Valutazione Critica delle Evidenze

1. Concetti generali 2
2. Validità interna 3
   - Bias di selezione 3
   - Performance Bias (conduzione dello studio) 3
   - Attrition Bias 3
   - Detection Bias 4
   - Reporting Bias 4
   - Imprecision Bias 4
3. La validità esterna dello studio (Directness) 4
   - I pazienti 5
   - I trattamenti 5
   - Gli outcomes 5
     - Un cenno sugli outcomes intermedi e surrogati 5
4. La rilevanza clinica (Relevance) 7
   - Rilevanza clinica negli Studi di superiorità 7
   - Rilevanza clinica negli studi di “non inferiorità” 8
   - Rilevanza clinica negli studi di Qualità della vita percepita dal paziente (PRO) ... 9
5. Consistenza 10
   - Tra le caratteristiche più importanti c’è la eterogeneità. 10
   - Pubblication Bias 11
6. Referenze 12

1. Concetti generali

I criteri per l'approccio critico variano in relazione al tipo di studio/quesito (epidemiologico, diagnotico, terapeutico, prognostico), ma seguono la stessa traccia comune.

- Lo studio è valido?
- I risultati dello studio sono clinicamente rilevanti?
- I risultati dello studio sono applicabili?

L'oggetto della valutazione critica (critical appraisal) è la qualità della ricerca clinica, una caratteristica multidimensionale costituita da 4 dimensioni:

1. **Validità interna (Confidence)**

   La **validità interna** misura quanto i risultati di uno studio siano corretti per il campione di individui che sono stati studiati. Essa viene detta «interna» perché si applica al campione, e non necessariamente agli altri individui della popolazione.
   Negli studi clinici, la validità interna dipende, quindi, dalla correttezza di impostazione dello studio stesso, dalla scelta di buone tecniche diagnostiche e da un loro corretto utilizzo, da una buona elaborazione dei dati ecc.
   E' strettamente legata al rigore metodologico seguito dai ricercatori nelle fasi di pianificazione, conduzione ed analisi dei risultati con l'obiettivo di minimizzare i bias (errori sistematici) che possono compromettere la stessa validità interna della ricerca.
   I requisiti metodologici che influenzano la validità interna variano nei differenti tipi di studi e, di conseguenza, condizionano i criteri per valutare criticamente i diversi articoli.

2. **Validità esterna (Directness)**
   La validità esterna è il grado di «generalizzabilità» delle conclusioni tratte da uno studio.
   Ad esempio, nel caso di uno studio epidemiologico clinico, essa risponde alla domanda "Supponendo che i risultati dello studio siano veri, essi si applicano anche ai miei pazienti?". In altre parole, la validità esterna misura il grado di verità dell'assunto secondo cui i pazienti studiati sono "uguali" ad altri pazienti affetti dalla stessa condizione.
   Definita anche applicabilità, è il grado con cui i risultati della ricerca possono essere applicati al paziente individuale.
   E' influenzata principalmente da:
   - criteri di selezione dei pazienti
   - setting assistenziale: organizzativo, tecnologico, professionale
   - motivazioni: staff, pazienti

3. **Rilevanza clinica (Relevance)**
   Il background deve spiegare bene le ragioni scientifiche che giustificano lo studio; non è etico sottoporre pazienti al rischio delle sperimentazioni che non sono giustificate dal punto di vista scientifico (Helsinki decl.)
   Fornisce una stima dell'entità e della precisione del beneficio ottenuto.
   Nel critical appraisal corrisponde alla **fase di analisi dei risultati**: ad esempio, negli studi terapeutici la rilevanza clinica dei risultati è condizionata da
   - end-point misurati (surrogati vs clinicamente rilevanti)
   - numero necessario di pazienti da trattare (NNT)
   - ampiezza limiti di confidenza (Imprecision)

4. **Consistenza**
   Definita anche riproducibilità della ricerca, si riferisce al fatto che i risultati di uno studio siano confermati da altri studi
   La consistenza, a differenza delle altre caratteristiche, non può essere valutata nel singolo studio: infatti, lo strumento ideale è costituito dalle revisioni sistematiche con meta-analisi di più studi.

Dunque, la validità esterna esprime il grado di generalizzabilità dei risultati ad altre popolazioni, mentre la validità interna esprime i provvedimenti messi in atto per ridurre il rischio di bias (errori sistematici) durante la pianificazione, esecuzione ed analisi dei dati di uno studio.
La validità interna è prerequisito della validità esterna.

In sintesi: la valutazione critica della letteratura (critical appraisal) è, quindi, un processo sequenziale che ha l'obiettivo di valutare metodologia, rilevanza ed applicabilità (validità esterna) della ricerca pubblicata.
La valutazione della consistenza non costituisce, invece, parte integrante del critical appraisal degli studi primari.

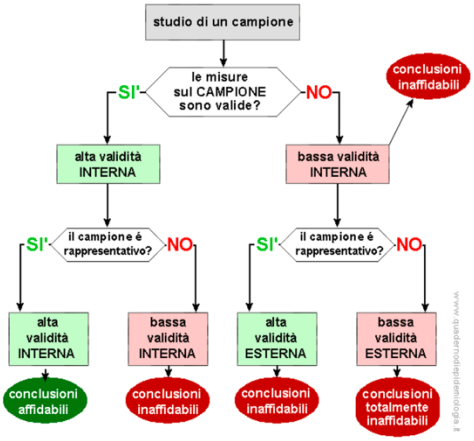

Figura 1 Validità interna e validità esterna nel Criticl appraisal

2. Validità interna

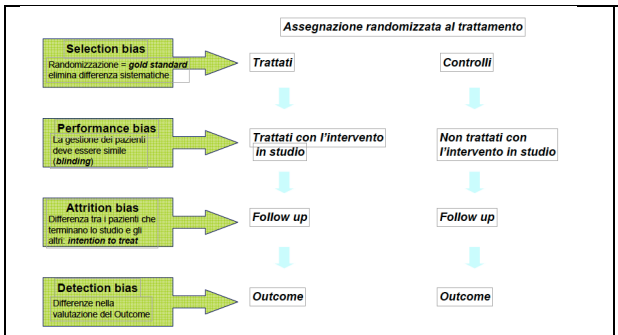

Figura 2 Validità interna e bias

Type of bias	Description	
Selection bias.	Systematic differences between baseline characteristics of the groups that are compared.	• Sequence generation. • Allocation concealment.
Performance bias.	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest.	• Blinding of participants and personnel. • Other potential threats to validity.
Detection bias.	Systematic differences between groups in how outcomes are determined.	• Blinding of outcome assessment. • Other potential threats to validity.
Attrition bias.	Systematic differences between groups in withdrawals from a study.	• Incomplete outcome data
Reporting bias.	Systematic differences between reported and unreported findings.	• Selective outcome reporting

