

PROFESSIONE NEFROLOGO

# Il valore delle utilità per la stima dei QALYs: un problema aperto



**Francesco S. Mennini**<sup>1,2</sup>

(1) CEIS, Economic Evaluation and HTA (EEHTA), Università "Tor Vergata", Roma

(2) Kingston University, London, UK

## Abstract

L'impiego degli anni di vita guadagnati corretti per la qualità (QALYs) come misura di efficacia di un intervento è una pratica ormai consolidata nelle valutazioni economiche.

I QALYs combinano in una singola misura gli anni di vita guadagnati a seguito di un intervento con la qualità della vita, essi devono rappresentare l'utilità dell'individuo associata ai diversi stati di salute.

## Introduzione

L'impiego degli anni di vita guadagnati corretti per la qualità (QALYs) come misura di efficacia di un intervento è una pratica ormai consolidata nelle valutazioni economiche.

I QALYs combinano in una singola misura gli anni di vita guadagnati a seguito di un intervento con la qualità della vita, essi devono rappresentare l'utilità dell'individuo associata ai diversi stati di salute.

Dalla letteratura vi sono quattro caratteristiche che devono essere rispettate affinché il QALY sia una misura di utilità attendibile:

1. i due attributi della quantità e qualità della salute sono mutuamente indipendenti;
2. gli individui scambiano quantità e qualità in maniera proporzionalmente costante;
3. gli individui sono neutrali al rischio
4. deve essere soddisfatta la condizione di indipendenza additiva lungo gli intervalli temporali nell'ipotesi di stati di salute temporanei.

Se le preferenze degli individui violano queste assunzioni la validità dei QALYs viene messa in discussione.

Ancora, è importante aggiungere che il valore attribuibile ad uno stato di salute non può prescindere dal contesto in cui quello stato di salute si verifica. Alcuni studi hanno testato la condizione di additività per stati di salute temporanei dimostrando la sua inconsistenza [1] [2]. Altri studi hanno cercato di esplorare i fattori che influenzano le preferenze temporali e che contribuirebbero alla violazione della condizione di additività [3].

Tuttavia, in letteratura esistono lavori empirici in cui si dimostra che l'ipotesi di *trade-off* proporzionalmente costante tra longevità e stato di salute possa fornire una ragionevole buona approssimazione delle preferenze, accompagnata da un'ottimale gestione della rac-

colta delle informazioni. Ad esempio, nello studio di Stalmeier [4] i valori per stati di salute in corrispondenza degli ultimi 10, 25 e 50 anni sono simili e in quello di Bleichrodt e Johannesson [5] i valori medi per stati di salute negli ultimi 10 anni e per stati di salute negli ultimi 30 anni sono pressoché identici tra loro.

Il criterio di massimizzazione QALY è fondato, in ogni caso, su una teoria di utilità ponderate che richiede il rispetto delle condizioni fin qui citate [6].

In questo contesto, molti fattori sociali e culturali (età, sesso, sottostante co-morbidità, status economico, stili di vita, educazione, capacità cognitive, modelli organizzativi, religione, ecc.) possono influenzare le preferenze dello stato di salute percepite dai pazienti [7] [8] [9] [10] (full text).

Infatti, anche se raccolti con una metodologia identica, sono stati ottenuti differenti valori di QALY per la stessa malattia in paesi diversi [11].

In un confronto che coinvolge sette studi condotti per valutare l'utilità media per pazienti sottoposti a dialisi con malattia renale allo stadio finale, Arnesen e Trommald hanno trovato, ad esempio, che i valori di utilità stimati in ciascuno studio evidenziavano una elevata variabilità (fino a 2,15 volte confrontando il valore minimo ed il massimo) [11]. La variabilità era attribuibile sia ad un fattore temporale (studi meno recenti stimavano utilità differenti rispetto a studi di anni successivi) sia ad un fattore spaziale (utilità stimate su pazienti americani erano notevolmente differenti rispetto a utilità stimate su pazienti canadesi).

Quanto sopra, ci porta a pensare come, al fine di fornire i servizi sanitari più appropriati, i decisori politici dovrebbero prendere la loro decisione sulla base di valutazioni economiche affidabili. Pertanto, la necessità della valutazione di programmi di utilità, nonché valori di QALY specificamente raccolti in ciascun paese (o aree geografiche con caratteristiche socio-economiche simili) dovrebbe essere percepita come una reale necessità ed urgenza.

Conseguentemente, la definizione e la costruzione delle utilità diviene un passaggio cruciale in termini di affidabilità del valore del QALY che scaturisce dalle analisi di valutazione economica.

