in un gruppo di controlli (B). In questi casi gli eventi presi in considerazione sono chiamati hard outcomes, per distinguerli dagli outcomes surrogati o indicatori indiretti di eventi maggiori. I cosiddetti hard outcomes sono generalmente variabili di risposta binarie, ossia con due possibili valori o livelli (successo/insuccesso, evento sì/no, malattia sì/no, morto/vivo). Al contrario, gli indicatori indiretti di evento sono variabili che possono assumere più valori o essere addirittura variabili quantitative continue (come i valori pressori, gli spessori parietali o i volumi ventricolari cardiaci). Il tipo di evento più adatto per la valutazione della rilevanza clinica del risultato di uno studio è costituito da hard outcomes e a questi ci riferiremo nel presente lavoro. Purtroppo a molti quesiti clinici la risposta viene data soltanto mediante studi di outcomes surrogati, (sicuramente meno costosi ma anche meno affidabili) anche quando è possibile studiare hard outcomes. Non sempre infatti un effetto favorevole su un outcome surrogato si traduce in paragonabili effetti su hard outcomes. Infine, bisogna tener presente che anche negli studi di hard outcomes il risultato di uno studio clinico può essere riportato utilizzando diverse misure di effetto e il tipo di misura scelta può enfatizzare il risultato osservato e, conseguentemente, influenzare il lettore. Ne deriva che l'impatto dell'intervento studiato può

essere sottostimato o sopravvalutato indipendentemente dai dati realmente osservati. È pertanto importante conoscere le misure di *malattia* e di *effetto* prodotte dagli studi di hard outcomes per imparare a pesare l'importanza dei risultati in termini clinici (1).

Misure di malattia

Il rischio

Il clinico è interessato a stime numeriche sia della frequenza delle malattie che dell'effetto di diverse esposizioni (fattori di rischio o trattamenti). Le misure di frequenza più utilizzate in epidemiologia clinica sono la prevalenza e l'incidenza (Fig. 1). La *prevalenza* è prodotta in studi trasversali (una sola rilevazione) e corrisponde al rapporto tra numero di soggetti malati (casi rilevati) in un preciso momento sul totale dei soggetti studiati (malati e non) in quel preciso momento. La prevalenza non permette di stabilire la probabilità di ammalare (rischio) e quindi non verrà trattata in questo lavoro. Questa probabilità è invece stimata con studi longitudinali che permettono di riferire le osservazioni alla dimensione temporale, ossia definendo i nuovi casi in rapporto al numero di soggetti a rischio (liberi da evento) in un certo intervallo di tempo (*incidenza*). Il rischio è la misura di malattia ideale sia per lo studio della frequenza di malattia che dell'effetto delle esposizioni. Quando si parla di *rischio* di malattia o di evento di un soggetto ci si riferisce alla probabilità che un soggetto sviluppi la malattia in questione. Questo rischio non è direttamente misurabile in una singola persona dal momento che una persona ha solo due possibilità: sviluppo o non sviluppo della malattia. Per questa ragione si descrive il rischio come proporzione di soggetti che sviluppa la malattia in un più largo gruppo di persone. Se in un campione composto da N soggetti, D soggetti sviluppano la malattia durante un determinato *intervallo di tempo*, allora il rapporto D/N (frequenza relativa) stima il rischio medio (detto anche tasso, rischio incidente, rischio assoluto, incidenza cumulativa) di malattia della popolazione da cui il campione è stato estratto nel periodo di tempo considerato. La stima del rischio richiede che tutti gli N soggetti siano liberi da malattia all'inizio dello studio (siano a *rischio*) e siano seguiti per tutta la durata dell'intervallo di tempo in cui il rischio viene misurato. Questo accade tanto più facilmente quanto più è rapida l'evoluzione della malattia e breve è lo studio. Il rischio individuale di malattia è anche chiamato *proporzione incidente* e, poiché stime del rischio medio della popolazione vengono applicate al singolo individuo, i due termini sono sinonimi per indicare l'occorrenza di morte o di qualsiasi altro evento binario (con due possibili valori, sì/no). Va notato che per interpretare un rischio è indispensabile definire un intervallo di tempo durante il quale il rischio si sviluppa, perché il rischio è la *velocità* con cui l'e-

Misure di frequenza di malattia

Prevalenza: misura di frequenza in un dato momento (t_0), è la quantità di malattia (D) attribuibile a ciascuna unità di denominatore (ciascuna di N persone),

$$P \text{ in } t_0 = D/N$$

Incidenza: stima la velocità con cui un dato evento (E) si propaga in una data popolazione (N) a rischio a partire da t_0 nell'intervallo di tempo (t_1-t_0),

$$I = E \text{ in } (t_1-t_0)/N$$

Fig. 1 - Definizioni di prevalenza ed incidenza.

vento evolve nel tempo. Non avrebbe senso parlare di rischio del 5% di sviluppare un infarto miocardico senza specificare un intervallo di tempo durante il quale l'evento potrebbe verificarsi. La stessa quantità sarebbe enorme o insignificante a seconda della durata più o meno breve dell'intervallo di tempo. Dire che il rischio di sviluppare uremia terminale in un gruppo di soggetti è 0.1 (10%) in 10 anni, e dire che ciascun individuo ha una probabilità di sviluppare uremia terminale di 0.1 (rischio del 10%) in 10 anni è la stessa identica cosa. Il rischio nell'intervallo di tempo t non ha dimensioni, ha un range $[0,1]$, è una frequenza relativa ("una parte del tutto" o quoziente) e può modificarsi nel tempo in relazione alla suscettibilità individuale a sviluppare la malattia (evento).

