



I metodi non parametrici per la stima dell'intervallo di confidenza del Rapporto incrementale di costo-efficacia

Non-parametric methods for estimating confidence intervals around the Incremental Cost-Effectiveness Ratio

Michele Petrinco,^{1,2} Dario Gregori,¹ Fulvio Lazzarato,³ Franco Brunello,⁴ Andrea Veltri,⁵ Paola Berchialla,¹ Eva Pagano³

¹ Dipartimento di sanità pubblica e microbiologia, Università di Torino

² Dipartimento di statistica e matematica applicata «D. de Castro», Università di Torino

³ Servizio di epidemiologia dei tumori, Università di Torino e CPO Piemonte

⁴ Divisione di gastro-epatologia, Ospedale San Giovanni Battista di Torino

⁵ Istituto di diagnostica e interventistica radiologica, Università di Torino

Corrispondenza: Eva Pagano, Servizio di Epidemiologia dei tumori, Università di Torino e CPO Piemonte, via Santena 7, 10126 Torino; tel. 011 6336855; e-mail: evapagano@yahoo.com

Riassunto

Negli ultimi anni, l'analisi costo-efficacia è sempre più frequentemente inserita come componente di studi clinici randomizzati. Dal punto di vista statistico, i maggiori sforzi di ricerca hanno interessato la stima del Rapporto incrementale di costo-efficacia (Incremental Cost Effectiveness Ratio, ICER) e del relativo intervallo di confidenza con metodi sia parametrici sia non parametrici. Obiettivo del presente lavoro è presentare brevemente i principali metodi non parametrici, basati su un approccio di tipo bootstrap. Le meto-

dologie presentate vengono applicate ai dati tratti da uno studio clinico randomizzato riguardante l'analisi comparativa di due tecniche per l'ablazione delle lesioni epatiche da carcinoma epatocellulare. L'esempio mostra come l'applicazione dei diversi metodi porti alla stima d'intervalli di confidenza significativamente differenti con conseguenti implicazioni nell'interpretazione dei risultati. (*Epidemiol Prev* 2008; 32(6): 307-313)

Parole chiave: analisi costo-efficacia, rapporto incrementale di costo-efficacia, metodi non-parametrici, bootstrap

Abstract

In recent years, cost-effectiveness analysis has become a frequent component of randomized clinical trials. In statistical terms, the major efforts addressed the method for estimating the Incremental Cost Effectiveness Ratio (ICER) and its confidence interval, both with parametric and non-parametric methods.

The goal of the present work is to briefly present the main non-parametric methods, based on a bootstrap approach. The methods which

have been considered were applied to the data of a randomized clinical trial comparing two alternative approaches to treat hepatocellular carcinoma. The example shows that the application of different methods leads to significantly different confidence intervals' estimates, with consequences on the interpretation of the study results. (*Epidemiol Prev* 2008; 32(6): 307-313)

Key words: cost-effectiveness analysis, incremental cost-effectiveness ratio, non-parametric methods, bootstrap

Introduzione

L'aumento degli studi di valutazione economica ha comportato lo sviluppo di appropriate metodologie statistiche designate a rispondere a questioni di politica sanitaria. L'inclusione dell'analisi prospettica di costo-efficacia come componente integrata di trial clinici randomizzati sta diventando una pratica sempre più diffusa,¹ permettendo di disporre di dati di efficacia e di costo a livello di singolo paziente. A partire dai primi anni Novanta, molteplici studi hanno affrontato il problema della valutazione dell'incertezza nella stima dei dati di sintesi degli studi di costo-efficacia, proponendo soluzioni alternative e dando vita a un acceso dibattito metodologico. I principali sforzi di ricerca riguardo la definizione di misure di costo-efficacia hanno interessato la stima del Rapporto incrementale di costo-efficacia (Incremental Cost Effectiveness Ratio, ICER) e del relativo intervallo di confidenza, con metodi parametrici e non parametrici, il metodo del Beneficio sanitario netto (Net Monetary Benefit) attra-

verso l'utilizzo di approcci frequentisti e bayesiani e la Curva di accettabilità (Acceptability Curve). Tra i metodi proposti, il Rapporto incrementale di costo-efficacia è quello con l'interpretazione più intuitiva e il maggior utilizzo.²

Un precedente contributo in questa sezione della rivista è stato dedicato alla descrizione dei diversi approcci parametrici per la stima dell'intervallo di confidenza dell'ICER.³ Obiettivo del presente lavoro è presentare brevemente la principale metodologia ad approccio non parametrico e le sue varianti a partire dai limiti riscontrati in letteratura nei metodi parametrici.