Le modalità di costruzione delle utilità assumono, quindi, un'importanza notevole. In genere il metodo che permette di assegnare un peso ai vari possibili livelli di qualità degli anni di vita guadagnati, circa la desiderabilità delle varie condizioni di salute ipotizzabili, si basa sul giudizio soggettivo di specialisti (i medici), dei soggetti interessati (i pazienti) o degli individui comuni (i consumatori).

Esistono numerosi metodi per la stima dell'utilità:

- il *Time Trade Off* (TTO),
- lo *Standard Gamble*,
- il *Visual Analogue Scale*
- e il *Rating Scale*.

La preferenza per ogni stato di salute viene "pesata" secondo la valutazione media calcolata su un collettivo con i metodi citati in precedenza, indicando la scelta del paziente in una scala che va da un minimo di zero (decesso) a un massimo di uno (perfetta salute): questo è il punteggio di utilità (*utility score*). Una importante malattia per esempio corrisponde a 0,5. Il punteggio viene poi moltiplicato per il periodo di tempo passato in quel determinato stato di salute.

Un anno in piena salute corrisponde a 1 QALY (cioè a 1 moltiplicato per 1). Due anni trascorsi con un serio problema di salute (cui corrisponde un punteggio di 0,5) danno 1 QALY (2 mol-

tiplicato per 0,5) e così via. Per un determinato paziente un anno passato in perfetta salute potrebbe essere preferibile a 2 anni con una malattia importante o viceversa.

Il TTO (metodo delle alternative temporali), tra i metodi per la misurazione dell'utilità, è il più comunemente utilizzato. Il TTO consiste nel chiedere ai pazienti quanta quantità di vita sono disposti a sacrificare per essere alleviati da un problema di salute.

La preferenza, dunque, a restare per un tempo  $t$  in una condizione cronica  $i$  o se passare un tempo  $x < t$  in condizione di salute. Il tempo  $x$  viene fatto variare finché la scelta tra le due alternative è indifferente. A questo punto il valore di preferenza richiesto per lo stato  $i$  sarà dato da  $h_i = x/t$ .

Idealmente, quindi, un peso attribuito ad un certo gruppo diagnostico dovrebbe essere riproducibile da chiunque. Ma, questo non sempre accade.

Sette studi in letteratura, ad esempio, hanno pubblicato una media dei pesi della qualità della vita per "malattia renale terminale in pazienti che subiscono emodialisi in ospedale" [12] [13] [14] [15] [16] [17] (full text) [18].

I pesi variano da 0,39 a 0,84. Uno qualsiasi dei pesi della qualità della vita pubblicati potrebbero essere utilizzati in uno studio costo-utilità per l'emodialisi. La scelta del peso potrebbe determinare se l'emodialisi possa essere considerata un intervento con rapporto di costo-efficacia o meno.

Questo esempio pone degli interrogativi riguardo l'uso di strumenti apparentemente precisi come il costo per QALY. Questi pesi non sembrano, dunque, rispondere alla caratteristica di moneta comune che dovrebbe garantire allorquando andiamo ad utilizzare studi esteri per valutazioni economiche in Italia.

In particolare: è corretto utilizzare misure di qualità della vita percepita su una popolazione per ottenere misure di valutazione associabili ad altre realtà? Può, quindi, una misura ottenuta su una popolazione con determinate caratteristiche socio-economiche rispecchiare le preferenze di vita di altre popolazioni che vivono in realtà molto differenti?

Nasce, quindi, la necessità di ottenere dei valori delle utilità espresse che siano più specifici con riferimento alla popolazione sulla quale si andrà a effettuare lo studio di Costi - Utilità. In altre parole, una soluzione al problema della percezione reale della QoL di una popolazione è quella di ottenere per ogni paese, o territorio con caratteristiche socio economiche simili, un suo definito valore di utilità e quindi un QALYs personale che rispecchi la realtà sul quale gli studi vengono effettuati. Anziché prendere per buoni i risultati degli altri paesi, farne dei propri e cercare di arrivare il più possibile vicini alla realtà in cui si sta lavorando.

Ma questo è possibile? Quale è la situazione al riguardo in Italia?

Recentemente [19] è stato pubblicato uno studio che aveva proprio l'obiettivo di studiare la fattibilità di una procedura standardizzata *Time Trade-off* (TTO) per quantificare la perdita di utilità negli stati di salute di pazienti affetti da patologie indotte da HPV in Italia.

È stato il primo studio che ha analizzato la perdita di utilità utilizzando una *computer-guided administration* di una procedura di TTO per l'elicitazione standardizzata delle preferenze dello stato di salute in Italia.