Densità di incidenza

Negli studi delle malattie acute, come alcune infezioni delle vie respiratorie o urinarie (IVU), o delle complicanze acute di alcune procedure come l'insufficienza renale acuta da mezzo di contrasto iodato, gli studi clinici arruolano una cosiddetta *coorte statica e chiusa*, ossia composta da un certo numero di soggetti per cui l'osservazione è completa (Fig. 2). In questi studi, ogni unità entra nello stesso momento (periodo di arruolamento) ed esce prima della fine dello studio solo se è avvenuto l'evento (*coorte fissa* o chiusa). In questi casi lo studio longitudinale permette di stimare il rischio incidente (incidenza cumulativa) nel periodo dello studio. Nell'esempio, 2 episodi di IVU osservati su 7 soggetti seguiti per 7 mesi, permettono di stimare un rischio di IVU di 0.285 per soggetto (28.5 per 100 soggetti), in 7 mesi. Nello studio delle malattie croniche le cose si complicano perché l'evoluzione più lenta del quadro clinico comporta, insieme ad una durata maggiore dello studio, anche una maggiore possibilità che alcune osservazioni diventino *incomplete*,

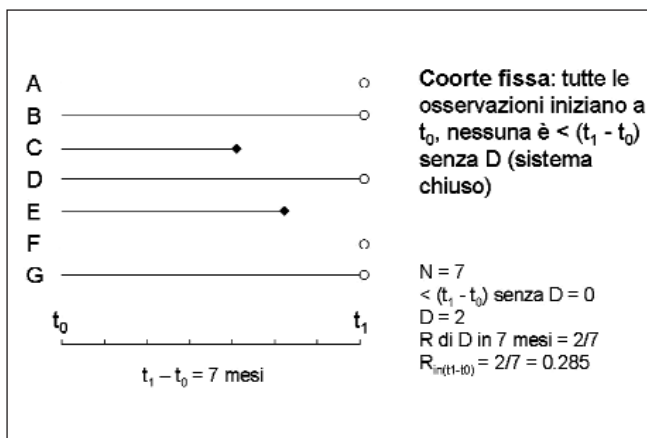


Fig. 2 - Definizione di coorte fissa o chiusa e stima del rischio durante l'intervallo t (rischio o proporzione incidente).

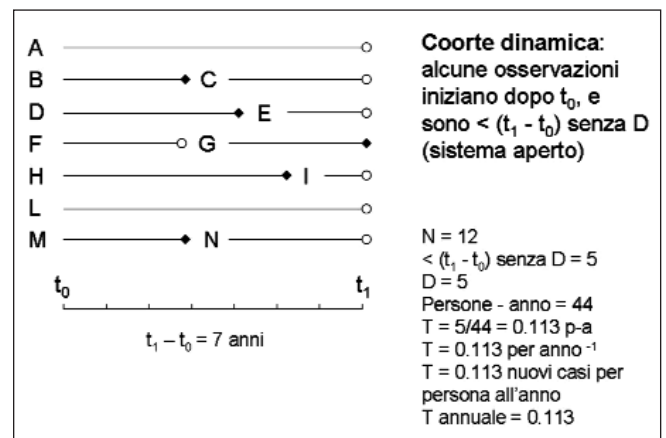


Fig. 3 - Definizione di coorte dinamica o aperta e stima del rischio annuo (dalla densità di incidenza).

ossia escano dalla coorte prima della fine dello studio senza che l'evento si sia verificato (per esempio per l'esistenza di rischi competitivi dovuti ad eventi di tipo diverso da quello in questione, per trasferimento presso altri centri o ritiro del consenso). Le osservazioni possono anche iniziare in tempi diversi (*coorte dinamica* o aperta, Fig. 3). In questi studi si perdono possibili eventi successivi al termine dell'osservazione e la coorte si riduce di dimensioni se altre unità non vengono arruolate. Per queste ragioni, il rischio viene stimato calcolando un diverso rapporto, la *densità di incidenza* (o tasso di incidenza). Si tratta del rapporto tra il numero di eventi nel periodo considerato e la massa di persone a rischio (esperienza tempo-persona). Il numeratore del rapporto è lo stesso utilizzato per il calcolo dell'incidenza cumulativa, ma al denominatore compare un nuovo termine: il numero di unità epidemiologiche (persone) per unità di tempo calcolata come somma delle unità di tempo cui ciascuna osservazione ha contribuito. Ad esempio nella Figura 3 il soggetto A contribuisce al denominatore per 7 unità tempo-persona, il soggetto B per 3, il C per 3, D per 4, E per 2 e così via per un totale di 44 persone-anno (massa a rischio o esperienza tempo-persona della coorte). Ciascuna unità di tempo è equivalente alle altre indipendentemente dal soggetto che ad essa ha contribuito. Nell'esempio, 5 eventi si sono verificati in 7 anni, su 44 persone-anno. Per cui il tasso di incidenza è 0.113 per ciascuna unità tempo-persona (persona-anno), ossia 0.113/anno oppure anche 0.113 anno⁻¹ (11.3 casi per 100 persone anno⁻¹). La densità di incidenza ha proprietà diverse dall'incidenza cumulativa. Innanzitutto questa stima del rischio non è più un vero e proprio rischio (probabilità) perché è comparsa la dimensione "tempo" al denominatore del rapporto (quoziente). Inoltre ha un range $[0, \infty]$, con valore dipendente non solo da quanto è elevato il rischio stimato ma anche dall'unità di tempo scelta. Infatti, come una velocità, il rischio incidente è un concetto istantaneo perché il deno-

minatore è un'unità di tempo e lo stesso numero di casi produrrà diversi valori di densità di incidenza per diverse unità di tempo. Per esempio, 10 casi per 1000 persone-mese (0.01 casi per persona-mese, 0.01 mese⁻¹) corrispondono alla "velocità" di 0.12 casi per persona-anno (0.12 anno⁻¹), nonostante i valori numerici siano diversi. Se scegliamo le decadi la densità di incidenza diventa > 1 (1.2 per 10 persone-anno o 12 secolo⁻¹). Pertanto l'unità di tempo del denominatore è arbitraria e serve per esprimere il valore medio di una misura istantanea sopra l'unità scelta. Inoltre va notato che la dimensione della densità di incidenza è il reciproco del tempo, il che implica che essa possa essere interpretata come l'inverso del tempo all'evento. Nell'esempio di prima, il tempo medio o atteso, all'evento è 1/0.12 anni ossia 8.33 anni. Per tutte queste ragioni, e poiché è molto più comodo ed intuitivo ragionare in termini di rischio (senza dimensioni) si usa moltiplicare la densità di incidenza per il tempo al fine di ottenere una stima più vicina alla probabilità. Per esempio se una popolazione di 10000 persone ha una densità di incidenza di insufficienza renale terminale di 6 casi per 10000 persone-anno (0.0006 anno⁻¹), allora la stima del rischio annuo diventa 0.0006 e per 6 mesi 0.0003. Il rischio annuo è la stima del rischio medio in un anno. Purtroppo questa approssimazione non è applicabile per l'intero range di valori della densità di incidenza, ma solo per valori inferiori a 0.2, mentre produce distorsioni progressivamente più evidenti per valori superiori. Fortunatamente nella maggior parte degli studi clinici i rischi sono bassi e l'approssimazione è valida, mentre per rischi maggiori, in genere, la durata dello studio è minore e i follow-up incompleti sono meno frequenti. Un altro problema di queste trasformazioni è la variabilità temporale della densità di incidenza. Ad ogni buon conto, per intervalli di tempo molto piccoli essa è relativamente costante e i rischi possono essere agevolmente derivati con il metodo delle analisi di sopravvivenza (2).