Il Rapporto incrementale di costo-efficacia (ICER)

L'analisi costo-efficacia è una valutazione comparativa dei costi e dell'efficacia di un trattamento o di una prestazione sanitaria rispetto a una o più sue alternative, in genere di una prestazione terapeutica sperimentale (T) verso il suo controllo (C).⁴

Conoscendo i costi (C) e l'efficacia (E) delle due alternative, l'analisi comparativa può risolversi in una delle seguenti situazioni:

- $C_T - C_C < 0$ e $E_T - E_C > 0$. La terapia sperimentale è meno costosa e più efficace, quindi preferibile (dominanza) rispetto a quella di controllo.
- $C_T - C_C > 0$ e $E_T - E_C < 0$. La terapia sperimentale è più costosa e meno efficace, quindi da rifiutare rispetto al trattamento di controllo (dominanza).
- $C_T - C_C > 0$ e $E_T - E_C > 0$. La terapia sperimentale è più costosa ma anche più efficace; l'assenza di dominanza di una delle due alternative evidenzia un trade-off e rende necessaria la valutazione del rapporto incrementale.
- $C_T - C_C < 0$ e $E_T - E_C < 0$. La terapia sperimentale è meno costosa ma anche meno efficace; l'assenza di dominanza di una delle due alternative evidenzia un trade-off e rende necessaria la valutazione del rapporto incrementale.

In caso di non dominanza di una delle due alternative, la decisione circa il trattamento da preferire deve essere effettuata al margine, in relazione ai costi e ai benefici incrementali.⁵ A tal fine va analizzato l'ICER, che viene stimato su dati campionari dal rapporto tra la differenza dei costi medi e dell'efficacia media delle due alternative. La misura dell'ICER è un rapporto che definisce i costi incrementali che devono essere sostenuti per un'unità di efficacia addizionale conseguente al trattamento sperimentale:

$$R = \frac{\overline{C}_T - \overline{C}_C}{\overline{E}_T - \overline{E}_C} = \frac{\Delta \overline{C}}{\Delta \overline{E}}$$

Il trattamento è da considerarsi costo-efficace se il valore dell'ICER stimato è inferiore a quello di un valore soglia fissato come massimo riferimento accettabile (R_{\max}) in base a considerazioni di «disponibilità a pagare» a livello di decisore politico o di società. Il trattamento statistico dell'ICER risulta problematico in quanto si tratta di una statistica non lineare per la quale non è immediato stimare in modo esatto la varianza. La derivazione di un intervallo di confidenza per l'ICER non è semplice come la sua stima puntuale, dal momento che la varianza di un rapporto tra variabili casuali non ha una forma analiticamente chiusa e la distribuzione di un rapporto è spesso asimmetrica.²

Per far fronte a questo problema sono stati sviluppati alcuni metodi parametrici e altri non parametrici, finalizzati alla stima dell'intervallo di confidenza dell'ICER. I primi poggiano su assunzioni circa la forma della distribuzione dello stimatore \hat{R} , e nella maggior parte dei casi sfruttano proprietà della distribuzione normale, mentre i metodi non parametrici sono basati su procedure bootstrap.

I metodi non parametrici bootstrap

Per ovviare ai problemi legati alle assunzioni circa la forma della distribuzione dello stimatore \hat{R} , sono stati proposti in letteratura metodi di tipo bootstrap⁵⁻⁷. Tale approccio utilizza la distribuzione osservata delle variabili campionarie (costi ed efficacia) per stimare, attraverso un campionamento ripetuto con reimmissione, una distribuzione empirica dell'ICER.

Infatti, ciascun ricampionamento produce una stima dell'ICER e dalla distribuzione empirica identificata è possibile calcolare gli intervalli di confidenza e costruire un test dell'ipotesi, senza effettuare alcuna assunzione circa le caratteristiche della distribuzione.