Questo studio pilota multicentrico ha dimostrato la fattibilità e l'appropriatezza di un TTO standardizzato per valutare l'utilità in pazienti affetti sia da CIN2-3 o verruche anogenitali o cancro cervicale invasivo in Italia, fornendo, quindi, per la prima volta un valore di utilità da poter utilizzare in maniera specifica per la popolazione italiana.

La percezione delle donne del loro stato di salute è elevato, risultando pari a  $0,91 \pm 0,21$ . Il valore di utilità medio è stato pari a  $0,73 \pm 0,22$ ,  $0,71 \pm 0,35$  e  $0,02 \pm 0,08$  per CIN2-3, verruche anogenitali e cancro cervicale invasivo, rispettivamente. L'analisi dei coefficienti di correlazione ha dimostrato che nessuna delle tre variabili considerate ha avuto un effetto significativo sulla misura dell'utilità.

Ma, un risultato altrettanto importante è quello relativo al confronto con quanto pubblicato da uno studio americano [20]. Infatti, i valori espressi dalle due popolazioni analizzate (USA per lo studio americano e ITA per lo studio italiano) sono molto differenti a conferma del fatto che non è propriamente corretto, per quanto spiegato precedentemente, utilizzare misure di qualità della vita percepita su una popolazione per ottenere misure di valutazione associabili ad altre realtà. Infatti, questo potrebbe portare a prendere delle decisioni che, basandosi su risultati provenienti da altri Paesi, non riescono a garantire il corretto equilibrio tra risorse "scarse" e accessibilità alle cure maggiormente efficaci e costo efficaci.

## Discussione

Si può, quindi, affermare che, da quanto emerso dalle analisi e dagli studi descritti, il confronto dei valori di utilità diviene ancora più complicato quando vengono considerate/confrontate popolazioni appartenenti a diversi paesi. Ancora, la valutazione di utilità può essere influenzata da diversi fattori, fundamentalmente legati a questioni sociali e culturali che devono essere aggiunti allo specifico potenziale bias procedurale comprese: le descrizioni dello stato di salute, onere cognitivo e il bias legato all'intervistatore.

La descrizione dello stato di salute potrebbe non essere un problema se l'intervistato ha sperimentato precedentemente problemi di salute indotti da una particolare malattia. Al contrario, uno stato immaginario di salute richiede spiegazioni e descrizioni che possono essere interpretate positivamente o negativamente.

Queste descrizioni possono essere significativamente differenti a seconda degli intervistatori [21] [22]. Inoltre, la descrizione fornita oralmente potrebbe sovraccaricare o indebolire la memoria degli intervistati.

Questi effetti di confusione possono essere evitati fornendo una descrizione dello stato salute in forma scritta e anche mostrando foto insieme al testo per dare informazioni standard e neutre nonché un richiamo alla memoria.

L'elicitazione delle utilità può anche essere una sfida cognitiva. Quindi, gli intervistati possono trovare alcune difficoltà con i numeri o anche con la comprensione degli anni di vita rimanenti proporzionali al *trade off* (allo scambio) [23]. Questi inconvenienti possono essere affrontati con una spiegazione accurata e standardizzata, utilizzando anche presentazioni figurative esplicite e anche testando la comprensione degli intervistati sulle loro attività. I metodi di elicitazione computerizzata possono migliorare notevolmente la valutazione dell'utilità [24] [25].

Conseguentemente, si suggerisce di ottenere dei valori delle utilità che siano più specifici con riferimento alla popolazione sulla quale si andranno a effettuare studi di HTA e Valutazione economica. Quindi, si sottolinea l'esigenza di effettuare studi che permettano di ottenere per ogni paese, o territorio con caratteristiche socio economiche simili, uno specifico valore di utilità e quindi un QALYs personale che rispecchi la realtà sulla quale gli studi vengono effettuati.

In questo modo si possono ottenere risultati ancora più robusti in grado di rappresentare, tanto dal punto di vista sociale che economico e con ridotti margini di incertezza, il reale impatto in termini di Costo Utilità delle tecnologie introdotte sul mercato.