Cochrane Handbook for Systematic Reviews of Interventions, version 5.1.0

Bias di selezione

Il bias di selezione è quello conseguente a un processo di assegnazione non casuale (non randomizzato) del trattamento in uno studio clinico:

- Generazione della sequenza di randomizzazione
- Mascheramento dell'assegnazione (allocation concealment)

Performance Bias (conduzione dello studio)

Il Performance Bias è legato alla possibile influenza che la conoscenza del trattamento assegnato (assenza di mascheramento) da parte del paziente e/o del medico può avere sulla rilevazione dell'effetto del trattamento…

Attrition Bias

L'Attrition Bias riguarda la possibile distorsione legata alla perdita (omogenea o disomogenea) di pazienti alla valutazione dell'outcome di interesse clinico.

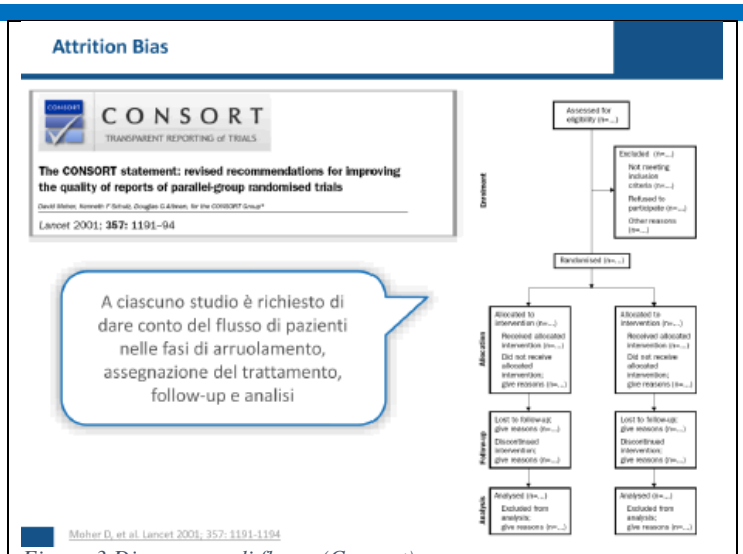

Figura 3 Diagramma di flusso (Consort)

- Un'attenta lettura del diagramma di flusso dei pazienti in uno studio può essere di valido aiuto per quantificare un'eventuale attrition (Moher, Schulz, e Altman 2001)
- La differenza tra i due gruppi non dovrebbe essere mai disomogenea (di maggiore entità in uno dei due bracci a confronto) e comunque per convenzione mai superiore al 10%(Vale, Tierney, e Burdett 2013).

Detection Bias

Il Detection Bias è legato alla possibile influenza che la conoscenza del trattamento assegnato (assenza di mascheramento) da ***parte del valutatore*** può avere sul giudizio di efficacia (o meno) del trattamento: ad esempio, *la conoscenza del trattamento in atto da parte del radiologo potrebbe influenzare la sua valutazione riguardo alla presenza o meno di progressione radiologica della malattia*.

Reporting Bias

Il Reporting Bias riguarda la comunicazione selettiva (più confacente allo scopo dello sperimentatore) dei risultati, tralasciando (o riportando in maniera meno evidente) quelli che potrebbero intralciare le conclusioni dello studio.
Ad. Esempio è il caso in cui le considerazioni conclusive possano venir centrate più su outcome accessori (per il quale lo studio non era stato pianificato), mentre può non essere presa in considerazione o addirittura non venir fatta menzione l'outcome primario risultato negativo e deludente per i ricercatori.

Imprecision Bias

L'affidabilità dei risultati di uno studio è legata anche alla precisione (il valore e l'intervallo di confidenza entro cui probabilisticamente questo valore rappresenta la realtà che stiamo studiando) con cui i risultati vengono riportati.
In presenza di un numero insufficiente di casi arruolati e/o di eventi e/o di intervalli di confidenza di eccessiva ampiezza si parla di *Imprecision*.
Poiché la quantificazione della *Imprecision* è legata anche alla valutazione della rilevanza clinica dei risultati (*Relevance*), si rimanda alla Rilevanza.

3. La validità esterna dello studio (Directness)

Considera la diretta trasferibilità, o meno, delle evidenze in esame al paziente o alla popolazione cui il ricercatore fa riferimento.

I fattori in gioco sono:

- le caratteristiche del Paziente (P)
- il tipo di Intervento in esame (I)
- l'alternativa terapeutica considerata (C)
- i parametri clinico-laboratoristici ritenuti essenziali per formulare la proposta terapeutica (O)

Non esiste una validità esterna assoluta;
Sta al giudizio dell'esperto valutare in quale misura

- i pazienti,
- i trattamenti,
- gli outcome

assomigliano alle condizioni della pratica clinica

I pazienti

- Selezione dei pazienti, setting, area geografica;
- Criteri di eleggibilità/esclusione: (anziani, donne, percentuale dei pazienti screenati ma non eleggibili, percentuale dei pazienti eleggibili ma non arruolati e cause);

1. *Erano rappresentativi della popolazione di pazienti che si incontrano nella pratica clinica routinaria ?*
2. *Il setting era simile a quello in cui lavoro io?*
3. *I criteri di inclusione erano troppo stretti o troppo larghi?*
4. *Quanti pazienti eleggibili non sono poi stati randomizzati?*
5. *E' stato tenuto un registro dei pazienti con la stessa malattia che non sono stati inclusi, le motivazioni della non inclusione e il loro esito?*

I trattamenti

(modalità di erogazione, durata, i trattamenti concomitanti, limitazioni nei cotrattamenti, la durata del follow up)

- Erano appropriati e riproducibili rispetto alla pratica clinica anche in altre realtà?
- Sono stati messi in atto gli opportuni trattamenti concomitanti?