Per ottenere le stime di R replicate con il bootstrap è necessario effettuare i seguenti passaggi operativi:

- dal campione di osservazioni del gruppo di trattamento T si ri-campionano con reimmissione n_T coppie di costo-efficacia, dalla cui distribuzione campionaria si ottengono le stime di \overline{C}_T^* e \overline{E}_T^* ;
- dal campione di osservazioni del gruppo di controllo C si ri-campionano con reimmissione n_C coppie di costo-efficacia, dalla cui distribuzione campionaria si ottengono le stime di \overline{C}_C^* e \overline{E}_C^* ;
- si calcola il valore replicato con il bootstrap di \hat{R}_B^* attraverso la formula:

$$\hat{R}_B^* = \frac{\overline{C}_T^* - \overline{C}_C^*}{\overline{E}_T^* - \overline{E}_C^*} = \frac{\Delta \overline{C}^*}{\Delta \overline{E}^*}$$

si ripetono i primi tre passaggi un numero di volte (B , pari ad almeno 1.000 repliche) così da ottenere i valori replicati di bootstrap $\hat{R}_1^*, \hat{R}_2^*, \dots, \hat{R}_B^*$.

Le stime di \hat{R} e della sua varianza sono ottenute dalla media e dalla varianza della distribuzione derivante dalle B repliche di bootstrap:

$$\overline{R}^* = \frac{1}{B} \sum_{b=1}^B \hat{R}_b^*$$

$$\text{var}(R^*) = \frac{1}{B-1} \sum_{b=1}^B (\hat{R}_b^* - \overline{R}^*)^2$$

Si noti che al crescere del numero di repliche B , la varianza della distribuzione non sembra necessariamente convergere,⁵ probabilmente a causa della non trascurabile possibilità di avere un denominatore dell'ICER prossimo allo zero. Pertanto, quando il coefficiente di variazione del ΔE è elevato, le stime dell'IC basate sulla varianza stimata con il metodo bootstrap possono non essere affidabili. Inoltre, l'ICER campionario è uno stimatore consistente ma distorto dell'ICER di popolazione;⁸ tale distorsione viene amplificata dal processo di bootstrap e la stima \overline{R}^* più distorta di \hat{R} .

Una volta definita la distribuzione campionaria di \hat{R}^* , sono applicabili i diversi metodi di stima dell'intervallo di confidenza:

Primo metodo

Assumendo che \hat{R} sia distribuito normalmente, è possibile calcolare l'intervallo di confidenza, attraverso la varianza di \hat{R}^* , come:

$$\hat{R} \pm z_{\alpha/2} \sqrt{\text{var}(\hat{R}^*)}$$

Tale intervallo è simmetrico intorno alla stima campionaria e richiedendo la normalità della distribuzione, in caso di violazione dell'assunto l'IC così calcolato non è corretto.

Secondo metodo

Il metodo dei percentili semplicemente utilizza il 100 ($\alpha/2$) e il 100 ($1-\alpha/2$) percentile della distribuzione \hat{R}_B^* come stima dei limiti superiore e inferiore di confidenza. Nel caso tale distribuzione sia non-normale, l'intervallo di confidenza così calcolato sarà asimmetrico attorno alla stima di \hat{R} . Quando invece la distribuzione è simmetrica, questo metodo dà un risultato analogo a quello precedente e i due metodi possono essere considerati intercambiabili. In caso di piccoli campioni, tuttavia, la distribuzione \hat{R}_B^* può essere asimmetrica e gli intervalli di confidenza calcolati con questi due approcci sono distorti, così come la stima di \bar{R}^* ; poiché la distribuzione tende ad avere una asimmetria positiva, il valore di \bar{R}^* risulta essere una sovrastima di \hat{R} . Dal momento che questo metodo non prende in considerazione la stima originaria dell'ICER, si genera un problema legato alla distorsione dello stimatore \bar{R}^* .⁷

Terzo metodo

Il metodo dei percentili corretti per la distorsione è più appropriato quando, a causa della non normalità della distribuzione di \hat{R}_B^* , la stima \bar{R}^* è distorta e risente dell'asimmetria. Il metodo si basa su una correzione algebrica dei percentili individuati come IC, tale correzione è proporzionale all'asimmetria presente nella distribuzione di \hat{R}_B^* . Gli aggiustamenti proposti⁵ sono finalizzati a correggere la distorsione dello stimatore tramite il calcolo di un IC al $(1-\alpha)\%$ per \hat{R} dato da $(\hat{R}^{*(\alpha_1)}, \hat{R}^{*(\alpha_2)})$ dove:

$$\alpha_1 = \Phi(2\hat{z}_0 + \hat{z}_{\alpha/2})$$

$$\alpha_2 = \Phi(2\hat{z}_0 + \hat{z}_{1-\alpha/2})$$

La correzione alla distribuzione per il bias dello stimatore viene portata da \hat{z}_0 , calcolato come:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{R}_b^* < \hat{R}\}}{B}\right)$$

dove il termine tra parentesi non è altro che la proporzione delle repliche bootstrap dove l'ICER stimato \hat{R}_B^* è inferiore alla stima campionaria \hat{R} .

