

# Early stage Healthcare Technology Assessment: valutazione delle tecnologie biomediche in fase di sviluppo

*Leandro Pecchia\**, *Paolo Melillo\*\**, *Umberto Bracale\*\*\**, *Michael Craven\*\*\*\**,  
*Marcello Bracale\*\*\*\*\**

## 1. IL CONCETTO DI TECNOLOGIA

Il termine *technology* ha in inglese un significato più ampio rispetto all'uso comune del termine italiano tecnologia. Mentre in Italia per la maggior parte delle persone, talvolta anche operanti nel Servizio Sanitario Nazionale (SSN), tecnologia evoca alla memoria principalmente un dispositivo, nella comunità scientifica internazionale per *healthcare technology* si intendono una gamma di metodi utilizzati per promuovere la salute, prevenire e curare le malattie e migliorare la riabilitazione e assistenza a lungo termine, tra cui:

1. farmaci;
2. dispositivi medici: ben definiti dalle direttive europee<sup>1</sup>;
3. procedure: quali le tecniche chirurgiche, l'agopuntura, la consulenza medica;
4. ambienti di cura: ambienti ospedalieri, ambulatoriali;
5. programmi sanitari: dagli screening ai programmi di dimissione protetta.

Per quanto riguarda l'inclusione del farmaco nel concetto di tecnologia, basti pensare alle moderne tecniche di *drug delivering*, che prevedono l'impiego di com-

\* Electrical Systems and optics research division, Faculty of Engineering, University of Nottingham, NG7 2RD, UK. Health Technology Assessment Division, The International Federation for Medical and Biological Engineering (IFMBE).

\*\* Dipartimento di Oftalmologia, Seconda Università degli studi di Napoli. Dipartimento di Elettronica, Informatica e Sistemistica, Alma Mater Studiorum - Università di Bologna.

\*\*\* Dipartimento di Scienze Chirurgiche, Anestesiologiche-Rianimatorie e dell'emergenza Giuseppe Zannini, Università degli studi di Napoli Federico II.

\*\*\*\* Electrical Systems and optics research division, Faculty of Engineering, University of Nottingham, NG7 2RD, UK.

\*\*\*\*\* Dipartimento di Ingegneria Biomedica, Elettronica e delle Telecomunicazioni, Università degli studi di Napoli Federico II.

<sup>1</sup> Direttiva 93/42/CEE (dispositivi medici in generale); Direttiva 2007/47/CE (modifica la 93/42 ed equipara il software medico stand alone ad un dispositivo); Direttiva 98/79/CE (dispositivi medico-diagnostics in vitro); Direttiva 90/385/CEE (dispositivi medici impiantabili attivi).



Fig. 1 – Processo produttivo in sanità.

petenze multi- ed inter-disciplinari per modificare il profilo di rilascio, di assorbimento, di distribuzione e di eliminazione di una sostanza farmacologica, al fine di migliorarne l'efficacia e la sicurezza. In base a tale definizione, per *Health Technology Assessment* (HTA) si intende un processo decisionale scientifico, multidisciplinare e multidimensionale che consente la valutazione comparata degli effetti desiderati ed indesiderati di una o più tecnologie per la cura della salute al fine di scegliere, tra quelle individuate, quella che massimizzi la soddisfazione dei bisogni (non solo medici) identificati, con lo scopo di supportare decisioni di politica sanitaria a tutti i livelli.

Tuttavia, la misurazione del soddisfacimento di un bisogno in sanità è un'operazione molto complessa, e talvolta anche difficile. Il miglioramento della salute, o il suo mantenimento, dipende solo in parte dall'azione diretta di una tecnologia ed anche il contributo di questa parte è molto complesso da misurare e valutare. Uno schema ampiamente diffuso che mette in relazione il processo produttivo del SSN ed altre fattori concomitanti che influiscono sulla salute del cittadino è riportato in figura 1.

Usando la schematizzazione riportata in figura 1 è utile richiamare tre concetti (Anthony and Young, 2003):

$$Efficacia = \frac{Output}{Input} \quad (1)$$

$$Efficienza = \frac{Outcome}{Output} \quad (2)$$

$$Performane = Efficacia * Efficienza = \frac{Outcome}{Input} \quad (3)$$

In base ai tre concetti introdotti, possiamo dire che lo scopo dell'HTA è quello di individuare la tecnologia con la migliore performance. L'HTA è dunque quel processo decisionale in base al quale nei paesi moderni si decide se adottare o meno una nuova tecnologia per la cura o il mantenimento della salute, in base alla sua efficacia ed efficienza.