Gli outcomes

- (outcome surrogati, outcome clinicamente rilevanti, scale per la misurazione degli outcomes, outcome patient centered, outcome combinati, lunghezza del follow up, completezza degli effetti indesiderati);

1. *Le misure di outcome erano adeguate per raggiungere le conclusioni riguardo ai trattamenti in studio ?*
2. *E' stata data troppa importanza agli endpoint surrogati?* •
3. *La durata del trattamento era adeguata per valutare l'efficacia e la tossicità del trattamento ?* •
4. *La durata del follow-up era adeguata per valutare l'efficacia e la tollerabilità del trattamento ?* •
5. *Sono state prese misure adeguate per la raccolta e al presentazione dei dati di tossicità ?*

Un cenno sugli outcomes intermedi e surrogati

Per definire un *outcome* intermedio come surrogato dell'*outcome* finale, l'outcome finale dovrebbe essere adeguatamente predetto dall'effetto che il trattamento dimostra sull'*outcome* intermedio…(Sargent 2011)

Cioè occorrerebbe che fosse innanzitutto dimostrata una correlazione tra l'*outcome* intermedio e l'*outcome* finale…
(se un singolo paziente ha un controllo della progressione di malattia, quello stesso paziente vivrà ragionevolmente più a lungo, correlazione *at patient level*).
Questo non dovrebbe essere correlato solo per gli outcome di pazienti in un determinato contesto, ma per essere realmente surrogato deve ***"predire la magnitudine dell'effetto del trattamento sull'endpoint "vero" in base all'effetto del trattamento sull'endpoint (candidato) surrogato. "***

Il surrogato è tale se la predizione è sufficientemente precisa. Cioè la dimostrazione di surrogacy viene effettuata attraverso una validazione che necessita dell'esistenza di più studi randomizzati attraverso un modello di regressione.
In parole semplice ma forse troppo approssimative: non basta che l'effetto di una terapia nel contesto di uno studio mi predica che il paziente vada meglio in un braccio che in un altro, ma occorre che questa verità sia tale anche in altri contesti e che la proporzionalità dell'effetto sia costante. Solo così si potrà dire che si tratta di una surrogacy.

Per fare un esempio banale, normalmente l'outcome intermedio di regressione obiettiva della lesione tumorale (cioè la risposta ovvero la sopravvivenza libera da progressione o da ripresa del PSA etc.. - di solito studi di fase II) è predittivo dell'outcome finale "sopravvivenza"; ma questo outcome intermedio non è sempre predittivo dell'outcome sopravvivenza e/o dello stessa entità nei contesti di differenti studi. La regressione di una lesione all'aumentare dell'efficacia di una terapia potrebbe contemporaneamente incrementare le morti per tossicità per cui il miglioramento sull'outcome sopravvivenza potrebbe essere ridotto perdendo quel valore predittivo di proporzionalità. Ovvero l'incremento della lesione (outcome intermedio) potrebbe non più essere espressione di peggioramento dell'effetto negativo sull'outcome finale (sopravvienza), perché dovuto ad una maggiore infiltrazione infiammatoria (si vede immunoterapia) difensiva e non una crescita tumorale.

Di qui la necessità di correlazione con modello regressivo in diversi studi (c.d. correlazione at trial level)

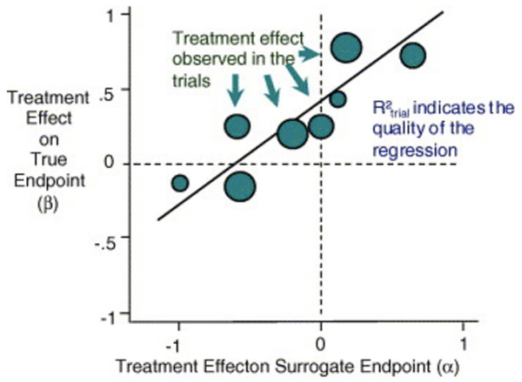

Figura 4

Nella Figura 4 viene verificata (*at trial level*) la correlazione tra effetto della radioterapia sull'*outcome* intermedio "*event-free survival*" ed effetto del trattamento sulla "*overall survival*". (Collette, Burzykowski, e Schröder 2006)