Nelle formule, Φ è la funzione cumulativa della normale standard e \hat{z}_α è il 100α percentile della stessa distribuzione. Il termine \hat{z}_0 , usato per correggere la distorsione misura il bias mediano di \hat{R}_B^* , ovvero la discrepanza tra la mediana di \hat{R}_B^* e di \hat{R} in unità normali.⁹ Nel caso in cui la distribuzione degli \hat{R}_B^* sia simmetrica intorno a \hat{R} , esattamente metà degli \hat{R}_B^* risultano minori o uguali a \hat{R} : in tali situazioni chiaramente \hat{z}_0 è nullo, quindi non viene portata alcuna correzione per la distorsione e questo metodo fornisce risultati equivalenti al precedente.

Quarto metodo

Al fine di migliorare le performance di quest'ultimo metodo riducendo l'errore di coverage associato, ne è stata proposta un'altra versione basata sulla stima di una costante d'accelerazione e definita come metodo dei percentili corretti accelerato.^{10,11} Esso fornisce una stima più accurata dell'intervallo di confidenza e migliori performance in termini di coverage rispetto al metodo standard dei percentili corretti.¹² Oltre al calcolo del termine \hat{z}_0 usato per correggere la distorsione, si rende necessaria la stima di una costante d'accelerazione a .

Con tale metodo l'intervallo di confidenza al $(1-\alpha)\%$ per \hat{R} è dato da $(\hat{R}^{*(\alpha_1)}, \hat{R}^{*(\alpha_2)})$ dove:

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + \hat{z}_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + \hat{z}_{\alpha/2})}\right)$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + \hat{z}_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + \hat{z}_{1-\alpha/2})}\right)$$

La costante d'accelerazione a fornisce l'aggiustamento per l'asimmetria della distribuzione; essa può essere calcolata in modi differenti, ne sono esempi piuttosto utilizzati la stima proposta da Hall e Martin¹¹ e quella da Efron e Tibshirani,⁹ che indicheremo rispettivamente come metodi Pca1 e Pca2. Il metodo dei percentili corretti standard è un caso particolare di quello dei percentili corretti accelerato, in cui nel calcolo di α_1 e α_2 la costante di accelerazione \hat{a} viene posta pari a zero.

Nel caso invece in cui sia \hat{a} sia \hat{z}_0 siano nulli, questo metodo è equivalente a quello dei percentili.

Quinto metodo

Dal momento che l'assunzione di normalità della distribuzione è spesso violata, una possibilità è ottenere i percentili

dalla distribuzione approssimata t formata dai ricampionamenti bootstrap invece che dalla normale standard.

Il metodo bootstrap-t ricorre alla trasformazione dei valori replicati del bootstrap \hat{R}_b^* in una variabile standardizzata t_b^* :

$$t_b^* = \frac{\hat{R}_b^* - \hat{R}}{s.e.(\hat{R}_b^*)}$$

L'errore standard dell'ICER stimato in ogni ricampionamento b è stato calcolato in questo lavoro seguendo l'espansione della serie di Taylor.

A partire dalla distribuzione dei valori t_b^* , una volta ordinata per trovare i percentili di ordine $(\alpha/2)$ e $(1-\alpha/2)$, $\hat{t}^{(\alpha/2)}$ e $\hat{t}^{(1-\alpha/2)}$, definiti come:

$$\alpha = \frac{\#\{t_b^* < \hat{t}^{(\alpha)}\}}{B}$$

l'IC $(1-\alpha)\%$ bootstrap-t è dato da:

$$(\hat{R} - \hat{t}^{(1-\alpha/2)} \sqrt{\text{var}(\hat{R})}, \hat{R} - \hat{t}^{(\alpha/2)} \sqrt{\text{var}(\hat{R})})$$

Ulteriori dettagli possono essere trovati nel libro di Efron e Tibshirani (1993).⁹

Un'applicazione empirica

I metodi sopra esposti sono stati applicati a dati tratti da uno studio clinico randomizzato condotto nel periodo gennaio 2001-settembre 2004 presso l'ASO San Giovanni Battista di Torino e analizzati in esteso in una specifica pubblicazione.¹³ Obiettivo dello studio era l'analisi comparativa dell'efficacia e dei costi delle tecniche percutanee di radiofrequenza (RF) e alcolizzazione (PEI) per l'ablazione delle lesioni epatiche da carcinoma epatocellulare in fase precoce in pazienti cirrotici non trattabili chirurgicamente. L'end-point primario dello studio era la completa risposta a un anno dal trattamento in termini di completa guarigione di tutti i noduli identificati al momento dell'ingresso nello studio.