Ecco perché è importante oggi considerare tali variabili molto prima che una

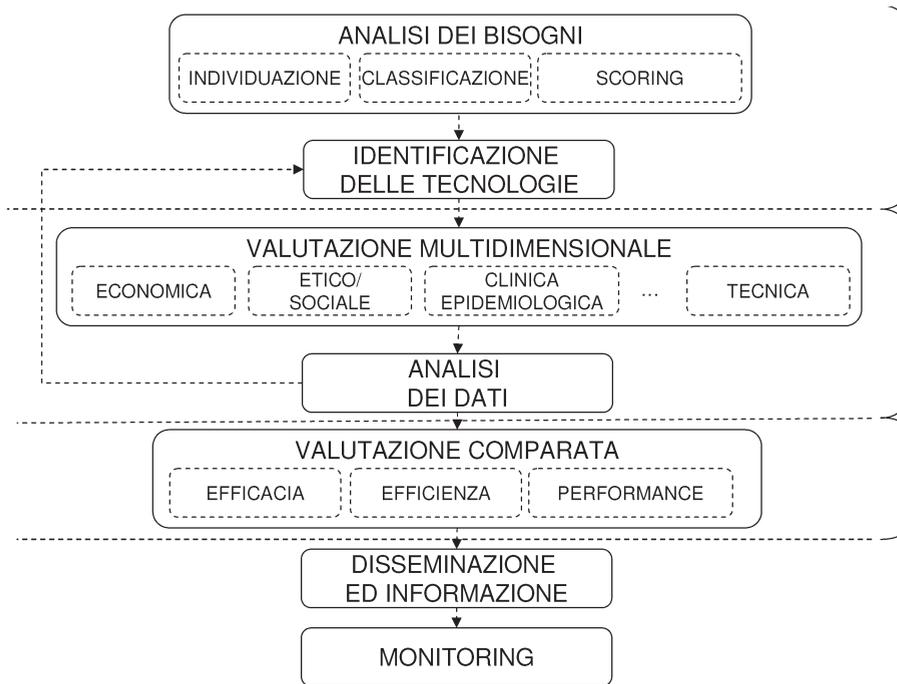


Fig. 2 – Il processo di HTA (Pecchia et al., 2009a).

nuova tecnologia arrivi sul mercato e, quando possibile, anche in fase di concezione, di ricerca e di sviluppo di una nuova tecnologia biomedica, al fine anche di massimizzarne l’impatto. Anche l’Organizzazione Mondiale della Sanità (OMS) sottolinea che la valutazione delle tecnologie è percepita come un ostacolo che rallenta l’introduzione di tecnologie innovative nei sistemi sanitari. Una delle ragioni individuate dall’OMS è proprio che il “problema” delle valutazioni viene considerato solo quando la tecnologia è prossima alla sua immissione nel mercato(OMS). L’importanza crescente dell’HTA per l’ingegneria biomedica è percepita con sempre maggiore chiarezza. Ad esempio, progetti Europei finalizzati all’armonizzazione ed alla standardizzazione dei percorsi formativi degli ingegneri biomedici<sup>2</sup> hanno per la prima volta nel 2010 incluso l’HTA tra gli insegnamenti fondamentali per l’ingegneria biomedica(Jarm et al., 2012, Pallikarakis et al., 2011). Su scala mondiale, poi, la Federazione Internazionale di Ingegneria Medica e Biomedica (IFMBE) ha creato la Divisione HTA, che oggi è al suo secondo mandato(IFMBE).

<sup>2</sup> CRH-BME “Curricula Reformation and Harmonisation in the field of Biomedical Engineering” Project Number: 144537-TEMPUS-2008-GR-JPCR (2008-4527).

## 2. HEALTH TECHNOLOGY ASSESSMENT: STANDARD DE FACTO

In base all'obiettivo della valutazione, alle risorse disponibili ed al tempo di cui si dispone, una valutazione di HTA può assumere varie formati. Tuttavia, i rapporti di HTA pubblicati sulle più rilevanti riviste scientifiche internazionali hanno una struttura molto regolare (NIHR), che definisce oramai uno standard de facto metodologico, e che rappresenta comunque un riferimento anche per valutazioni più sintetiche. Tale struttura può essere riassunta come segue:

1. definizione del problema medico: analisi dei bisogni ed identificazione delle tecnologie;
2. valutazione delle conseguenze sulla base della migliore evidenza clinica disponibile;
3. valutazione delle risorse necessarie attraverso l'analisi dei costi;
4. analisi comparata dei costi rispetto alle conseguenze.