## Bibliografia

- [1] MacKeigan LD, O'Brien BJ, Oh PI et al. Holistic versus composite preferences for lifetime treatment sequences for type 2 diabetes. *Medical decision making : an international journal of the Society for Medical Decision Making* 1999 Apr-Jun;19(2):113-21
- [2] Richardson J, Hall J, Salkeld G et al. The measurement of utility in multiphase health states. *International journal of technology assessment in health care* 1996 Winter;12(1):151-62
- [3] Kongnakorn T (2004). Development and test of a new method for preference measurement for multistate health profiles. PhD thesis. Georgia Institute of Technology.
- [4] Stalmeier PF, Bezembinder TG, Unic IJ et al. Proportional heuristics in time tradeoff and conjoint measurement. *Medical decision making : an international journal of the Society for Medical Decision Making* 1996 Jan-Mar;16(1):36-44
- [5] Bleichrodt H, Johannesson M The validity of QALYs: an experimental test of constant proportional tradeoff and utility independence. *Medical decision making : an international journal of the Society for Medical Decision Making* 1997 Jan-Mar;17(1):21-32
- [6] Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Operations Research* 1980; 28: 206-224
- [7] Loomes G. Disparities between health state measures: is there a rational explanation? In: Gerrard Bed. *The economics of rationality*. London, Routledge 1993; pag: 149-178
- [8] Mehrez A, Gafni A Preference based outcome measures for economic evaluation of drug interventions: quality adjusted life years (QALYs) versus healthy years equivalents (HYEs). *PharmacoEconomics* 1992 May;1(5):338-45
- [9] Tramarin A. Le basi metodologiche e le implicazioni cliniche dell'approccio economico alla farmacoterapia. *Medical Publisher International, Milano, 1999; pag. 57-58*
- [10] Sassi F, Archard L, Le Grand J et al. Equity and the economic evaluation of healthcare. *Health technology assessment (Winchester, England)* 2001;5(3):1-138 (full text)
- [11] Arnesen T, Trommald M Roughly right or precisely wrong? Systematic review of quality-of-life weights elicited with the time trade-off method. *Journal of health services research & policy* 2004 Jan;9(1):43-50
- [12] Hornberger JC, Redelmeier DA, Petersen J et al. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. *Journal of clinical epidemiology* 1992 May;45(5):505-12
- [13] Churchill D, Morgan J, Torrance G. Quality of life in endstage renal disease. *Peritoneal Dialysis Bulletin* 1984; 4: 20-23
- [14] Churchill DN, Torrance GW, Taylor DW et al. Measurement of quality of life in end-stage renal disease: the time trade-off approach. *Clinical and investigative medicine. Medecine clinique et experimentale* 1987 Jan;10(1):14-20
- [15] Molzahn AE, Northcott HC, Hayduk L et al. Quality of life of patients with end stage renal disease: a structural equation model. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1996 Aug;5(4):426-32
- [16] Lenert LA, Hornberger JC Computer-assisted quality of life assessment for clinical trials. *Proceedings : a conference of the American Medical Informatics Association / ... AMIA Annual Fall Symposium. AMIA Fall Symposium 1996;:992-6*
- [17] Sesso R, Yoshihiro MM Time of diagnosis of chronic renal failure and assessment of quality of life in haemodialysis patients. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 1997 Oct;12(10):2111-6 (full text)
- [18] Hamel MB, Phillips RS, Davis RB et al. Outcomes and cost-effectiveness of initiating dialysis and continuing aggressive care in seriously ill hospitalized adults. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. *Annals of internal medicine* 1997 Aug 1;127(3):195-202
- [19] Mennini FS, Panatto D, Marcellusi A et al. Time trade-off procedure for measuring health utilities loss with human papillomavirus-induced diseases: a multicenter, retrospective, observational pilot study in Italy. *Clinical therapeutics* 2011 Aug;33(8):1084-1095.e4
- [20] Kulasingam SL, Myers ER Potential health and economic impact of adding a human papillomavirus vaccine to screening programs. *JAMA : the journal of the American Medical Association* 2003 Aug 13;290(6):781-9
- [21] Llewellyn-Thomas H, Sutherland HJ, Tibshirani R et al. Describing health states. Methodologic issues in obtaining values for health states. *Medical care* 1984 Jun;22(6):543-52
- [22] McNeil BJ, Pauker SG, Tversky A. On the framing of medical decisions. In D. E. Bell, H. Raiffa, & A. Tversky (Eds.). *Decision making: Descriptive, normative, and prescriptive interactions*. (). Cambridge University Press 1988; pag. 562-568
- [23] Woloshin S, Schwartz LM, Moncur M et al. Assessing values for health: numeracy matters. *Medical decision making : an international journal of the Society for Medical Decision Making* 2001 Sep-Oct;21(5):382-90
- [24] MK. Goldstein, DE. Miller, S. Davies, and A. M. Garber. Quality of life assessment software for computer-inexperienced older adults: multimedia utility elicitation for activities of daily living. *Proc AMIA Symp* 2002; 295-299.
- [25] LA. Lenert and JC. Hornberger. Computer-assisted quality of life assessment for clinical trials. *Proc AMIA Annu Fall Symp* 1996; 992-996.