Test di verifica

1) Il rischio di malattia o morte:

- È una frequenza relativa o quoziente (una parte sopra un tutto)
- È la probabilità dell'evento
- Richiede che sia specificato l'intervallo di tempo in cui l'evento può verificarsi
- Non ha dimensioni
- Tutte le precedenti

2) La densità di incidenza:

- È un rapporto che ha lo stesso numeratore del rischio
- Stima il rischio medio nell'unità di tempo
- È diversa dal rischio perché ha la dimensione del reciproco del tempo
- Ha un range da 0 a + infinito, mentre il rischio va da 0 a 1.

e) tutte le precedenti

3) Il rischio di un evento è stimato dalla densità di incidenza perché:

- Perché si studiano campioni e non la popolazione obiettivo
- È difficile che tutti i soggetti studiati siano seguiti in modo completo fino all'evento
- Perché è necessaria la dimensione temporale
- Perché il rischio è la prevalenza mentre noi siamo interessati all'incidenza
- Tutte le precedenti

La risposta corretta alle domande sarà disponibile sul sito internet www.sin-italy.org/gin e in questo numero del giornale cartaceo dopo il Notiziario SIN

Misure di effetto

Negli studi clinici ciò a cui siamo realmente interessati è il contrasto tra diverse esposizioni o interventi. Infatti un soggetto può essere esposto ad un fattore di rischio e poi sviluppare la malattia senza che tra questa e il fattore di rischio ci sia alcuna relazione. Altrettanto può accadere per un trattamento preventivo e il mancato sviluppo di una complicanza. Per misurare l'associazione (positiva o negativa) tra esposizione e malattia, ed esplorare possibili effetti causali, è necessario confrontare l'esperienza dei soggetti esposti con ciò che sarebbe successo agli stessi soggetti in assenza di esposizione. Il confronto ideale sarebbe infatti tra soggetti esposti e gli stessi soggetti non esposti allo stesso tempo, una situazione non realistica. Quello che in pratica si fa è sostituire questi non esposti "ideali" con un altro gruppo di soggetti che permette di stimare ciò che sarebbe accaduto se fossero stati non esposti i soggetti esposti. Questo processo di sostituzione è il problema cruciale di tutti gli studi clinici. Il modo per raggiungere una

Misure di effetto

Differenza di rischio: $RD = [R_0 - R_1]$

Rischio relativo: $RR = [R_1 / R_0]$

Effetto relativo: $RE = [R_0 - R_1] / R_0$
 $RE = [RR - 1]$, se $R_1 > R_0$
 $RE = [1 - RR]$, se $R_1 < R_0$

NNT: $1/RD$

Fig. 4 - Definizione delle principali misure di effetto.

sostituzione *valida* per promuovere la confrontabilità e il perché *associazione* non significa necessariamente *causalità*, saranno discussi in successive revisioni dedicate agli aspetti metodologici degli studi clinici.

Assumendo che i soggetti esposti e non esposti siano altrimenti uguali, le misure di malattia che abbiamo precedentemente conosciuto vengono confrontate per valutare l'effetto dell'esposizione. Possiamo per esempio stimare la differenza o il rapporto tra i rischi degli esposti (rischio negli esposti, R_1) e dei non esposti (rischio basale o rischio nei non esposti, R_0). La differenza di rischio (risk difference, RD) è semplicemente la differenza aritmetica $[R_0 - R_1]$. Il rapporto (o rischio relativo, RR) è $[R_1 / R_0]$. In assenza di effetto RD è zero, mentre RR è 1. Se $R_1 < R_0$ (l'esposizione protegge) la RD è positiva e il RR è < 1 . Se $R_1 > R_0$ (l'esposizione aumenta il rischio) la RD è negativa e il RR è > 1 . Un'altra misura di effetto interessante è l'effetto relativo (riduzione o aumento relativo del rischio, relative effect RE), calcolata come il rapporto tra RD e R_0 (Fig. 4). Per esempio un RE di 2.1 implica un RR di 3.1 e un rischio negli esposti superiore a tre volte quello dei controlli. La RD ha un range $[-1, +1]$, mentre il RR un range $[0, \infty]$. Poiché abbiamo a che fare con stime del rischio piuttosto che con i rischi veri, è bene tener presente che il range della differenza tra densità di incidenza è $[-\infty, +\infty]$ e che quello del rapporto delle densità di incidenza è $[0, \infty]$. Come già accennato, sotto certe condizioni le differenze e i rapporti tra stime danno accettabili approssimazioni dei veri RD e RR. Se per esempio abbiamo a che fare con bassi valori di densità di incidenza e brevi intervalli di tempo (T), allora $RR = [R_1/R_0] = [I_1 * T / I_0 * T] = [I_1/I_0]$. Al ridursi di T (tendente a zero), anche le densità di incidenza tendono a zero, ma il loro rapporto resta diverso da zero e si avvicina quello del RR. Nelle analisi di sopravvivenza, il rapporto delle