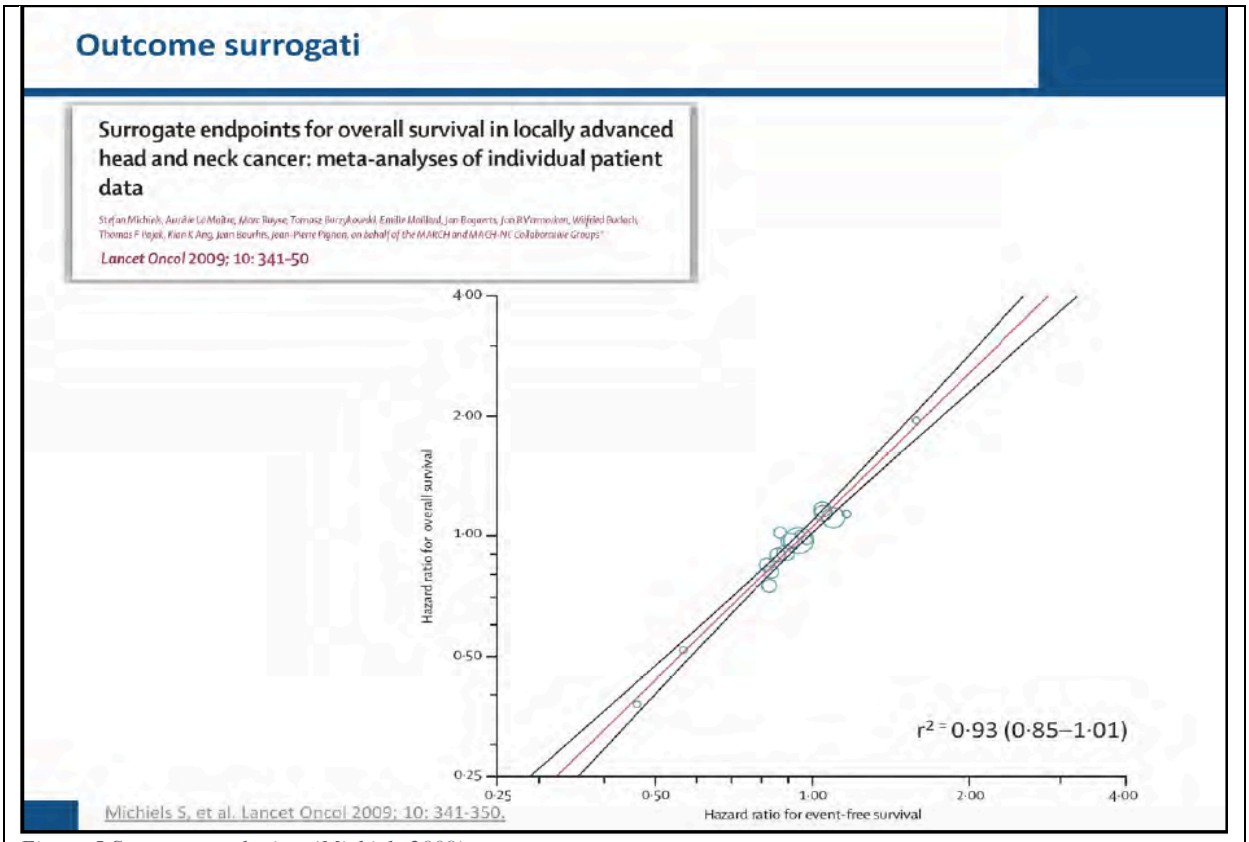

Figura 5 Surrogate endpoints (Michiels 2009)

Come si evince dal grafico e dall'indice di correlazione r2 del trial di Michelis(Michiels et al. 2009) (Figura 5 Surrogate endpoints (Michiels 2009)), che correla l'endpoint finale sopravvivenza con la sopravvivenza libera da eventi, si può parlare di surrogacy, in quanto il 93% dell'effetto del trattamento sulla sopravvivenza è spiegato dall'effetto dello stesso trattamento sulla prevenzione degli eventi sfavorevoli. (*Presa dal corso del dott. Pappagallo (2017) per i radio-oncologi Corso ECM FAD "Metodologia degli studi clinici di terapia combinata nelle neoplasie testa-collo" http://fad.accmed.org Accademia Nazionale di Medicina).*

La distanza delle curve in azzurro rappresentano l'intervallo di confidenza nei vari contesti e più è piccolo, più preciso della verità è il valore trovato, più affidabile è la surrogacy.
Riassumendo,

- il raggiungimento di un *outcome* intermedio (per es. *event-free survival*) è spesso associato a una specifica evoluzione della malattia. Rappresenta quindi ***un fattore prognostico di esito.***
- ***Un surrogato*** è invece un sostituto per un *endpoint* di ***riconosciuta valenza clinica nella pianificazione di uno studio clinico***.

Per dirla con le parole del dott. Pappagallo:

- il fattore prognostico è da considerarsi utile per un singolo paziente (nella pratica clinica) – (patient level)
- l'endpoint surrogato è da considerarsi utile (risparmio di tempo e risorse) nella pianificazione di uno studio clinico (trial level)

4. La rilevanza clinica (Relevance)

Nel disegnare uno studio viene calcolato la numerosità dei campioni da confrontare così da avere un beneficio minimo giudicato di rilevanza clinica tale da giustificare lo studio e la modifica del nostro atteggiamento terapeutico.
Si possono ipotizzare da un punto di vista di rilevanza clinica tre situazioni:

- Il caso di studi di superiorità
- Il caso di studi di "non inferiorità"
- Il caso di studi che misurano la qualità percepita dei pazienti attraverso questionari.

In tutti i casi occorre definire quanto il risultato sia rilevante da un punto di vista di utilità clinica.
(Ocana e Tannock 2011) Ocana e Tannock puntualizzano "…We would define a positive trial as one in which the predefined value of Δ represents a clinically important difference in an endpoint that directly reflects benefit (mainly OS or quality of life) to patients and for which the results provide a best estimate of the difference that exceeds that predefined value of Δ….. " ed inoltre "… "The high cost of new medications adds weight to the requirement that studies should be designed to detect only important clinical benefits. A positive study should be based on a meaningful increase in survival or in quality of life for the patient, or both, but not on a statistically significant *P* value…."