In letteratura esistono diversi algoritmi che riassumono in maniera più o meno formale le fasi principali di una HTA. Nella figura 2 si riporta uno schema di principio sufficientemente generale e successivamente si descrivono sinteticamente le singole fasi, sebbene ognuna meriti un'approfondita trattazione individuale (Pecchia et al., 2009b, Pecchia et al., 2009a, Pecchia et al., 2009c).

### 2.1. Definizione del problema medico

Scenari diversi possono influenzare la definizione del problema medico, che spesso è determinato da contingenze, e cambia a seconda della dimensione del problema stesso. Pertanto non è facile individuare un metodo standard della scelta del problema medico.

#### 2.1.a. Analisi dei bisogni

Il primo passo per effettuare un HTA è definire formalmente il problema medico (in genere clinico o epidemiologico) che la tecnologia in esame deve soddisfare, mediante una approfondita analisi dei bisogni. In questo caso per analisi si intende: individuazione dei bisogni; la loro organizzazione in classi omogenee (ad esempio clinica, economica, tecnologica etc.); l'attribuzione ad ogni bisogno ed ad ogni classe di bisogni di un peso che ne rifletta l'importanza relativa. L'importanza relativa delle singole dimensioni e dei bisogni che le compongono dovrà poi essere quantificata. Quest'ultimo passaggio è molto delicato e deve essere massimamente trasparente e strutturato a priori, visto che dal peso relativo che si attribuisce alle singole dimensioni dell'analisi una tecnologia può risultare più o meno performante di altre. Inoltre, tale fase si basa su una parte di informazioni prevalentemente qualitativa che deve essere quantificata e resa massimamente oggettiva (al fine di consentire la riproducibilità della valutazione) prima di entrare in un processo decisionale analitico e scientifico, come un HTA (Pecchia et al., 2011b). Diversi metodi sono stati utilizzati in letteratura per la valutazione scientifica dei bisogni come ad esempio: *conjoint analysis* (CA) (Bridges et al., 2012), *discrete choice experi-*

*ments* (de Bekker-Grob et al., 2012), *best-worst scaling* (Gallego et al., 2011), e di recente lo *Analytic Hierarchy Process* (AHP) (Pecchia et al., 2011b). Al termine di questa fase, dovrebbero essere stati individuati i più importanti bisogni da soddisfare, le più rilevanti dimensioni dell'analisi (es.: clinica/epidemiologica, economica, tecnica etc.), e quantificata la loro relativa importanza.

### 2.1.b. *Identificazione delle tecnologie*

In questa fase si cerca di individuare tutte le tecnologie che possono contribuire a soddisfare i bisogni individuati. È buona norma che in questa fase non si consideri aprioristicamente informazioni parziali sulla performance delle stesse. La maggiore difficoltà è rappresentata dal fatto di includere tutte le tecnologie che concorrono a soddisfare i bisogni individuati. L'identificazione risente molto della scala della valutazione che può essere nazionale, regionale o aziendale ed in base alla scala si parla di macro-, meso- e micro-HTA. Ad esempio, una valutazione di un dispositivo medico a livello di Azienda o Presidio Ospedaliero potrà avere come presupposto l'acquisto delle nuove attrezzature. Lo stesso *assessment* potrebbe, a livello Regionale o Nazionale, avere come obiettivo l'indirizzo strategico della sanità sul territorio. Nel secondo caso competenze mediche più orientate alla ricerca gioverebbero maggiormente, si pensi alla necessità di individuare le tecnologie idonee mediante anche studi epidemiologici o ricerche avanzate, come una meta-analisi su database internazionali. L'accesso a riviste elettroniche di settore, la conoscenza di algebre booleane per l'utilizzo di motori di ricerca scientifici, o la sola padronanza della lingua inglese inficiano l'analisi stessa. Al termine di questa fase dovrebbero essere state individuate tutte le tecnologie, concorrenti ed alternative, che possono soddisfare tutti o parte dei bisogni precedentemente identificati, senza per ora farsi influenzare dalla importanza relativa degli stessi.