I costi sono stati stimati rispetto all'intero ciclo di trattamento, comprensivo di una o più sedute, con riferimento ai soli costi ospedalieri, valutati attraverso le tariffe DRG (in euro) attribuite alle schede di dimissioni ospedaliere relative ai diversi episodi di trattamento. Siccome le tariffe DRG non catturano completamente la differenza di costi tra i due trattamenti, i costi relativi agli aghi utilizzati sono stati aggiunti al costo totale di ogni paziente.

L'ICER è stato calcolato sulla base della completa risposta al trattamento a un anno di follow-up.

Nella tabella 1 sono riportati i dati di costo e di efficacia me-

dia dei due gruppi di trattamento e la stima puntuale dell'ICER. Attraverso l'applicazione dei metodi bootstrap standard, bootstrap percentili, bootstrap percentili corretti e bootstrap-t, effettuata attraverso delle routine di R (www.r-project.org), sono state ottenute le diverse stime dell'intervallo di confidenza, riportate nella tabella 2 e relative a 10.000 ricampionamenti bootstrap per ogni metodo considerato.

Il costo incrementale affinché un paziente abbia una completa risposta a un anno di follow-up, utilizzando la tecnica RF rispetto alla PEI, è risultato essere di 8.525€. L'intervallo di confidenza al 95% intorno a tale stima indica che la spesa addizionale può arrivare fino a quasi 25.000€, valore ottenuto con l'utilizzo del metodo bootstrap standard.

Il valore negativo ottenuto da quest'ultimo metodo e da quello basato sulla distribuzione t nella stima dell'estremo inferiore dell'intervallo di confidenza, porterebbe a non escludere la possibilità della dominanza di una delle due alternative. Tuttavia l'analisi dei valori ricampionati con procedura bootstrap (figura 1) indica che soltanto in 2 casi su 10.000 ricampionamenti, i valori corrispondenti dell'ICER risultano inferiori allo 0. La figura 1 riporta la distribuzione dei valori bootstrap nel piano di costo-efficacia, dove l'asse-x rappresenta la differenza negli effetti tra le due terapie e l'asse-y la differenza nei costi. L'ICER di ogni punto, che a sua volta rappresenta una replica bootstrap, è definito come l'intercetta della retta passante per l'origine e per il punto di riferimento I due metodi citati (standard e bootstrap-t), nonostante presentino un estremo inferiore negativo dell'IC per l'ICER, non riescono a comprendere i due valori bootstrap inferiori allo zero, essendo questi ultimi molto elevati in valore assoluto (precisamente uguali a -319.254€ e a -54.629 €), a causa di un denominatore (ΔE) prossimo allo zero.

I risultati derivanti dall'applicazione degli altri metodi, che nella più recente letteratura risultano essere maggiormente utilizzati, sottolineano la dominanza in termini di costo-efficacia della terapia RF rispetto alla PEI.

Le stime dell'IC al 95% sono molto simili per quanto riguarda il metodo dei percentili, quello dei percentili-corretti e quello dei percentili-corretti-accelerati nella versione di Efron e Tibshirani⁹. Rispetto a esse l'IC derivante dall'applicazione del metodo dei percentili-corretti-accelerati nella versione di Hall e Martin¹¹ risulta più spostato verso sinistra, mantenendosi in ogni caso al di sopra del valore zero, rappresentante la non dominanza della terapia in questione.

Il criterio principale per valutare le performance dei diversi metodi nel calcolo dell'IC dell'ICER, è stata la probabilità complessiva di miscoverage, la quale non è altro che la percentuale dei campioni bootstrap in cui il corrispondente ICER cade al di fuori dell'intervallo di confidenza calcolato con una data metodologia. La procedura col miglior calcolo dell'intervallo di confidenza sarà quella col livello di miscoverage più vicino possibile al valore target del 5%.

La simmetria del miscoverage è stata utilizzata come ulteriore criterio per la valutazione dei diversi metodi (tabella 3).

