

La rilevanza clinica dei trial



A. Cartabellotta

nino.cartabellotta@gimbe.org

Presidente Fondazione GIMBE

Parole chiave: Randomized controlled trial, outcomes, number needed to treat, confidence intervals

Key words: Trial controllato randomizzato, misure di esito, numero necessario (di pazienti) da trattare, limiti di confidenza

Il Giornale di AMD, 2012;15:101-104

Sommario

L'outcome (end-point) primario di un trial clinico e gli eventuali outcome secondari influenzano la rilevanza clinica dello studio. In quale misura i trial che utilizzano end-pointsurrogati hanno una limitata rilevanza clinica? Esiste una metodologia standardizzata per validare un end-point surrogato? Come definire in maniera oggettiva e riproducibile i criteri di misurazione degli outcome? Il follow-up è sufficientemente lungo per rilevare gli outcome definiti? Qual è il minimum data set sulle modalità di definizione e misurazione degli outcome che devono fornire i ricercatori? Oltre che dagli end-point misurati, la rilevanza clinica di un RCT è influenzata dai numeri che stimano quantitativamente benefici e rischi degli interventi sanitari. La pillola riporta definizioni, formule, vantaggi e limiti delle misure di efficacia/sicurezza più affidabili: il number needed to treat (NNT) ed il number needed to harm (NNH). Indispensabile anche valutare i limiti di confidenza per stimare la precisione di NNT/NNH.

Abstract

The primary and the secondary end-points affect the relevance of clinical trials. Do trials with surrogate end-point have a limited clinical relevance? Are there methodologies to validate the surrogate end-points? How to measure end-points in objective way? Does follow-up last enough to capture the established outcomes? What minimum data set should the researchers establish to define and measure outcomes? More than outcomes, the clinical relevance of trials is affected by quantitative estimates of intervention's benefits and harms. This article reports definitions, formulae, benefits and limits of the most reliable measures of efficacy and safety: the number-needed-to-treat (NNT) and the number-needed-to-harm (NNH), more than the confidence intervals.

La rilevanza clinica di un trial (RCT) è influenzata da quattro variabili: l'end-point misurato, il numero di pazienti da trattare per ottenere un beneficio (NNT), il numero di pazienti da trattare per osservare un effetto avverso (NNH) e la precisione dei risultati di NNT/NNH espressa dai limiti di confidenza. Questo articolo è dedicato alla selezione degli outcome.

1. Definizione Outcome, end-point.

Nella metodologia della ricerca clinica possono essere considerati sinonimi: entrambi hanno il significato di "misura di esito".

2. Classificazione

Epstein e Sherwood (AnnInternMed, 1996) classificano gli outcome in tre categorie.

Clinici. Mortalità (totale o causa specifica), eventi morbosi (infarto, stroke, frattura), misure fisiologico-metaboliche (colesterolo LDL, pressione arteriosa, densità minerale ossea).

Economici. Diretti (ospedalizzazioni, visite ambulatoriali, test diagnostici, farmaci), indiretti (giornate lavorative perse), intangibili (grado di sofferenza, stress psicofisico).

Umanistici. Qualità di vita, soddisfazione dell'utente, sintomi, status funzionale.

Considerato che qualunque intervento sanitario (preventivo, diagnostico, terapeutico, riabilitativo, palliativo) sperimentato in un RCT modifica in varia misura outcome appartenenti alle tre categorie, idealmente un trial dovrebbe misurare almeno un outcome per ciascuna categoria. In realtà, vengono sempre selezionati gli outcome clinici, molto raramente quelli economici e - solo recentemente in alcuni settori (oncologia, malattie croniche) - quelli umanistici.