### 2.2. *Valutazione delle risorse e delle conseguenze*

Anche se molti autori suggeriscono di valutare i costi e le conseguenze (desiderate e non) di una tecnologia sanitaria in diverse dimensioni la maggior parte degli studi concentrano principalmente sulle dimensioni cliniche ed economiche. Questo può apparire come una limitazione, tuttavia queste due dimensioni sono sufficienti per la maggior parte degli studi HTA, e questi saranno più estensivamente trattate in questo capitolo.

#### 2.2.a. *La dimensione economica*

Insieme alla dimensione clinica, la dimensione economica è quella che maggiormente viene considerata in un HTA. Invero, molte pubblicazioni pur richiamando tra gli obiettivi quello di effettuare un HTA, effettuano valutazioni limitatamente a queste due sole dimensioni, le cosiddette valutazioni economiche. In questa fase dell'analisi bisogna analizzare i costi di tutte le risorse economiche impiegate

per l'erogazione della tecnologia *valutanda*. Senza entrare nel dettaglio delle metodiche specifiche, per le quali si rimanda a testi dedicati, lo scopo di questa fase è quello di attribuire un valore economico ad ogni risorsa impiegata direttamente o indirettamente nella produzione del servizio sanitario erogato. Per risorse si intendono la complessità di risorse umane, ambienti, tecnologie, forniture, ammortamenti, pulizie, assicurazioni etc.; per costi diretti si intendono i costi di quelle risorse dedicate esclusivamente alla tecnologia in esame; diversamente si intendono per costi indiretti i costi di quelle risorse che vengono impiegate solo in parte. Ad esempio, il personale dell'accettazione di un ospedale ricade solo in parte sui costi di sala operatoria, e dunque, in un'analisi molto ampia è da considerarsi come un costo indiretto. Ancora, è utile distinguere tra costi fissi e costi variabili, che rispettivamente si riferiscono ai costi legati con la quantità di bene/servizio erogato e quelli invece indipendenti da essa. Ad esempio, il costo di locazione è in genere indipendente dal numero di prestazioni erogate. Anche se in un ambulatorio non si effettuasse nemmeno una prestazione, bisognerebbe, a meno di situazioni particolari, continuare a pagarne l'affitto. Anche un'analisi economica si articola in almeno tre fasi: individuazione delle risorse necessarie, classificazione delle stesse, monetizzazione del loro valore. Ulteriori complicazioni in questa dimensione sono introdotti quando l'analisi effettuata copre periodi temporali lunghi, in cui il valore della moneta cambia, soprattutto se l'ambito dell'analisi è tale da dover convertire il valore in una moneta comune.

### 2.2.b. *La dimensione clinica/epidemiologica*

Questa è la fase centrale di un HTA. In questa fase si valutano gli effetti clinici/epidemiologici delle tecnologie mediante la misurazione dell'*outcome*. Naturalmente, l'*outcome* varia in considerazione di vari fattori, tra cui: la scala dell'*assessment* (aziendale, regionale, nazionale, internazionale), il tipo di tecnologia che si sta analizzando (apparecchiatura, farmaco, processo sanitario, tecnica chirurgica, ecc), la finalità della tecnologia (diagnosi, cura, terapia, prevenzione, screening), la tipologia di pazienti interessati (per gruppo nosologico, per età, per specializzazione medica, ecc.). È molto comune classificare uno studio proprio in base al metodo scelto per misurare l'effetto di una tecnologia. Esempi classici di misura degli effetti di una tecnologia sono: la misura dell'efficacia; la rilevazione dell'utilità; la stima dei benefici.

Per efficacia si intende la misura di quanto una tecnologia soddisfi uno specifico *outcome* clinico (es. riduzione della mortalità, del dolore, ecc). Il limite principale di questa tipologia di analisi è che essa misura gli effetti in maniera fondamentalmente monodimensionale. Si può misurare l'efficacia in più *outcome*, ma sostanzialmente non si fa altro che ripetere un'analisi monodimensionale. Rimarrà comunque la difficoltà di dover poi stabilire quale *outcome* si ritiene essere più importante soprattutto nel caso in cui non ci sia una tecnologia pienamente dominante, in altre parole una tecnologia che superi le altre in tutti gli *outcome*. Il vantaggio di tale tipologia di analisi è che l'*outcome* viene misurato direttamente ed in scala naturale, cioè non si stima un risultato mediante un effetto correlato e non si introducono scale complesse.