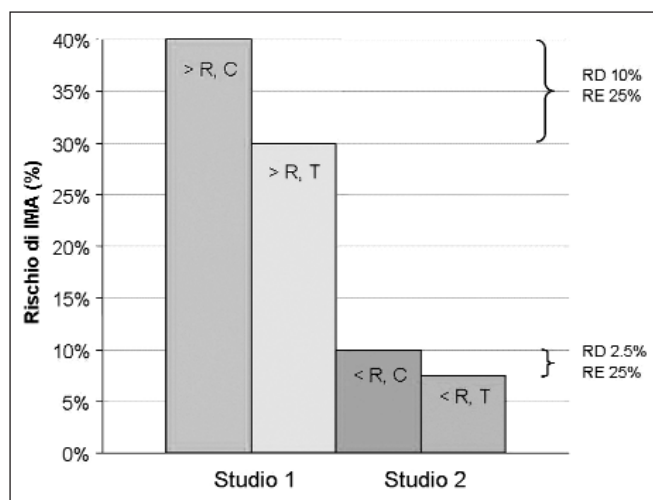


Fig. 5 - Studio 1 (colonne di sinistra): studio di un trattamento cardio-protettivo in soggetti con elevato rischio basale di IMA (40% a cinque anni). Studio 2 (colonne di destra): rischio basale basso (10% a cinque anni). Lo stesso trattamento (e lo stesso effetto relativo, del 25%) produce diverse differenze di rischio e diversi valori di NNT (10 nel primo caso e 40 nel secondo). C = controlli; T = trattati. R = rischio basale. RD differenza di rischio; RE effetto relativo.

densità di incidenza è il valore limite di RR con T, durante il quale i rischi si applicano, tendente a zero. Ecco perché il rapporto delle densità di incidenza è un RR istantaneo, anche definito azzardo relativo (hazard ratio, HR).

Rilevanza clinica

A questo punto il lettore potrebbe giustamente chiedersi: quale valore di RD, RR o RE dobbiamo considerare clinicamente importante? Supponiamo che uno studio clinico abbia dimostrato un RE di 0.3 (riduzione del 30% del rischio di eventi) tra gli esposti rispetto ai controlli (RR = 0.7). Se lo studio è statisticamente significativo, è necessariamente anche clinicamente rilevante? Un effetto relativo del 30% è sufficiente per autorizzarci a ritenere clinicamente importante il beneficio di un trattamento? Il primo studio che ha sollevato questo problema e di cui si è ampiamente discusso in letteratura risale al 1989. Grazie a questo studio, un gruppo di ricercatori Canadesi della McMaster University ad Hamilton (Ontario, CA) ha proposto una nuova misura di effetto, oggi ancora non sufficientemente utilizzata in letteratura, il number needed to treat (NNT). Questo studio randomizzato e controllato ha dimostrato una riduzione del 44% del rischio di infarto miocardico tra i soggetti esposti ad aspirina (325 mg a dì alterni) vs. placebo per 5 anni (3). Nonostante il RR di tutto rispetto (0.56) la differenza assoluta di rischio (RD) era bassa (0.009). Negli anni successivi è stato suggerito che uno dei problemi di RE e RR è che essi non riflettono il rischio basale (R_0) e, pertanto, non permettono di discriminare tra effetti

importanti e non. Al contrario RD chiaramente descrive la distanza tra R_1 e R_0 e, quindi, è più utile al clinico (4). Per esempio lo stesso RE del 44% può risultare da un trial in cui R_0 è 0.022 e R_1 è 0.012 (come nel trial dell'ASA) e da un trial ipotetico in cui i rischi sono 20 volte maggiori (R_0 è 0.44 e R_1 è 0.25). Nell'ultimo caso anche RD è 20 volte maggiore, così come l'impatto clinico del risultato. Purtroppo il significato di RD è difficile da tradurre nella pratica clinica quotidiana. Al contrario il suo reciproco ($1/RD$), oltre ad essere un numero intero (e non decimale, quindi più familiare al clinico), coincide col numero di pazienti che è necessario trattare per evitare un evento aggiuntivo per effetto del trattamento. In altre parole, il reciproco di RD ha la proprietà di informarci sul numero di pazienti che dobbiamo sottoporre al trattamento sperimentale per la durata del trial al fine di prevenire un evento aggiuntivo (number needed to treat to benefit, NNTB). Nell'esempio precedente dell'ASA, NNTB per 5 anni per prevenire un infarto per effetto dell'aspirina corrisponde a 100, mentre una RD 20 volte maggiore si traduce in un NNTB di 5. Inoltre, nel trial dell'ASA il 3.3% tra i soggetti trattati e il 2.2% dei non trattati ha presentato sanguinamento gastrointestinale durante il trial (NNT to harm, NNTH di 90). Il calcolo di NNTB e NNTH sono diventati un metodo interessante per stabilire in termini di rischio-beneficio i vantaggi di un intervento sanitario, anche se altre misure come la qualità della vita, i costi e le perdite di produttività andrebbero considerate in aggiunta. Secondo la relazione $RD = [R_0 * RE]$ e, quindi, $NNT = 1/[R_0 * RE]$, minore RE e /o il rischio basale (R_0) e maggiore NNT. Per cui se la riduzione relativa del rischio (RE) e i rischi di effetti indesiderati del trattamento sono gli stessi nei pazienti a basso ed alto rischio di eventi, allora, dovendo scegliere, l'intervento dovrebbe essere riservato a pazienti con maggior R_0 al fine di massimizzare i vantaggi (minimizzare NNTB) rispetto al rischio di potenziali effetti indesiderati (NNTH). In Figura 5 si vede come a parità di effetto relativo, se il rischio basale è maggiore allora la differenza dei rischi è maggiore, e NNT è minore.

Questa misura di rilevanza clinica che Laupacis e colleghi hanno definito NNT (4) ha le seguenti proprietà. Innanzitutto offre un riferimento scientifico per prendere la decisione di trattare alcuni pazienti e non altri pur affetti dallo stesso disturbo ma con diverso rischio basale per l'evento da prevenire. Poi NNT offre una base per esprimere i costi del trattamento e il suo monitoraggio in termini biologici ed economici (benefici ed effetti indesiderati). Infine NNT è un utile strumento per il confronto di diversi trattamenti per diversi disordini, cosa che potrebbe avere un interesse in termini di programmazione sanitaria (5, 6). Tuttavia, dobbiamo ricordare che NNT, come i rischi, deve essere espresso insieme al periodo di tempo del trattamento (durata). Per confrontare diversi trattamenti di diversa durata sulla base di NNT dovremmo accettare l'assunto che RE rimanga costante nel tempo. Se

così fosse, si potrebbe calcolare NNT per una durata ipotetica come il prodotto tra NNT osservato e [tempo osservato/tempo ipotetico]. Per esempio, volendo esprimere NNT per un anno da uno studio che dimostra un NNT di 5 per 2 anni, otteniamo un NNT di 10 per un anno. Ma le cose non sono così semplici.