3. Outcome primario e outcome secondari

In un RCT i ricercatori devono definire l'outcome primario e, eventualmente, uno o più outcome secondari. È bene precisare che l'aggettivo "primario" non caratterizza la rilevanza clinica dell'outcome, ma riveste solo un significato statistico: infatti, l'outcome primario è quello utilizzato per stimare la dimensione del campione. Pertanto, anche se può apparire paradossale, esistono trial con outcome primario surrogato e outcome secondari clinicamente rilevanti perché non è obbligatorio definire un end-point rilevante, nemmeno a fini regolatori.

4. Outcome clinicamente rilevanti e surrogati

Gli end-point clinici vengono storicamente dicotomizzati in:

Outcome clinicamente rilevanti. Comprendono, oltre la mortalità, tutti gli eventi clinici morbosi che possono essere rilevati nel corso della storia naturale/post-terapeutica di una malattia.

Outcome surrogati. Sono variabili anatomico-fisiologico-metaboliche misurate con test di laboratorio / strumentali. Vengono utilizzati con la presunzione di dimostra-

re l'efficacia del trattamento su un end-point clinicamente rilevante senza misurarlo direttamente! Tale presupposto, teoricamente ineccepibile, è stato ripetutamente smentito: infatti, numerosi trattamenti efficaci su end-point surrogati (antiaritmici nel post-infarto, farmaci inotropi nello scompenso, terapia sostitutiva ormonale, fluoruri nell'osteoporosi, etc.), sono stati abbandonati perché adeguate evidenze hanno dimostrato che, misurando l'end-point clinicamente rilevante, i pazienti non trattati avevano risultati migliori. Purtroppo, nell'ultimo ventennio, la rilevanza clinica della ricerca si è progressivamente ridotta anche per l'uso smodato degli end-point surrogati che - dalla prospettiva di chi finanzia la ricerca - presentano indiscutibili vantaggi: permettendo, infatti, di ridurre durata del follow-up e dimensione del campione, diminuiscono notevolmente i costi del trial. Senza entrare nel merito della spinosa problematica, le autorità regolatorie non hanno mai vincolato la rilevanza clinica dell'end-point all'approvazione di nuovi farmaci, anzi l'FDA utilizza gli studi condotti su end-point surrogati per il processo di *accelerated approval*!

Il "calderone" degli end-point surrogati, per definizione, contiene anche outcome con una certa rilevanza clinica. Ad esempio, per valutare l'efficacia di un trattamento con antivirali nei pazienti con epatite cronica B, i livelli di transaminasi, la viremia e il miglioramento del quadro istologico sono end-point surrogati per definizione, ma la loro rilevanza clinica è differente: infatti, solo la risposta istologica è predittiva di una riduzione delle complicanze della cirrosi. Questi outcome surrogati, validati da studi che ne dimostrano la predittività su end-point rilevanti, vengono definiti **outcome intermedi**.

In definitiva, piuttosto che essere dicotomizzati in surrogati e clinicamente rilevanti, gli outcome dovrebbero essere stratificati attraverso una scala di rilevanza clinica. In questa direzione va anche il GRADE - il nuovo sistema per effettuare il grading degli studi clinici ai fini della produzione di raccomandazioni cliniche nelle linee guida - che definisce una gerarchia quantitativa degli outcome clinici, utilizzando uno score numerico:

- Score 1-3: outcome poco rilevanti per le decisioni cliniche e/o di limitata importanza per il paziente.
- Score 4-6: outcome importanti per il paziente, ma non critici per le decisioni cliniche.
- Score 7-9: outcome critici per le decisioni cliniche.

5. La validazione degli outcome surrogati

Esiste la possibilità di validare end-point surrogati ed utilizzarli come "veri sostituti" dei corrispondenti outcome clinicamente rilevanti? La risposta alle *Users' Guides to Biomedical Literature* - "testo sacro" del *critical appraisal* - che hanno definito per gli end-point surrogati tre livelli di validazione.