	RF*	PEI**	Differenza (RF - PEI) (IC 95%)
n. pazienti	69	69	
costo medio in euro (deviazione standard)	6.541,13 (4.098,99)	4.069,95 (4.539,35)	2.471,17 (1.014,96 - 3.927,38)
proporzione di pazienti con completa risposta ad un anno	0,65	0,36	0,29 (0,13-0,45)
ICER (Rapporto incrementale di costo-efficacia: euro per completa risposta ad un anno)	8.525,54		
* Tecnica di radiofrequenza (Radiofrequency).			
** Tecnica di alcolizzazione (Percutaneous Ethanol Injection).			
[†] completa risposta ad un anno dal trattamento in termini di guarigione di tutti i noduli identificati al momento dell'ingresso nello studio.			

Tabella 1. Costi, efficacia media e Rapporto incrementale di costo-efficacia (ICER) per i 2 gruppi di trattamento.

Table 1. Mean costs, effects and Incremental Cost-Effectiveness Ratio (ICER) for the 2 treatment groups.

Metodo Bootstrap (numero di repliche = 10000)	ICER* (IC 95%)
standard	8.525,54 (-7.916,21 - 24.967,29)
percentili	8.525,54 (3.335,23 - 24.346,04)
percentili-corretti	8.525,54 (3.426,35 - 24.313,68)
percentili-corretti-accelerati 1 (Pca1)	8.525,54 (1.252,18 - 17.345,48)
percentili-corretti-accelerati 2 (Pca2)	8.525,54 (3.341,40 - 23.050,22)
bootstrap-t	8.525,54 (-2.132,01 - 18.083,94)
* Rapporto incrementale di Costo-Efficacia (Incremental Cost-Effectiveness Ratio)	

Tabella 2. Valore vero e IC 95% dell'ICER secondo diverse metodologie di stima bootstrap.

Table 2. True value and 95% CI for the ICER according to different bootstrap estimation methods.

	Metodo Bootstrap					
	standard	bootstrap-t	percentili	percentili corretti	pca1*	pca2**
miscoverage [†] (%)	2,7	5,9	6	6,3	7,3	6,7
grado di simmetria (%)	3,70	1,69	53,33	55,56	4,11	47,76
* Percentili-corretti-accelerati 1						
** Percentili-corretti-accelerati 2						
[†] % dei campioni bootstrap in cui l'ICER cade al di fuori dell'IC calcolato con un dato metodo						

Tabella 3. Miscoverage e grado di simmetria del miscoverage dei 6 metodi.

Table 3. Miscoverage and Degree of symmetry of miscoverage of the 6 methods.

Si ha miscoverage simmetrico quando il limite superiore dell'IC cade al di sotto del valore dell'ICER con la stessa frequenza con cui il limite inferiore dell'IC cade al di sopra del valore dell'ICER. Nella tabella 3 il grado di simmetria del miscoverage è stato calcolato come rapporto tra il numero di volte in cui l'ICER cade al di sotto dell'estremo inferiore dell'IC su il numero totale di volte in cui l'ICER non risulta compreso nel corrispondente IC. Un valore del grado di simmetria uguale al 50% rappresenta miscoverage simmetrico; un valore superiore al 50% indica un IC che sovrastima il vero ICER; inferiore al 50% rappresenta un IC che sottostima l'importanza del vero ICER.¹⁴

Analizzando congiuntamente i valori del miscoverage e del corrispondente grado di simmetria per ciascuno dei metodi considerati, si possono notare alcune caratteristiche riguardanti l'applicazione dei diversi metodi. La metodologia standard, che fornisce come risultato un ampio IC ha, conseguentemente, una bassa percentuale di miscoverage associata a una forte sottostima dell'IC del vero ICER. Simili conclusioni si possono trarre sul metodo bootstrap-t, che presenta un valore di miscoverage

vicino al target del 5% ma il più alto livello di asimmetria tra i metodi considerati. Anche la tecnica Pca1 non soddisfa contemporaneamente i due indicatori presi a riferimento, con in particolare un valore elevato dell'indice di asimmetria. Al contrario, il metodo dei percentili, quello dei percentili corretti e il Pca2 presentano buone performance in termini di miscoverage (compreso tra 6% e 6,7%) e sono gli unici ad avere simmetria nel miscoverage (grado di simmetria tra 47,76% e 53,33%). Sembrano quindi i più adatti nel fornire risultati realistici nell'analisi di questi dati.