Test di verifica

1) Una misura di effetto è:

- Una misura di tendenza centrale, quindi un parametro
- Una misura che aiuta a quantificare la modificazione di un risultato dovuta al trattamento
- Solo la differenza di rischio e non l'effetto relativo
- L'effetto relativo e non il rapporto dei rischi
- Soltanto il rischio relativo

2) La differenza di rischio permette di apprezzare la rilevanza clinica di un effetto perchè:

- Riflette il rischio basale
- Ha un range che va da -infinito a +infinito
- Si può esprimere in termini di NNT
- Le differenze sono più informative dal punto di vista clinico rispetto ai rapporti
- Tutte le precedenti

3) Il NNT:

- È il reciproco della differenza di rischio
- Come il rischio va espresso insieme ad un intervallo di tempo (durata del trattamento)
- È un numero intero che rappresenta il numero di pazienti che è necessario trattare sia per ottenere benefici (NNTB) che effetti indesiderati (NNTI) dalla terapia in questione
- Come le altre misure di effetto può variare nel tempo
- Tutte le precedenti

La risposta corretta alle domande sarà disponibile sul sito internet www.sin-italy.org/gin e in questo numero del giornale cartaceo dopo il Notiziario SIN

Problemi con NNT

Così come per rischi e tassi di incidenza, anche NNT può variare nel periodo di esposizione. Qualsiasi variazione in termini di effetto o di rischio basale comporta modificazioni di NNT nel tempo. Di tutto ciò bisogna tener conto quando si fanno confronti tra effetti di trattamenti diversi, specie se stimati in trials di durata diversa. Se infatti l'efficacia aumenta nel tempo un trial di breve durata tende a sovrastimare il vero NNT, mentre se essa tende a ridursi allora la stima di NNT sarà troppo ottimistica.

Un altro problema di NNT è mostrato in Figura 6. In questo studio ipotetico di durata inferiore ad un anno,

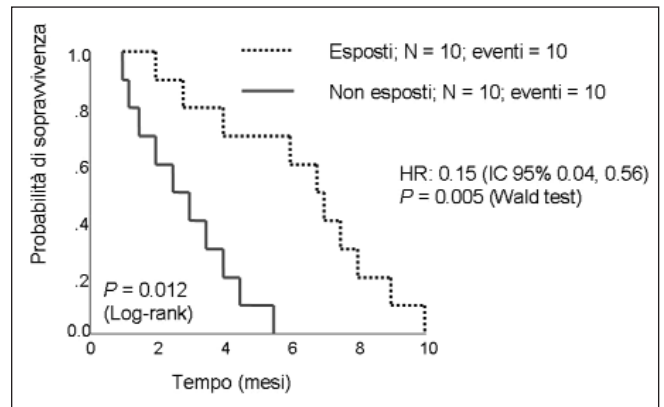
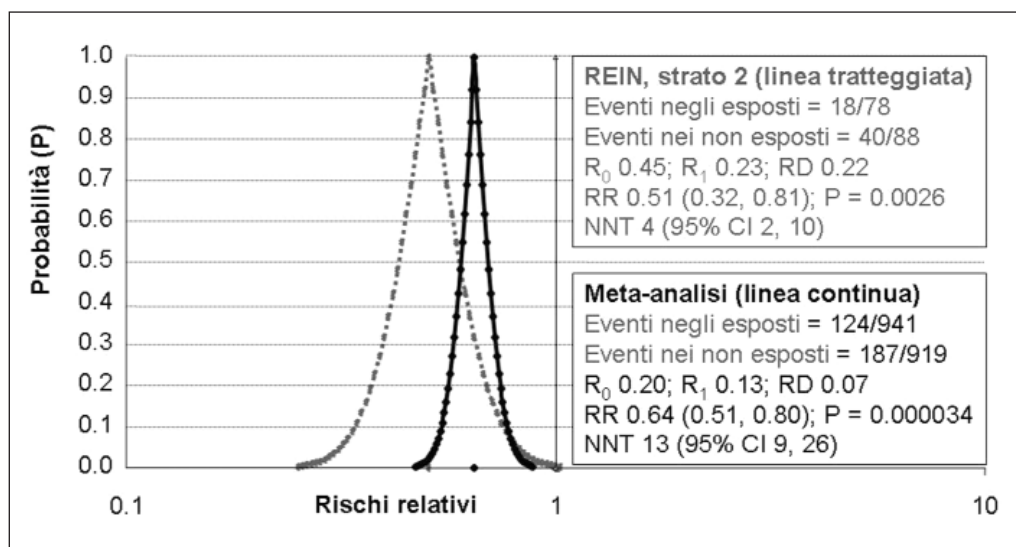


Fig. 6 - Esempio di curva di sopravvivenza nel caso di una malattia rapidamente fatale. HR azzardo relativo stimato con la regressione di Cox; IC intervallo di confidenza.

viene confrontato l'effetto di un trattamento vs. placebo. Non ci sono stati persi al follow-up, il rischio basale è molto alto e tutti i 20 pazienti hanno sperimentato l'evento. In questo caso la RD al termine dello studio non può essere calcolata come abbiamo imparato, nonostante le curve di sopravvivenza siano ben distinte e i tests statistici (log-rank sull'uguaglianza delle probabilità di sopravvivenza e Wald test di azzardi uguali) siano significativi (l'ipotesi nulla è rigettata). L'HR è 0.15 e supporta una forte associazione tra esposizione e outcome, nonostante la stima sia abbastanza imprecisa (pochi pazienti e ampio IC al 95%). È possibile calcolare NNT in questi casi? Sì, ma con alcuni limiti. Possiamo individuare alcuni momenti temporali di interesse e calcolare un NNT specifico per quel momento come $1/[(1-RR)*R_0]$ se $RR < 1$; come $1/[(RR-1)*R_0]$ se $RR > 1$. Nel nostro esempio a fine studio $NNT = 2$ (IC al 95% da 1 a 3, arrotondando per eccesso). Oppure possiamo calcolare NNT prima dell'ultimo evento nel gruppo di controllo, utilizzando le probabilità di sopravvivenza in quel momento. Nell'appendice sono riportati alcuni calcoli utili allo scopo e recentemente proposti da Doug Altman (7).