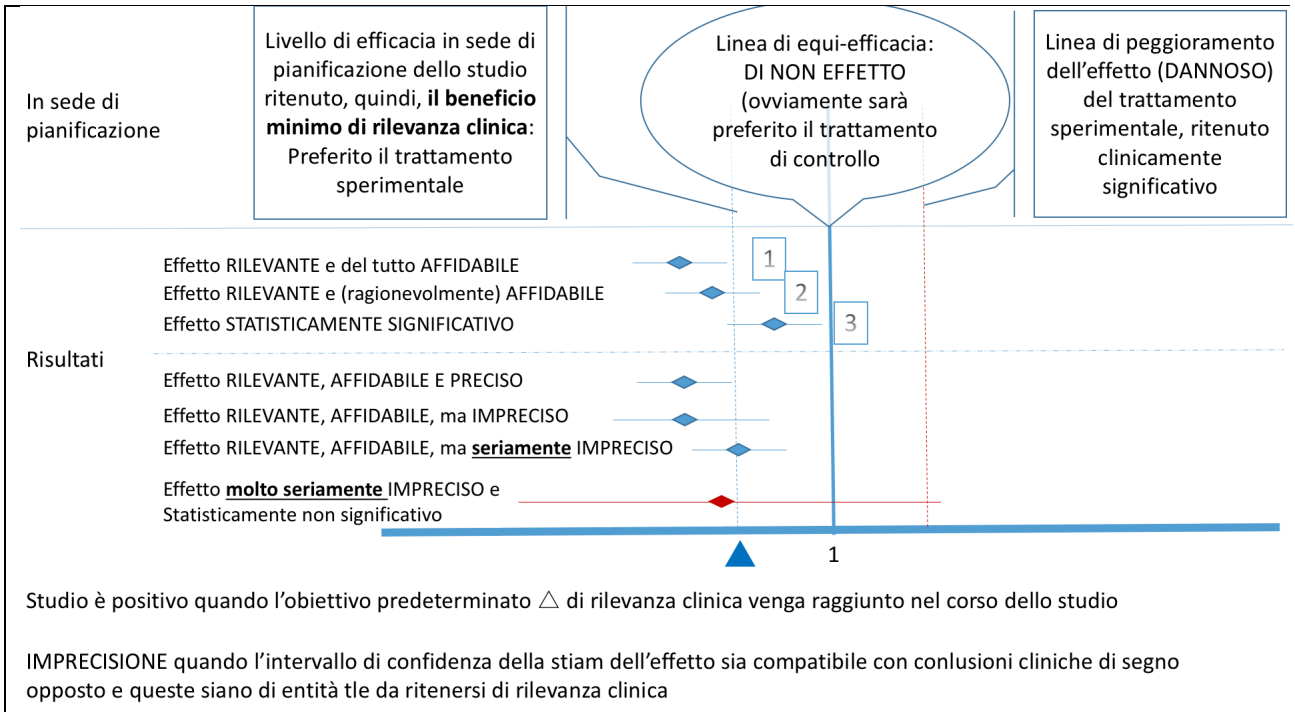

Figura 6

Rilevanza clinica negli Studi di superiorità

Gli studi di superiorità rispondono al quesito: ***Si ritiene che il trattamento sperimentale abbia le potenzialità per migliorare il trattamento standard almeno di una quantità Δ?***

Intuitivamente analizzando la Figura 6 si può osservare che fissati in fase di pianificazione il beneficio minimo (ed eventualmente l'effetto dannoso clinicamente rilevante negli studi di non inferiorità) si possono configurare alcune situazioni nell'analisi dei risultati: Negli esempi numerati "1,2 e 3" esiste in maniera statisticamente significativa una differenza tra il trattamento sperimentale e il controllo (l'estremità destra dell'intervallo di confidenza non interseca la linea di non-effetto continua, a indicare che la differenza osservata nel campione non è verosimilmente -con probabilità di risultato falsamente positivo <5%) dovuta al caso.
Tuttavia da questo tipo di interpretazione non si ricava alcuna informazione sulla rilevanza clinica dell'effetto (linea tratteggia con in ascisse il simbolo Δ). Le considerazioni di utilità clinica che si possono fare nei tre casi sono le seguenti:

- Caso 1: sia la stima puntuale (quanto osservato nello studio) sia l'intervallo di confidenza (indicatore di variabilità della stima) risiedono a sinistra della soglia di rilevanza clinica: il risultato è quindi rilevante e del tutto affidabile. Tuttavia questa è una situazione di riscontro assai raro.
- Caso 2: la stima puntuale si colloca a sinistra della soglia di rilevanza clinica, mentre l'intervallo di confidenza la interseca: il risultato è quindi rilevante e affidabile in misura dipendente dal grado di sovrapposizione tra intervallo di confidenza e linea di rilevanza clinica. In ogni caso questa situazione dovrebbe rappresentare il requisito minimo per accettare un risultato come clinicamente interessante.
- Caso 3: la stima puntuale si colloca a destra della linea di rilevanza clinica, e (naturalmente) l'intervallo di confidenza la interseca: il risultato è quindi (semplicemente) statisticamente significativo. Ciò a indicare che in termini numerici la differenza osservata a favore del trattamento sperimentale non è ragionevolmente dovuta al caso, tuttavia la dimensione della stessa non è tale da venire considerata come clinicamente rilevante.

Per quanto riguarda i parametri di confidenza e in particolar modo alla imprecisione dei risultati (Imprecision) si parla di ***Imprecision quando l'intervallo di confidenza della stima di effetto sia compatibile con conclusioni cliniche di segno opposto, e queste siano di entità tale da ritenersi di rilevanza clinica***.
In effetti gli esempi sottostanti a quelli numerati riportano i diversi livelli di imprecisione sulla base dell'intervallo di confidenza e della soglia (o soglie) di rilevanza clinica prestabilita (si veda anche Figura 7).

Rilevanza clinica negli studi di "non inferiorità"

Gli studi di non-inferiorità vogliono dimostrare che il trattamento sperimentale non è clinicamente peggiore del trattamento attivo di controllo.
Gli studi di "non inferiorità" rispondo al quesito: ***Vista la migliore tollerabilità (o il minor costo o altro vantaggio) del trattamento sperimentale, si è disposti ad accettarne una eventuale minore efficacia rispetto al trattamento standard purché questa non vada oltre un margine M***

The JAMA Network

From: **Reporting of Noninferiority and Equivalence Randomized TrialsExtension of the CONSORT 2010 Statement**
JAMA. 2012;308(24):2594-2604. doi:10.1001/jama.2012.87802

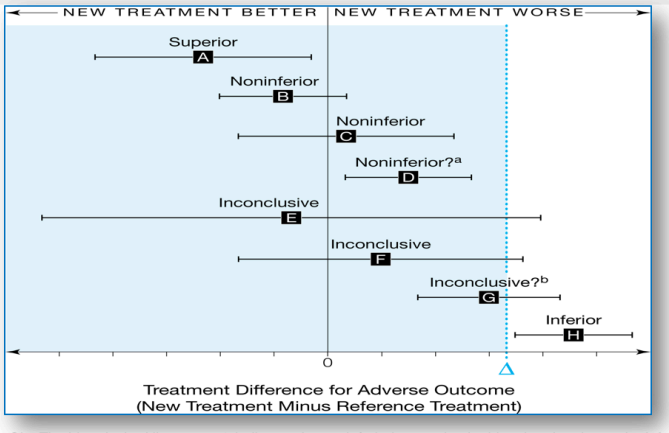

Figure Legend:

Error bars indicate 2-sided 95% CIs. The blue dashed line at x = Δ indicates the noninferiority margin; the blue tinted region to the left of x = Δ indicates the zone of inferiority. **A**, **If the CI lies wholly to the left of zero, the new treatment is superior**. **B and C**, **If the CI lies to the left of Δ and includes zero, the new treatment is noninferior but not shown to be superior**. ***D***, If the CI lies wholly to the left of Δ and wholly to the right of zero, the new treatment is noninferior in the sense already defined but also inferior in the sense that a null treatment difference is excluded. This puzzling circumstance is rare, because it requires a very large sample size. It also can result from a noninferiority margin that is too wide. ***E and F***, **If the CI includes Δ and zero, the difference is nonsignificant but the result regarding noninferiority is inconclusive.** ***G***, ***If the CI includes Δ and is wholly to the right of zero, the difference is statistically significant but the result is inconclusive regarding possible inferiority of magnitude Δ or worse.*** ***H***, ***If the CI is wholly above Δ, the new treatment is inferior.***

[a]This CI indicates noninferiority in the sense that it does not include Δ, but the new treatment is significantly worse than the standard. Such a result is unlikely because it would require a very large sample size. [b]This CI is inconclusive in that it is still plausible that the true treatment difference is less than Δ, but the new treatment is significantly worse than the standard. Adapted from Piaggio et al.

Figura 7

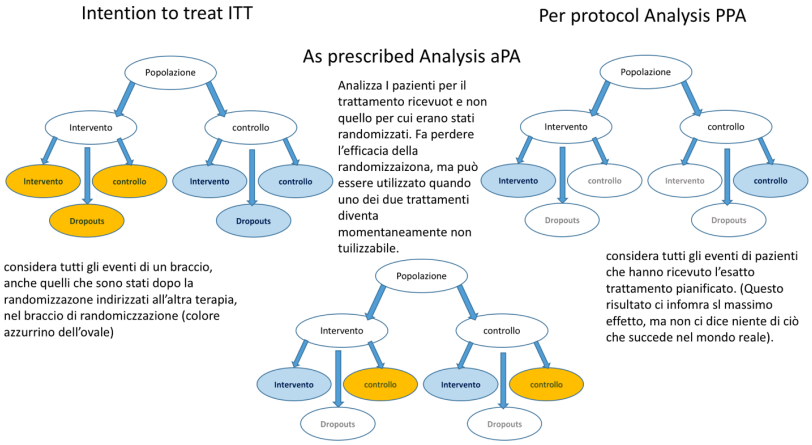

Figura 8

In questi studi l'attenzione va rivolta all'intervallo di confidenza, indipendentemente da dove si colloca la stima puntuale dell'effetto.

Quattro regole per gli studi di non inferiorità

1. Definizione e giustificazione del margine di non inferiorità o equivalenza
   – E' un compito del medico!
2. Descrizione del calcolo del campione
   – Deve tener conto del margine scelto
3. • Presentazione delle analisi sia ITT (intention to treat) che "per protocol"(vedi Figura 8)
   – Entrambi i risultati devono andare nella stessa direzione (come raccomandato da EMA e FDA); questo per verificare che un'eccesiva deviazione di trattamento prevista non vada a mascherare una possibile inferiorità del trattamento sperimentlae rispetto al trattamento di controllo.
4. • Presentazione dei IC del risultato
   – I IC devono essere più piccoli possibile, ma il giudizio finale deve sempre essere clinico

Rilevanza clinica negli studi di Qualità della vita percepita dal paziente (PRO)

Nel caso di punteggi ottenuti da questionari somministrati ai pazienti (per es. qualità di vita), il valore numerico è di per sé di difficile interpretazione clinica.

La MID (Minimal Interesting Difference) fornisce ai clinici un chiaro riferimento per l'interpretazione di un effetto rilevato sull'outcome in termini di successo/insuccesso, e quindi permette di definire un paziente come "responder".