Discussione

I dati di costo ottenuti da singoli individui in uno studio di economia sanitaria presentano, quasi nella totalità delle situazioni reali, una distribuzione altamente asimmetrica, dato che vi è una porzione relativamente piccola di soggetti con alti costi medici se paragonati al resto della popolazione. Un'altra caratteristica di molti studi di costo-efficacia è la scarsa disponibilità di dati, per cui si ha spesso a che fare con dimensioni campionarie relativamente piccole, motivo che spinge alla con-

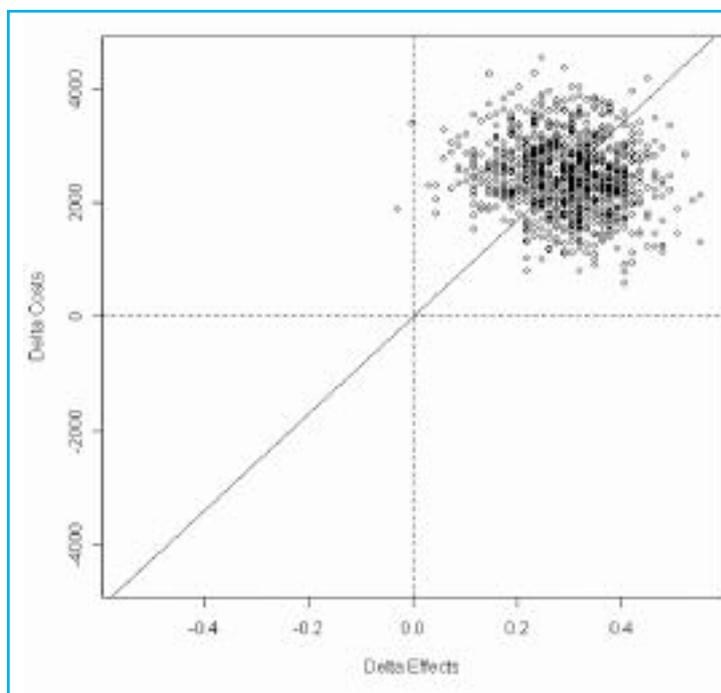


Figura 1. Distribuzione dei valori calcolati con procedura bootstrap (10000 repliche) nel piano di costo-efficacia.

Figure 1. Distribution of the values calculated with bootstrap procedure (10000 samples) on the Cost-Effectiveness plan.

Legenda: nel grafico l'asse-x rappresenta la differenza (RF-PEI) negli effetti tra i 2 trattamenti, l'asse-y la differenza nei costi. Ogni punto corrisponde ad un ricampionamento bootstrap.

duzione di trial multicentrici dove la raccolta dei dati avviene in centri differenti e l'analisi viene condotta con appropriate tecniche statistiche. L'asimmetria a destra della distribuzione dei costi e le piccole dimensioni campionarie a disposizione hanno portato a criticare fortemente l'utilizzo di metodi statistici basati sulla distribuzione normale. Tuttavia, può risultare difficile considerare appropriati modelli probabilistici per la distribuzione dei costi a livello della popolazione.

Considerate queste difficoltà, una possibile soluzione consiste nel basare le analisi su modelli statistici non-parametrici, che possono essere applicati senza assunzioni riguardo la forma della distribuzione della popolazione.

Tuttavia anche i metodi non parametrici non sono esenti da problematiche metodologiche.

Una critica comune riguardo l'uso della procedura bootstrap è che gli ICER stimati usando ricampionamenti bootstrap possono non essere veramente realistici.

Per esempio, se vengono generati un campione bootstrap con un effetto incrementale positivo molto piccolo e un altro campione bootstrap con un effetto incrementale negativo molto piccolo, assumendo che il costo incrementale sia positivo in entrambi i campioni, ne consegue che il primo campione porterà a un ICER positivo molto grande mentre il secondo a un ICER negativo molto elevato. I due campioni dell'esempio sono concettualmente simili, poiché presentano entrambi un costo incrementale positivo e un effetto incrementale di poca rilevanza, ma le associate stime bootstrap dell'ICER sono drasticamente differenti: la differenza tra i due ICER non riflette appropriatamente le differenze tra i 2 campioni.

I metodi bootstrap risentono in generale del problema della

discontinuità dell'ICER, per $\Delta E = 0$.¹⁵

Anche nel nostro esempio, i metodi considerati non riescono a «catturare» nel calcolo dell'IC i valori bootstrap dell'ICER corrispondenti a un ΔE prossimo allo zero.