Per ottenere una stima di NNT è necessario disporre della stima delle probabilità di sopravvivenza in ciascun gruppo a tempi fissi (ricordando che i rischi di evento sono pari a $1 - P$ di sopravvivenza), oppure della stima di HR insieme alla stima della P di sopravvivenza nel gruppo di controllo per derivare il rischio basale. Quando è riportato il numero di pazienti a rischio a tempi fissi (pazienti sotto osservazione ancora senza evento) oppure gli errori standard delle stime, è possibile calcolare anche l'IC al 95% di NNT. Sfortunatamente i risultati di molti lavori non riportano le informazioni necessarie per il calcolo.

Fig. 7 - Funzioni P-IC dal REIN study e dalla metanalisi di Jafar. L'effetto (forza dell'associazione tra esposizione e malattia) è maggiore nel REIN study (rischio basale R_0 maggiore) ma la stima è meno precisa (minor numero di pazienti).



Esempi

Applichiamo i concetti precedentemente illustrati ad un frequente quesito clinico che il nefrologo si pone nel trattamento delle nefropatie croniche progressive: qual è la rilevanza clinica del trattamento con ACE inibitori nelle nefropatie dei soggetti non diabetici? Che impatto ha il trattamento con bloccanti del recettore dell'angiotensina II sulla progressione delle nefropatie nei soggetti con diabete di tipo II? In altre parole, dato il risultato positivo riportato (significatività statistica), quale rilevanza clinica ci dobbiamo aspettare dall'utilizzo di questi farmaci?

Consideriamo l'effetto del ramipril nel II strato del REIN study (8) e la metanalisi di Jafar (9) che include il REIN study. In entrambi gli studi, i rischi dell'evento combinato "raddoppio della creatinina sierica" o uremia terminale sono confrontati nel gruppo assegnato a trattamento attivo e a placebo. Il follow-up medio è simile (26 mesi). Nella Figura 7 sono riportati i valori di RR e NNTB insieme alle funzioni P-IC (la funzione 'valore di P e intervalli di confidenza' verrà discussa in una successiva rassegna dedicata alla significatività statistica) sui valori di RR calcolati a fine studio. Bisogna notare che i soggetti arruolati nel REIN study avevano una proteinuria di 3 o più g al dì con simili valori pressori tra trattati e non (144/88). Al contrario i soggetti della metanalisi avevano una proteinuria basale di 1.8 [SD 2.3] g al dì, maggiore età (52 vs. 49) e minore prevalenza di malattie glomerulari (33 vs. 62%) e, inoltre, i pazienti in trattamento attivo avevano valori pressori significativamente inferiori ($P < 0.001$) durante il follow-up rispetto ai pazienti in placebo (139/85 vs. 144/87). Si può notare che i due rischi basali (R_0) sono diversi, mentre nel secondo studio RE è inferiore e NNT maggiore. Nonostante il peggior controllo pressorio nel gruppo placebo, l'effetto del trattamento aggiustato per i valori di PA è paragonabile a quello non aggiustato qui riportato (RR

0.64, IC 95% da 0.5 a 0.8). Dalle curve di sopravvivenza con i pazienti a rischio riportate nel REIN study possiamo inoltre calcolare (più correttamente) i valori di NNT a 12 e 24 mesi, con i loro IC al 95% (Fig. 8). Questi valori sono, rispettivamente 9 pazienti da trattare per un anno (da 4 NNTB a 53 NNTH) e 6 da trattare per due anni (da 3 NNTB a 48 NNTH). La definizione di "NNTH" del limite superiore dell'IC di NNTB è la conseguenza del valore negativo di RD (di cui NNT è il reciproco): quando la stima puntuale di RD non è significativa, il limite inferiore della stima intervallare è inferiore a 0 (numero negativo) e il suo reciproco va in direzione opposta a NNTB, quindi diventa NNTH invece di NNTB. Queste stime di NNT implicano una forte associazione tra trattamento e risultato (NNT è piccolo), anche se non significativa prima dei due anni di terapia (perché il numero di eventi totali è basso). A fine studio, il risultato diventa statisticamente significativo, anche se l'intervallo resta impreciso perché nel frattempo il numero di soggetti a rischio si è ridotto (alcuni pazienti hanno lasciato lo studio prima dell'evento di interesse). A fine studio possiamo calcolare gli IC in due modi: assumendo (ottimisticamente) che tutti i pazienti senza evento siano ancora a rischio a fine studio (IC da 3 a 11); ma non siamo certi che chi ha abbandonato lo studio anticipatamente sia realmente ancora a rischio; oppure (più ragionevolmente) assumendo che i pazienti a rischio a 24 mesi siano ancora a rischio a fine studio (IC da 5 a 77). Lo stesso calcolo si può fare dagli HR riportati dagli autori (NNTB 7, da 4 a 38). Pertanto, dopo il secondo anno di terapia, possiamo stimare che $7 \times (26.4/12) = 16$ sono i pazienti che dobbiamo trattare per un anno al fine di evitare un evento aggiuntivo per effetto del trattamento. In figura 8 sono riportate anche le RD calcolate dalla metanalisi di Jafar e colleghi (9) coi rispettivi NNTB. In questo lavoro non sono riportati i pazienti a rischio per l'evento combinato ma solo per l'evento uremia. Il calcolo delle RD basa-