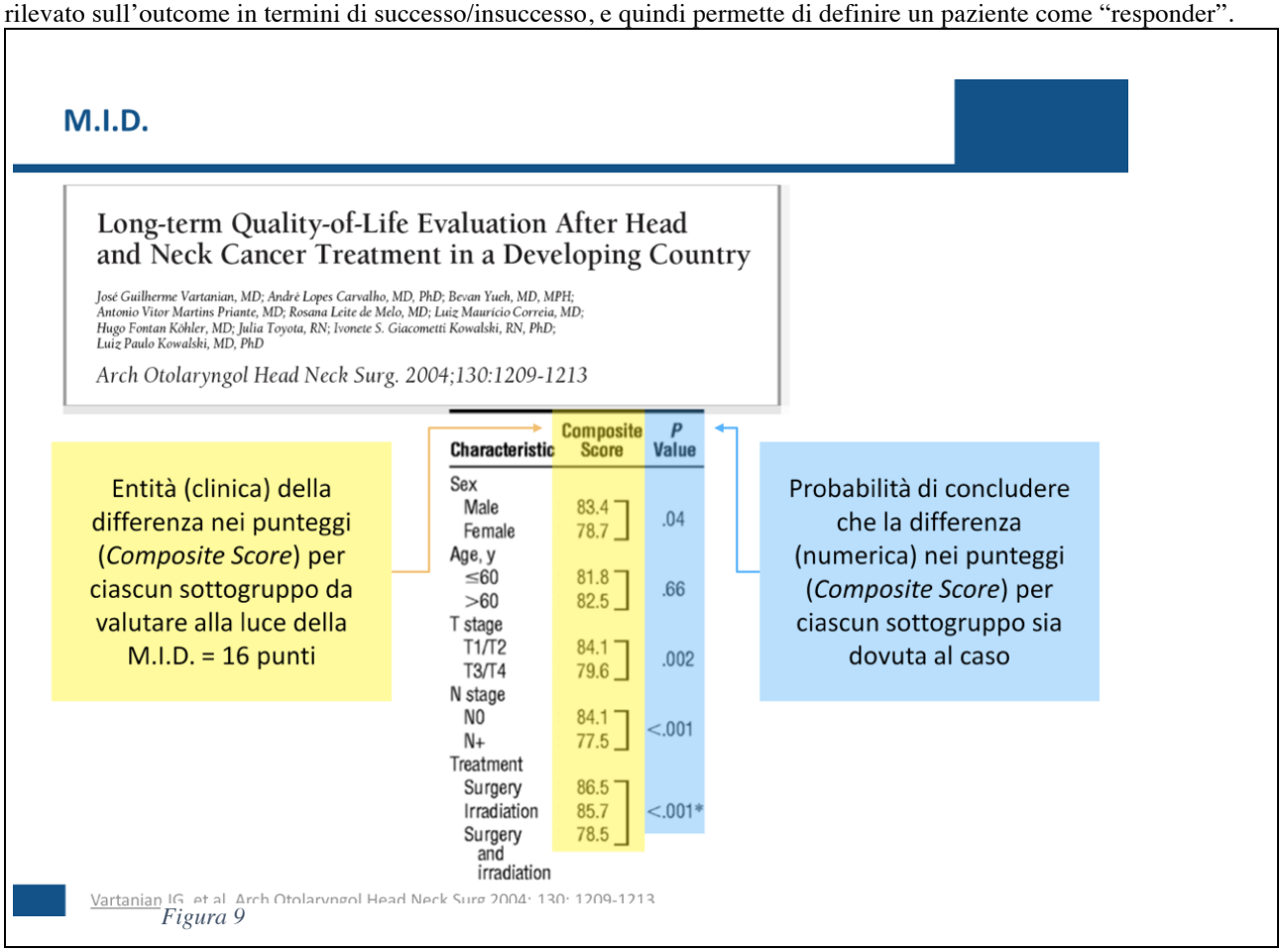

Figura 9

Nel caso descritto nella figura a lato tutte le differenze nel Composite Score del questionario UW-QoL-R risultate statisticamente significative, sono tuttavia da considerarsi come non rilevanti, in quanto l'entità delle differenze osservate è sempre inferiore alla M.I.D. =16 punti.

5. Consistenza

Definita anche riproducibilità della ricerca, si riferisce al fatto che i risultati di uno studio siano confermati da altri studi La consistenza, a differenza delle altre caratteristiche, non può essere valutata nel singolo studio: infatti, lo strumento ideale è costituito dalle revisioni sistematiche con meta-analisi di più studi.

Tra le caratteristiche più importanti c'è la eterogeneità.

L'Eterogeneità è costituita da ogni sorta di variabilità tra gli studi di una revisione sistematica e precisamente la eterogeneità delle stime di effetto tra gli studi che non trova spiegazione logica con la sola casualità.: ad es, diversità di intervento o della composizione del campione in studio.

1. Una prima valutazione può essere fatta sulla forest plot guardando (Figura 11):

– la variazione della dimensione dell'effetto per studio (Le stime puntuali variano ampiamente attraverso gli studi.
– Gli intervalli di confidenza si accavallano poco o nulla.