Per utilità si intende la misurazione indiretta dell'effetto di una tecnologia mediante una scala che in genere è il risultato di una funzione multidimensionale di più *outcome*. Ad esempio è molto comune misurare l'effetto di una tecnologia definendo il concetto complesso di qualità della vita, ovvero si stima l'efficacia mediante una funzione composta da scale funzionali di dimensioni indipendenti come: la sfera emotiva; sensoriale; cognitive; il livello di autosufficienza; l'assenza di dolore. Un esempio molto comune di funzione utilità è quello della QALY (*Quality Adjusted Life Years*), con cui si pesano gli anni di vita con un indice di qualità che assume valore 1 per un anno di vita in perfetta forma, e valori tra lo zero ed uno per anni di vita con una qualità minore. Tali valori vengono poi sommati per anno al fine di valutare il valore complessivo dei QALY nel periodo di osservazione. In pratica con queste metodiche si mettono sullo stesso piano due anni vissuti al 50% e un anno vissuto in perfetto stato di salute. Questo è una misura indiretta perché consente di stimare lo stato globale di salute del paziente, mediante metodi standard per valutare il valore che il paziente stesso dà al proprio stato di salute. I principali metodi sono:

1. il *time trade off*, in cui agli individui malati intervistati viene chiesto di scegliere tra il permanere nella condizione di malattia per un certo numero di anni o l'essere guariti e riportati a condizioni di buona salute, ma avendo un'aspettativa di vita minore;

2. lo *standard gamble*, in cui ai malati intervistati viene chiesto di scegliere tra il permanere nelle condizioni attuali o sopportare un intervento chirurgico che ha probabilità di farli morire o ripristinare uno stato di buona salute;

3. il *rating scale*, o *visual analogue scale*, in cui agli intervistati è chiesto di assegnare un valore tra 0 (morte) a 100 (salute) ai diversi stati di una malattia.

Il maggior limite di questa metodica è proprio nel modo in cui viene stabilito uno specifico stato. Tale valutazione è spesso soggettiva e difficilmente generalizzabile. Esistono anche altre metodiche più strutturate per l'autovalutazione dello stato di salute. Sfortunatamente, nessuna scala proposta finora ha raggiunto consenso unanime e dunque utilizzo universale. Questo è il maggiore limite delle analisi basate sull'utilità. Per superare tale limite si introduce il concetto di misurazione del beneficio.

Per beneficio si intende in genere la traduzione in termini economici dell'effetto di una tecnologia. È comunque comune usare il termine beneficio per esprimere il più generico concetto di conseguenza di una tecnologia, generando ovvie confusioni. In pratica si effettua una quantificazione monetaria del valore salute, diffusamente con il ricorso a due tecniche, che sono:

1. il metodo del capitale umano (*Human Capital*, HC);
2. la disponibilità a pagare (*Willingness To Pay*, WTP).

Nel metodo standard del capitale umano si presume che il valore per la società della vita di un individuo si misura dal futuro potenziale di produzione, calcolato come valore della retribuzione del lavoro previsto, riportato al valore presente. Questo metodo è implicitamente basato sulla massimizzazione della produzione futura individuale per la società attuale. Il principale problema di questo approccio è che riduce il valore della salute al solo contributo che la persona fornisce al Prodotto Interno Lordo (PIL), trascurando il valore intangibile della vita per sé. In un paese con un SSN pubblico e con una costituzione che ribadisce l'uguaglianza dei

cittadini, indipendentemente da quanto questi contribuiscano al PIL, che comunque è un indicatore discutibile della ricchezza di un paese, tale metodica è quantomeno inapplicabile.

Il metodo della disponibilità a pagare calcola, nel caso di una patologia, quanto la società sarebbe disposta a pagare per evitare di contrarre la malattia. In altre parole, si calcola il valore aggregato che una popolazione a rischio investirebbe sui programmi che salvano statisticamente delle vite, o la somma degli importi che singoli individui sono disposti a pagare ex-ante per comprare piccole riduzioni nella probabilità della loro morte. Un esempio può aiutare a illustrare questo punto. Si supponga che ogni persona in una popolazione di 100.000 abitanti, sia disposto a pagare 25 euro ( $25 \times 100.000 = 2,5$  milioni di euro) per un programma che si prevede riduca la probabilità complessiva di morte dallo 0,09% a 0,08%. Poiché questo è equivalente a una riduzione del tasso di mortalità da 90 per 100.000 a 80 per 100.000, il valore implicito per ciascuno delle 10 vite salvate è di 250.000 euro (2,5 milioni diviso 10 vite salvate). Naturalmente stabilire se una vita valga più o meno di questa cifra è un compito decisamente arduo.