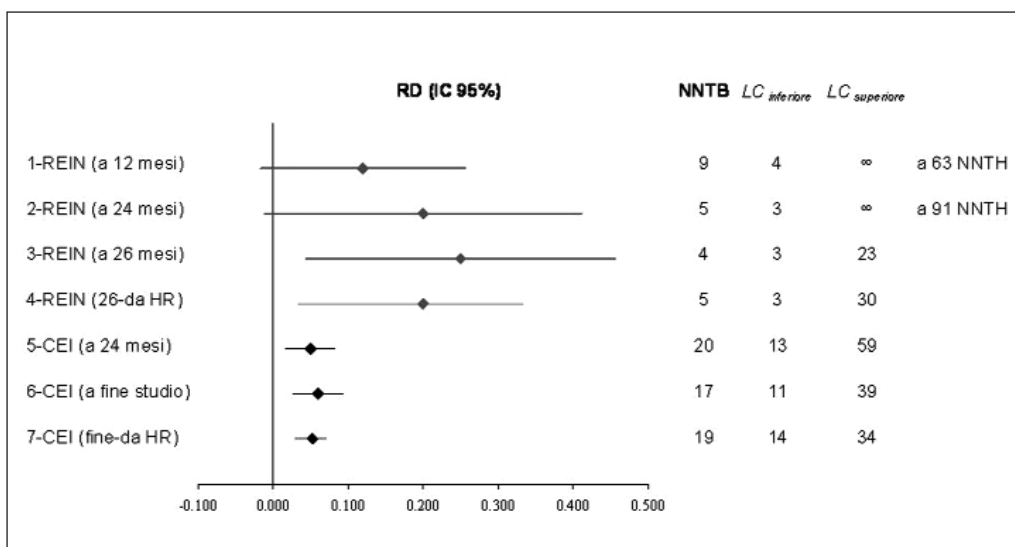


Fig. 8 - Differenze di rischio (RD) e NNT a 12 mesi, 24 mesi e fine studio calcolate con diverse modalità (vedi testo ed appendice).

to su questi valori di pazienti a rischio produce valori di NNT (per la durata dello studio) per l'evento combinato pari a 17 (da 10 a 27). Assumendo che tutti i pazienti precedentemente censurati siano ancora a rischio a fine studio, NNT diventa 14 (da 10 a 27) e, calcolando da HR, diventa 17 (da 10 a 31). Da notare che ogni stima nella metanalisi è più precisa (maggiore numerosità campionaria) rispetto al REIN study. Inoltre, la forza dell'associazione tra esposizione e malattia è più debole rispetto al REIN study, per il minor rischio basale dei pazienti inclusi nella metanalisi. Questi NNTB indicano che la progressione delle nefropatie croniche ha una prognosi modificabile con l'intervento con ACE-I ma non così drasticamente. I risultati implicano che trattando per oltre 2 anni 100 soggetti con rischio basale simile a quello dei soggetti della metanalisi dobbiamo aspettarci che solo da 3 a 10 pazienti eviteranno l'evento combinato per effetto del trattamento. Il vantaggio clinico appare un poco migliore se i nostri pazienti hanno un maggiore rischio basale, come quello del REIN study (i pazienti che eviterebbero l'evento per effetto dell'ACE-I sarebbero da 2 a 25 su 100). Naturalmente resta molto difficile prevedere il destino della funzione renale successivamente alla durata del trattamento (dopo 26 mesi) se gli stessi pazienti continuassero ad essere trattati: la differenza di rischio potrebbe rimanere la stessa, ridursi od aumentare (e altrettanto NNTB).

Simili NNTB (e anche maggiori) sono calcolabili dalle curve di sopravvivenza, dai pazienti a rischio e dagli HR riportati dai trials randomizzati sull'effetto di losartan e irbesartan rispetto al placebo nella riduzione dei rischi di raddoppio della creatinina, uremia terminale (vedi definizioni!!!) o morte nei pazienti con nefropatie croniche progressive e diabete tipo II (10, 11). In questi studi, il NNTB dopo 2 anni di trattamento è 26 (da 14 a 241) con losartan e 25 (da 15 a 415) con irbesartan. Dopo 3 anni i valori diventano rispettivamente 19 (da 10 a 178) e 17 (da 9 a 252).

Pertanto, non c'è dubbio circa l'esistenza di un effetto (statisticamente significativo) di ACE-I e inibitori del recettore angiotensinico nel ridurre il rischio di progressione delle nefropatie croniche. Purtroppo la rilevanza clinica di questo effetto è moderata, anche se essi rappresentano ad oggi il trattamento anti-ipertensivo di prima scelta nei nefropatici e il presidio nefro- e cardio-protettivo più efficace oggi disponibile (12). D'altra parte, visto che NNT è il reciproco di RD, un basso valore di NNT (ad esempio inferiore a 4, corrispondente a un RD di almeno 0.25) può risultare da un alto rischio basale e/o da un eccezionale effetto relativo, entrambi difficilmente compatibili con il decorso delle malattie cronico-degenerative, come le nefropatie croniche.

Infine va ricordato un concetto epidemiologico fondamentale. NNT dovrebbe essere calcolato sulla base dei rischi e non dalle densità di incidenza. Come per i rischi anche la stima accurata di NNT richiede un follow-up completo della coorte studiata. I soggetti che abbandonano lo studio anticipatamente senza evento potrebbero presentare l'evento dopo l'uscita dallo studio. Allo stesso modo non sapremo nulla di coloro "ancora senza evento" a fine studio. Queste osservazioni incomplete richiedono estrema prudenza nel calcolo di misure come NNT per stimare la rilevanza clinica dell'effetto di un intervento. Ecco perché uno studio con pochi eventi e, soprattutto con molti pazienti persi al follow-up è poco affidabile.