Per valori di ΔE tendenti a zero (graficamente prossimi all'asse delle ordinate del piano di costo-efficacia, figura 1) e valori di ΔC non trascurabili, l'ICER presenta valori molto elevati, positivi o negativi.

I valori negativi, nell'ordinamento degli \hat{R}_B^* , dovrebbero essere posti oltre i valori elevati positivi; il loro posizionamento derivante dall'ordinamento matematico può distorcere le stime di entrambi i limiti.¹⁶

Per ovviare alle difficoltà derivanti da valori di ΔE generati dal processo di bootstrap tendenti a zero, Cook e Heise¹⁷ hanno proposto un approccio detto della trasformazione angolare che, nonostante possa presentare qualche problema di ritrasformazione, evita in queste situazioni una sovrastima degli IC.

La situazione risulta esasperata quando la differenza d'efficacia tra i 2 trattamenti è nulla, portando a zero il denominatore del rapporto e facendo tendere l'ICER a $+\infty$ o a $-\infty$, a seconda del segno del costo incrementale. Statisticamente, un intervallo di confidenza infinito riflette l'ampia variazione del rapporto causata da un denominatore molto vicino allo zero e suggerisce che sulla base dei dati a disposizione potrebbe non esserci un'adeguata fiducia nel determinare dei limiti di variabilità del vero ICER.²

Nella nostra applicazione a dati reali, la costruzione di indicatori di performance riguardanti i metodi considerati ha evidenziato prestazioni superiori di tre metodi (Percentili, Per-

centili corretti e Pca2), i quali forniscono risultati maggiormente affidabili e risultano quindi i più adeguati per l'analisi di questi dati. A nostro avviso, il calcolo dell'IC con le diverse metodologie bootstrap andrebbe sempre accompagnato da un'analisi della bontà di performance di ogni singolo approccio, ottenendo in tal modo un supporto nella fase di valutazione della metodologia più adeguata.

Ringraziamenti: il presente lavoro è stato condotto nell'ambito del Progetto «Valutazione economica e di qualità negli studi di epidemiologia clinica», finanziato dalla Compagnia di San Paolo.

Bibliografia

1. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ* 2004; 13(5): 461-75.
2. Fan MY, Zhou XH. A simulation study to compare methods for constructing confidence intervals for the incremental cost-effectiveness ratio. *Health Serv Outcomes Res Method* 2007; 7(1-2): 57-77.
3. Pagano E, Silvestri L, Braga M, Gregori D. La presentazione dei risultati nelle valutazioni economiche di tipo stocastico. *Epidemiol Prev* 2004; 28(4-5): 272-78.
4. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for economic evaluation of health care programmes*. New York, Oxford University Press, 1997.
5. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ* 1997; 6(4): 327-40.
6. Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Stat Med* 1996; 15(13): 1447-58.
7. Tambour M, Zethraeus N. Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results. *Health Econ* 1998; 7(2): 143-47.
8. Stinnet AA. Adjusting for bias in *c/e* ratio estimates. *Health Econ* 1996; 5(5): 470-72.
9. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. Chapman & Hall, CRC press, 1993.
10. Efron B. Better bootstrap confidence intervals. *J Am Stat Assoc* 1987; 82: 171-85.
11. Hall P, Martin M. On the bootstrap and two-sample problems. *Austral J Statist* 1988; 30A: 179-92.
12. Briggs AH, Mooney CZ, Wonderling DE. Constructing confidence intervals for cost-effectiveness ratios: an evaluation of parametric and non-parametric techniques using Monte Carlo simulation. *Stat Med* 1999; 18(23): 3245-62.
13. Brunello F, Veltri A, Carucci P et al. Radiofrequency ablation versus Ethanol injection for early hepatocellular carcinoma: a randomised controlled trial. *Scand J Gastroenterol*; In press.
14. Polsky D, Glick HA, Willke R et al. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Econ* 1997; 6(3): 243-52.
15. Heitjan DF, Moskowitz AJ, Whang W. Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Econ* 1999; 8(3): 191-201.
16. Heitjan DF, Moskowitz AJ, Whang W. Problems with interval estimates of the incremental cost-effectiveness ratio. *Med Decis Making* 1999; 19(1): 9-15.
17. Cook JR, Heyse JE. Use of an angular transformation for ratio estimation in cost-effectiveness analysis. *Stat Med* 2000; 19(21): 2989-3003.

abbonamenti@inferenze.it
tel. 02-48702283

EPIDEMIOLOGIA
& PREVENZIONE

2009

ABBONARSI E' UTILE