### 2.2.c. *La dimensione etica e sociale*

La considerazione di queste dimensioni nei report di HTA è molto limitata (Sacchini et al., 2009). Per la loro trattazione si rimanda a testi specifici (Reiser, 1988). Tuttavia è utile richiamare lo schema di Heitman (Heitman, 1998) che propone una organizzazione in cinque categorie della dimensione etica:

1. aspetti relativi a concetti e definizioni del processo dei HTA stesso;
2. aspetti legati alle procedure diagnostiche (limiti, rischi, impiego, ecc);
3. aspetti legati alle strategie preventive (per esempio, la gestione del rischio di patologie) e terapie (evidence, efficacia, appropriatezza, ecc);
4. aspetti relativi alla ricerca (tutela dei soggetti arruolati negli studi, il consenso informato, ecc);
5. aspetti relativi all'allocazione delle risorse (giustizia distributiva, meccanismi di razionamento, le valutazioni economiche, ecc.).

### 2.2.d. *La dimensione tecnica*

Per quanto riguarda la dimensione tecnica, in essa ricadono quegli aspetti prevalentemente tecnologici che impattano solo parzialmente o indirettamente nella dimensione economica o clinica. Essa include, ad esempio, gli aspetti legati alla modalità di somministrazione di un farmaco che non impattino direttamente sul paziente (preparazione, conservazione etc.). Inoltre l'*asset* tecnologico di un'Azienda influisce sulla dimensione tecnica. Ad esempio, in un istituto di cura a vocazione scientifica, l'impatto di una tecnologia che richiede competenze più spinte è minore che in un Presidio Ospedaliero. È il caso, al momento, delle chirurgie robotiche per gli aspetti impiantistici e di *learning curve*. Infatti, per questo tipo di chirurgia la preparazione della sala chirurgica, sebbene talvolta non contemplata dalle pubblicazioni scientifiche, influenza indirettamente i costi della metodica e l'efficacia dell'intervento.

### 2.3. *Analisi dei dati: il contributo delle meta-analisi*

Individuate le dimensioni dell'analisi, le tecnologie da valutare e prima ancora i bisogni da soddisfare, si può passare all'analisi dei dati. Per analisi si intende: la raccolta dei dati, la loro elaborazione, la sintesi e la rappresentazione. Il metodo più usato per raccogliere i dati necessari ad un HTA è quello delle meta-analisi (Sutton, 2000). Questa metodica ha come scopo quello di aggregare risultati provenienti da studi condotti indipendentemente, da ricercatori di strutture diverse, con metodiche statistiche. L'obiettivo è quello di avere da più studi condotti su un numero contenuto di pazienti un risultato che sia una stima consistente dei dati potenzialmente ottenuti da un unico studio su tutti i pazienti coinvolti. Naturalmente l'analisi ha dei limiti, ma bisogna partire dal presupposto che non sarebbe altrimenti possibile condurre direttamente uno studio su grandi numeri di pazienti, per diverse ragioni, tra cui quelle economiche, ma anche per il numero stesso dei casi trattati dai singoli centri. L'individuazione degli studi viene di norma effettuata indipendente da più ricercatori con un metodo strutturato e che poi confrontano i loro risultati al fine di raggiungere il consenso su quali studi ammettere alla fase successiva dell'analisi, ovvero alla loro elaborazione. Il metodo di inclusione degli studi viene definito a priori, scegliendo appositi criteri, tra cui alcuni finalizzati alla scelta di studi scientifici di alto grado di evidenza. Si prediligono gli studi clinici controllati e randomizzati (RCT), poi quelli con pseudo randomizzazioni per poi arrivare agli studi di coorte (prospettivi, prospettivi con gruppo di controllo retrospettivo, retrospettivi) o agli studi retrospettivi (Guyatt et al., 1995). Individuati gli studi si procede poi a classificarli, pesarli, e ad estrapolarne da essi i risultati. Infine, i risultati dei singoli studi vengono aggregati pesandoli sulla base dell'importanza dello studio stesso. Tale importanza dipende dal numero di casi trattati e dalla precisione dei risultati. Tali risultati vengono poi rappresentati, solitamente in forma grafica, per essere comunicati al più ampio numero di decisori, che possono anche non essere estremamente confidenti con metodiche matematiche avanzate. Nella figura 3 si presentano risultati di una meta-analisi condotta su un farmaco biologico per valutarne l'efficacia sulla qualità della vita dei pazienti trattati rispetto ad un gruppo di controllo.