- Livello 1. Esiste un'associazione stretta, indipendente e consistente tra l'end-point surrogato e quello clinicamente rilevante? Questo primo livello, necessario ma non sufficiente, viene fornito da studi osservazionali che dimostrano l'inferenza tra i due outcome.
- Livello 2. Esistono evidenze sperimentali condotte con altre classi di farmaci, dove il miglioramento dell'end-point surrogato ha determinato un miglioramento di quello clinicamente rilevante?

- Livello 3. Esistono evidenze sperimentali condotte con farmaci della stessa classe, dove il miglioramento dell'end-point surrogato ha determinato un miglioramento di quello clinicamente rilevante?

Ad esempio, tutti gli anti-ipertensivi sono validati ai livelli 1 e 2, mentre solo per alcune classi di farmaci esistono evidenze di validazione al livello 3.

6. Le raccomandazioni del CONSORT Statement

L'item 7 del CONSORT Statement - linea guida di riferimento per il reporting dei trial clinici - raccomanda ai ricercatori di "definire chiaramente l'outcome primario e quelli secondari e, quando applicabile, descrivere i metodi utilizzati per garantire la qualità della loro misurazione". Ecco i dettagli:

- Precisa definizione e descrizione degli outcome che, per garantire una certa rilevanza clinica al trial, dovrebbero essere clinicamente rilevanti o almeno intermedi.
- Definizione di criteri standard per la misurazione degli outcome, che dovrebbe idealmente essere oggettiva.
- Quando esistono inferenze soggettive (segni, sintomi, interpretazione di imaging, istologia, etc.), è necessario prevedere strategie metodologiche per la loro minimizzazione: *blinding* (se possibile), valutazioni multiple, training standardizzato degli *outcome assessors*.
- Definizione della frequenza nella misurazione degli outcome e di eventuali "aggiustamenti".
- Reporting di tutti i metodi statistici, in particolare quando gli outcome primari sono multipli.

Oltre che dagli end-point misurati, la rilevanza clinica di un RCT è influenzata dai numeri che stimano quantitativamente benefici e rischi degli interventi sanitari: NNT, NNH e relativi limiti di confidenza.

NNT: Number Needed to Treat

Definizione. È il numero di pazienti da trattare per ottenere un beneficio terapeutico. Il NNT è realmente informativo solo se "accompagnato" da tutti i seguenti elementi: intervento sperimentale, intervento di controllo, outcome, durata del trattamento, limiti di confidenza.

Formola. Il NNT è il reciproco della riduzione del rischio assoluto [$1/ARR$], arrotondato per eccesso al numero intero.

Interpretazione. Al diminuire del NNT aumenta l'efficacia del trattamento (figura), per cui 1 è il NNT ideale: un successo terapeutico per ciascun paziente trattato.

Vantaggi. Rispetto ad altre misure di efficacia utilizzate nei trial (RRR, RR, OR, ARR) - oggetto di approfondimento della prossima "pillola" - il NNT è molto semplice da comprendere e, misurando i benefici assoluti, permette facilmente di confrontare l'efficacia del trattamento su outcome differenti.

Limiti. Il calcolo del NNT, semplice per gli outcome dicotomici, è molto complesso per quelli continui; i NNT derivati dalle meta-analisi possono risultare particolarmente "insidiosi", specialmente quando il rischio basale non è omogeneo tra i diversi studi; i NNT degli outcome *event-based* (eventi ricorrenti come esacerbazioni di BPCO, attacchi di emicrania, etc.) determinano una sovrastima dell'efficacia terapeutica; i NNT derivati dai trial possono essere sensibilmente diversi da quelli ottenuti in contesti assistenziali reali; nonostante le raccomandazioni del CONSORT Statement, pochi trial riportano il NNT; diffidare, infine,