Test di verifica

1) La rilevanza clinica di un trattamento:

- a. Non può prescindere dalla significatività statistica
- b. È facilmente quantificabile dal rischio relativo
- c. Dipende dalla modificazione di una misura di malattia conseguente all'esposizione ed è influenzata dal

- rischio basale (rischio nei controlli)
- d. Non va considerata se il risultato non è statisticamente significativo
 - e. tutte le precedenti
- 2) Nelle malattie cronico-degenerative:**
- a. È più facile misurare il rischio vero
 - b. È più facile disporre di trattamenti con efficacia eccezionale
 - c. È più facile completare studi di follow-up a lungo termine
 - d. Il rischio basale nei controlli è altissimo
 - e. Nessuna delle precedenti
- 3) Per pazienti a rischio si intende:**
- a. Il numero di persone che non hanno ancora presentato l'evento ma sono suscettibili di manifestarlo e sono sotto osservazione ad un certo punto temporale
 - b. Il numero di soggetti che hanno già presentato l'evento
 - c. I soggetti censurati
 - d. I pazienti persi al follow-up
 - e. Tutte le precedenti

La risposta corretta alle domande sarà disponibile sul sito internet www.sin-italy.org/gin e in questo numero del giornale cartaceo dopo il Notiziario SIN

Conclusione

Un risultato statisticamente significativo non implica un effetto clinicamente rilevante. L'importanza clinica di un trattamento è meglio quantificabile con la differenza di rischio indotta e meglio apprezzabile dal clinico mediante il suo reciproco, il NNT. In ogni caso va tenuto presente che ciascuna di queste stime tanto più tende ad allontanarsi dal valore vero quanto maggiore è la quota di pazienti persi durante lo studio. Stime delle differenze di rischio (e, quindi di NNT) potrebbero dimostrarsi non costanti nel tempo e sono estremamente approssimate quando proiettate in follow-up superiori alla durata degli studi che le hanno prodotte.

Riassunto

Lo studio critico delle evidenze prodotte in letteratura comprende la valutazione dell'importanza clinica dei risultati degli studi clinici. La rilevanza dell'effetto di un trattamento dovrebbe coincidere col minimo effetto clinicamente importante atteso dai ricercatori nella pianificazione del lavoro. Negli studi di hard outcomes la rilevanza dell'effetto di un'esposizione può essere quantificata come differenza assoluta dei rischi di malattia (tra esposti e non esposti) e il suo reciproco, il numero di pazienti che è necessario

trattare per evitare un evento (NNTB). Anche il numero di pazienti da trattare per produrre effetti indesiderati della terapia (NNTH) va considerato insieme al NNTB. Infine, va ricordato che queste misure di effetto sono ottenute da misure di malattia che stimano i rischi veri e sono pertanto stime grezze dei veri parametri della popolazione. Questo succede non solo perché il lavoro è condotto su campioni ma anche perché il follow-up dei pazienti è incompleto. Gli assunti e gli aspetti metodologici del disegno dello studio sono particolarmente importanti per valutare la rilevanza clinica di ogni risultato ottenuto.

Appendice

1) NNT da due curve di sopravvivenza (S_1 e S_0) e dal numero di pazienti a rischio (n_1 and n_0), ricordando che la probabilità di non evento (S) è complemento a 1 della probabilità di evento ($1 - S$). Per prima cosa si calcola l'errore standard della RD (SE di RD):

$$SE_{RD} = \sqrt{[S_1^2 * (1 - S_1)/n_1 + S_0^2 * (1 - S_0)/n_0]}$$

Poi l'IC al 95% di RD $[(1 - S_0) - (1 - S_1)]$ come:

$$IC\ 95\% = RD \pm 1.96 * SE_{RD}$$

Quindi NNT è il reciproco di RD.

2) NNT dal rischio basale e da HR (IC 95%, HR_L e HR_U):
 $NNT = 1/[S_0^{HR} - S_0]$, da $1/[S_0^{HR_U} - S_0]$ a $1/[S_0^{HR_L} - S_0]$

Se non riportati esplicitamente i valori di HR possono essere ottenuti dai coefficienti di regressione b (e da SE_b). Infatti $h = e^b$, da $e^{b - 1.96 * SE_b}$ a $e^{b + 1.96 * SE_b}$

3) Se SE di S_1 e S_0 sono disponibili, allora SE_{RD} è $\sqrt{[SE_{S_0}^2 + SE_{S_1}^2]}$

Se gli IC 95% di S_1 e S_0 sono riportati invece dei loro SE, allora ciascun SE è approssimativamente un quarto dell'ampiezza di IC.

Ringraziamenti

Il Dr. Ravani è vincitore di una delle borse che la Società Italiana di Nefrologia ha messo a disposizione per la formazione e la ricerca clinica per l'anno 2005. Entrambi gli Autori sono grati al Dr. Francesco Casino per i preziosi suggerimenti espressi durante la revisione del lavoro.

Indirizzo degli Autori:

Dr. Pietro Ravani

Divisione di Nefrologia e Dialisi

Azienda Istituti Ospitalieri di Cremona

Largo Priori, 1

26100 Cremona

e-mail: p.ravani@ospedale.cremona.it

Bibliografia

1. Rothman KJ, ed. Measuring disease occurrence and causal effects. In: *Epidemiology: an introduction*. New York: Oxford University Press 2002; pp. 24-56.
2. Tripepi G, Catalano F. Kaplan-Meier analysis. *G Ital Nefrol* 2004; 21: 540-6.
3. Steering Committee of the Physicians' Health Study Research Group. Final report of the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* 1989; 321: 129.
4. Sackett D, Haynes BR, Guyatt GH, Tugwell P, eds. Deciding for the best therapy. *Clinical Epidemiology, a basic science for clinical medicine*. Lippincott Williams & Wilkins, 1991; pp. 187-248.
5. <http://www.jr2.ox.ac.uk/bandolier/band50/b50-8.html>
6. <http://www.cebm.utoronto.ca/glossary/nnts.htm#table>
7. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999; 319: 1492-5.
8. The GISEN Group (Gruppo Italiano di Studi Epidemiologici in Nefrologia): Randomised placebo-controlled trial of effect of ramipril on decline in glomerular filtration rate and risk of terminal renal failure in proteinuric, nondiabetic nephropathy. *Lancet* 1997; 349: 1857-63.
9. Jafar TH, Schmid CH, Landa M, et al. Angiotensin-converting enzyme inhibitors and progression of nondiabetic renal disease. A meta-analysis of patient-level data. *Ann Intern Med* 2001; 135: 73-87.
10. Brenner BM, Cooper ME, De Zeeuw D, et al. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. *N Engl J Med* 2001; 345: 861-9.
11. Lewis E, Hunsicker LG, Clarke WL, et al. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. *N Engl J Med* 2001; 345: 851-60.
12. Rossert JA, Wauters JP. Recommendations for the screening and management of patients with chronic kidney disease. *Nephrol Dial Transplant* 2002; 17 (Suppl 1): S19-28.