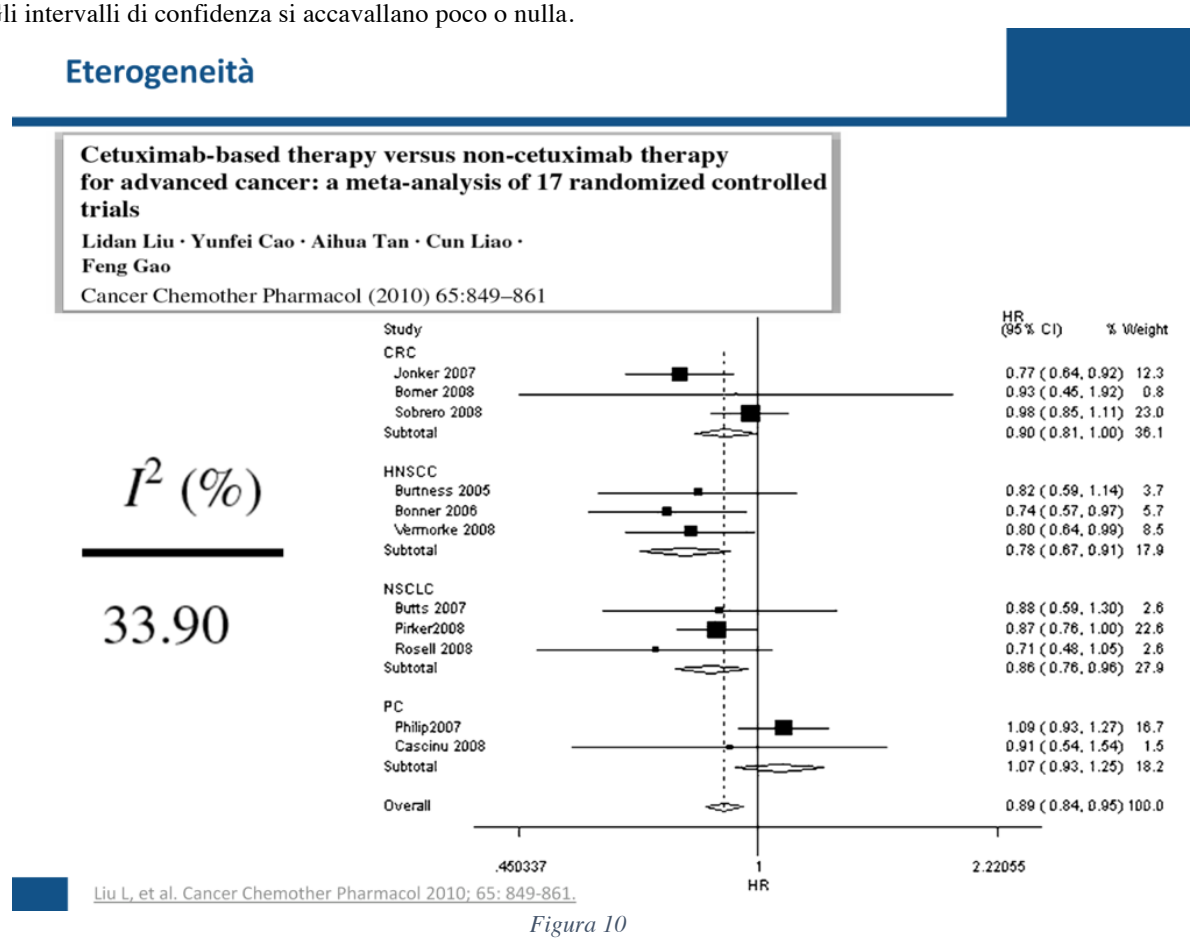

Figura 10

2. Utilizzando Cochrane's Q: per valutare se le differenze osservate nei risultati sono compatibili con la distribuzione χ^2 chi quadro; p=<10 indica eterogenità
3. Utilizzando **l'indice di eterogenità I^2** (vedi Figura 11) è considerarsi indicatore di sostanziale eterogeneità quando maggiore del 50% (quando cioè più del 50% della variabilità dei risultati dei singoli studi non appare ragionevolmente dovuta al caso).

Eterogeneità

What is Heterogeneity?

- **Any kind of variability among studies in a systematic review may be termed heterogeneity.**
- **I-squared (I^2)** $I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$, where Q is the chi-squared statistic and df is its degrees of freedom
  - describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance).
  - thresholds for the interpretation of I^2:
    - 0% to 40%: might not be important;
    - 30% to 60%: may represent moderate heterogeneity;
    - **50% to 90%: may represent substantial heterogeneity;**
    - **75% to 100%: considerable heterogeneity.**

Figura 11

Pubblication Bias

Le revisioni precoci sono quelle che sono affette maggiormente da questo tipo di bias perché i lavori con esito negativo tardano ad essere pubblicati, mentre quelli positivi sono di dimensioni piccoli e spesso sponsorizzate dalle dite farmaceutica.

Nel funnel plot si riportano sull'asse orizzontale le misure degli effetti di trattamento

Sull'asse verticale gli errori standard (SE) degli effetti stimati dell'intervento (misure della precisione della stima: SE è determinata dalla dimensione del campione, dal numero dei partecipanti che vanno incontro all'evento negli studi con outcome dicotomici, mentre dalla deviazione standard delle riposte negli studi con eventi a variabile continua. Mentre la PRECISIONE della stima aumenta con le dimensioni del campione e saranno più in alto, gli studi piccoli e imprecisi si distribuiranno alla base.

Se la distribuzione asseconda la casualità il funnel plot dovrebbe essere simmetrico, altrimenti sarà asimmetrico con aree vuote.(vedi figura)

Il funnel plot può aiutarci a desumere l'esistenza di lavori ad esito negativo non pubblicati a patto che la distribuzione casuale possa essere valutata e cioè quando ci siano almeno 10 lavori pubblicati.

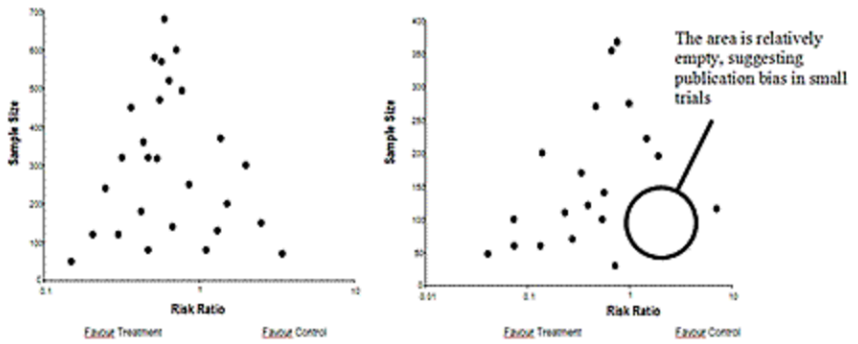

Figura 12 Funnel Plot

6. Referenze

Collette, Laurence, Tomasz Burzykowski, e Fritz H. Schröder. 2006. «Prostate-specific antigen (PSA) alone is not an appropriate surrogate marker of long-term therapeutic benefit in prostate cancer trials». *European Journal of Cancer* 42 (10): 1344–50. doi:10.1016/j.ejca.2006.02.011.

Michiels, Stefan, Aurélie Le Maître, Marc Buyse, Tomasz Burzykowski, Emilie Maillard, Jan Bogaerts, Jan B Vermorken, et al. 2009. «Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data». *The Lancet Oncology* 10 (4): 341–50. doi:10.1016/S1470-2045(09)70023-3.

Moher, David, Kenneth F Schulz, e Douglas G Altman. 2001. «The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials». *The Lancet* 357 (9263): 1191–94. doi:10.1016/S0140-6736(00)04337-3.

Ocana, Alberto, e Ian F. Tannock. 2011. «When Are "Positive" Clinical Trials in Oncology Truly Positive?» *JNCI: Journal of the National Cancer Institute* 103 (1): 16–20. doi:10.1093/jnci/djq463.

Sargent, Dan J. 2011. «Appropriate Endpoints for Clinical Trials and Source of U.S. Food and Drug Administration/Oncologic Drugs Advisory Committee Controversy | 2011 ASCO Annual Meeting | Virtual Meeting | Meeting Library». http://meetinglibrary.asco.org/content/43424?media=sl.

Vale, Claire L., Jayne F. Tierney, e Sarah Burdett. 2013. «Can Trial Quality Be Reliably Assessed from Published Reports of Cancer Trials: Evaluation of Risk of Bias Assessments in Systematic Reviews». *BMJ* 346 (aprile): f1798. doi:10.1136/bmj.f1798.