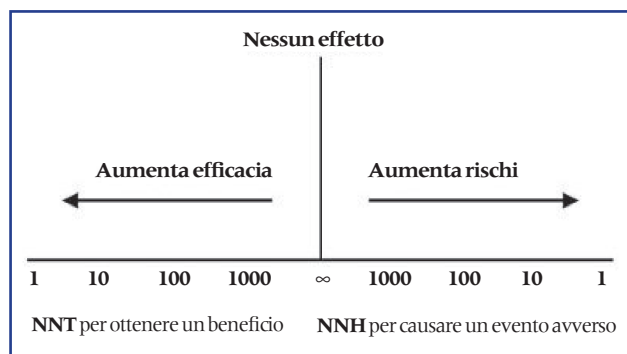
dal NNT/anno come “unità di misura” di efficacia per confrontare trattamenti diversi.

NNH: Number Needed to Harm

Definizione. È il numero di pazienti da trattare per osservare un effetto avverso del trattamento. Analogamente al NNT la sua interpretazione richiede: intervento sperimentale, intervento di controllo, outcome, durata del trattamento, limiti di confidenza.

Formola. Il NNH è il reciproco dell’incremento del rischio assoluto [$1/ARI$], arrotondato per difetto al numero intero.

Interpretazione. All’aumentare del NNH si riduce la probabilità di eventi avversi e aumenta la sicurezza del trattamento (figura), per cui il NNH ideale tende all’infinito, documentando l’assenza di eventi avversi.



Vantaggi. Analoghi al NNT.

Limiti. Il NNH può essere calcolato solo quando la frequenza dell’evento avverso viene riportata dai trial che - oltre a minimizzare frequenza e gravità degli eventi avversi - spesso non rilevano differenze significative tra i due gruppi di tali eventi, specialmente se poco frequenti; inoltre i NNH derivati dai trial sono generalmente più elevati (terapie più sicure) rispetto ai contesti assistenziali reali.

Limiti (intervallo) di confidenza

I risultati di un RCT vengono definiti “stima puntiforme”, perchè ottenuti da un campione di pazienti e non dall’intera popolazione. I limiti (o intervallo) di confidenza al 95% - *confidence intervals* (CIs) - delimitano il range di valori che contiene il valore reale per l’intera popolazione, con una probabilità del 95%. In altre parole, ripetendo lo studio 100 volte su campioni di popolazione differenti, i risultati (95/100 volte) sarebbero contenuti all’interno del CI, la cui ampiezza stima la precisione/incertezza dei risultati dello studio per le inferenze relative al campione studiato. Il CI è utile anche quando la differenza tra due trattamenti non è statisticamente significativa: infatti alcuni RCTs non raggiungono il fatidico $p < 0.05$ per una scarsa numerosità del campione studiato. In questi studi - falsamente negativi - il limite inferiore dell’IC del NNT può suggerire la potenziale rilevanza clinica di RCTs statisticamente non significativi. Analogamente, il limite superiore dell’IC del NNT può sminuire la rilevanza clinica di RCT statisticamente significativi. In conclusione, NNT e NNH esprimono la stima puntiforme dei risultati - favorevoli e sfavorevoli - del trial mentre i corrispondenti IC forniscono la precisione di tali stime.

Case study: il PEP trial

Nei pazienti sottoposti a intervento chirurgico per frattura di femore o ad artroplastica in elezione, l’aspirina (ASA) a basse dosi - 160 mg/die per 5 settimane - rispetto al placebo, riduce l’incidenza di eventi tromboembolici (TE)?

A questa *research question* risponde il PEP trial¹: “nei pazienti trattati con aspirina si riduce del 43% il rischio relativo di embolia polmonare (EP) e del 29% quello di trombosi venosa profonda (TVP). Un evento TE è confermato in 105/6679 pazienti trattati con aspirina e in 165/6677 pazienti assegnati al gruppo placebo, che corrisponde a una riduzione cumulativa del 36% di eventi tromboembolici (TE)”.

Proviamo a riportare i risultati con il NNT.

	Evento sì	Evento no
Trattati	105	6574
Controlli	165	6512
CER = $165/6677 = 0.025$		
EER = $105/6679 = 0.016$		
ARR = CER - EER = 0.009		
NNT = $1/ARR = 111$ (73-237)		

In altri termini per evitare un evento TE bisogna trattare con ASA 111 pazienti con ampi limiti di confidenza (73-237). Tuttavia, considerato che la terapia con ASA è poco costosa e semplice da somministrare e che la durata del trattamento è solo di 5 settimane, la sua rilevanza clinica rimane comunque elevata. Viene, tuttavia, ridimensionata valutando il NNH relativo a effetti avversi maggiori (ematemesi e melena): trattando 101 pazienti con ASA si osserva un grave effetto collaterale, seppur con ampi limiti di confidenza (66-222).

	Evento sì	Evento no
Trattati	205	6474
Controlli	139	6538
CER = $139/6677 = 0.021$		
EER = $205/6679 = 0.031$		
ARI = EER - CER = 0.009		
NNH = $1/ARI = 101$ (66-222)		

Abbreviazioni

CER= *Control Event Rate*: incidenza dell’evento nel gruppo di controllo
 EER= *Experimental Event Rate*: incidenza dell’evento nel gruppo sperimentale
 ARR= *Absolute Risk Reduction*: riduzione del rischio assoluto
 ARI= *Absolute Risk Increase*: aumento del rischio assoluto

¹ Pulmonary Embolism Prevention (PEP) Trial Collaborative Group. Prevention of pulmonary embolism and deep vein thrombosis with low dose aspirin: Pulmonary Embolism Prevention (PEP) trial. *Lancet* 2000;355: 1295-302

Articolo riprodotto con permesso da: GIMBEnews 2009;2:33-4 e GIMBEnews 2009;2:39-40

BIBLIOGRAFIA

Letture introduttive

1. Gebiski V, Marschner I, Keech AC. Specifying objectives and outcomes for clinical trials. *Med J Aust* 2002;176:491-2
2. Bucher HC, Guyatt GH, Cook DJ, et al. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. *JAMA* 1999;282:771-8.
3. Epstein RS, Sherwood LM. From outcomes research to disease management: a guide for the perplexed. *Ann Intern Med* 1996;124:832-7.
4. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996;125:605-13.
5. Moore A. What is an NNT. *Hayward Medical Communications*, 2009.
6. McAlister FA. The "number needed to treat" turns 20 — and continues to be used and misused. *CMAJ* 2008;179:549-553.
7. Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998;317:1309-12.
8. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995;310:452-4.
9. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988;318:1728-33.

Approfondimenti

1. Garrido MV, Mangiapane S. Surrogate outcomes in health technology assessment: an international comparison. *Int J Technol Assess Health Care* 2009; 25: 315-22

2. Guyatt GH, Oxman AD, Kunz R, et al. GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? *BMJ* 2008; 336: 995-8
3. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Aff (Millwood)* 2005; 24: 67-78
4. Psaty BM, Weiss NS, Furberg CD, et al. Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease. *JAMA* 1999; 282: 786-90
5. Stang A, Poole C, Bender R. Common problems related to the use of number needed to treat. *J Clin Epidemiol* 2010; 63: 820-5
6. Aaron SD, Fergusson DA. Exaggeration of treatment benefits using the "event-based" number needed to treat. *CMAJ* 2008;179:669-71
7. Massel D. The number needed to harm: is it too optimistic? *Can J Cardiol* 2003;19:1490-2
8. Nuovo J, Melnikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. *JAMA* 2002;287:2813-4
9. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses — sometimes informative, usually misleading. *BMJ* 1999; 318:1548-51

Software e Database

1. EBM calculator. Centre for Evidence-based Medicine, Oxford
2. Stats Calculator. Centre for Evidence-based Medicine. University of Toronto
3. NNT: table of examples. Centre for Evidence-based Medicine. University of Toronto