Questa meta-analisi è stata condotta su 5 studi (uno per riga) (Furst et al., 2003, van de Putte et al., 2003, Weinblatt et al., 2003, Keystone et al., 2004, van de Putte et al., 2004). In ogni studio sono stati osservati gli effetti su un gruppo di trattamento di numerosità  $N_t$  (da 67 a 419 per studio), tali effetti sono stati confrontati con un gruppo di controllo di  $N_c$  pazienti (da 62 a 314 per studio). La qualità della vita del paziente è stata misurata, nei singoli studi, con lo *Health Assessment Questionnaire* (HAQ), che è una scala per misurare la qualità della cura molto usata in studi su pazienti affetti da psoriasi, misurata prima e dopo il trattamento nei due gruppi (con placebo nel gruppo di controllo). I valori dell'HAQ, che indica un miglioramento dello stato di salute, è riportata nella seconda e nella quarta colonne, rispettivamente per il farmaco e per il placebo, in valore medio e deviazione standard in parentesi. Nella penultima colonna, viene riportato il peso relativo di ogni studio. Tale peso cresce con la popolosità dei due gruppi ( $N_c$  ed  $N_t$ ) e diminuisce con l'ampiezza della deviazione standard (SD) dell'effetto. Si nota, infatti, che lo studio di Keystone (quinta riga) condotto su 619 casi (419+200) e con una SD di

Study	Adalimumab N <sub>t</sub>	Adalimumab mean (SD)	Control N <sub>c</sub>	Control mean (SD)	WMD (fixed) (95% CI)	Weight (%)	WMD (fixed) (95% CI)
van de Putte, 2003	71	-0.45 (0.46)	70	-0.04 (0.37)	—	12.99	-0.41 (-0.55 to -0.27)
van de Putte, 2004	225	-0.38 (0.61)	110	-0.07 (0.49)	—	16.71	-0.31 (-0.43 to -0.19)
STAR	312	-0.51 (0.56)	314	-0.26 (0.46)	—	36.07	-0.25 (-0.33 to -0.17)
ARMADA	67	-0.62 (0.63)	62	-0.27 (0.57)	—	5.75	-0.35 (-0.56 to -0.14)
Keystone, 2004	419	-0.60 (0.56)	200	-0.25 (0.56)	—	27.68	-0.35 (-0.44 to -0.26)
Total (95% CI)	1094		756		◆	100.00	-0.31 (-0.36 to -0.26)
Test for heterogeneity: $\chi^2 = 4.90$ , $df = 4$ ( $p = 0.30$ ), $I^2 = 18.4\%$							
Test for overall effect: $z = 12.41$ ( $p < 0.00001$ )							

-1   -0.5   0   0.5   1  
Favours   Favours  
adalimumab   control

Fig. 3 – Differenza media dell'effetto dell'adalimumab sull'HAQ nel gruppo di controllo e nel gruppo trattato. Nelle prime cinque righe, sono riportati i risultati di altrettanti studi. Il risultato finale è riportato nell'ultima riga.

0,56 sia nel gruppo di controllo che nel gruppo trattato, ha un peso relativo, normalizzato a cento di 24,30, che è inferiore al peso relativo dello studio STAR, condotto su un numero maggiore di casi e con una SD minore, almeno nel gruppo di controllo. Infatti, lo STAR ha coinvolto 626 (312+314) pazienti e ha riportato una SD di 0,56 nel gruppo trattato, ma di 0,48 nel controllo. Infine, nell'ultima colonna si riporta, per ogni studio, la differenza media tra gruppi (controllo meno trattato) e tra parentesi l'intervallo di confidenza al 95%. Nell'ultima riga, si riporta la Differenza Media Pesata (*Weighted Mean Difference*, WMD), in altre parole la differenza che si stima si sarebbe ottenuta se fosse stato condotto un unico studio su 1094 pazienti trattati con il farmaco e 756 pazienti con placebo. Simili statistiche esistono per variabili dicotomiche, per le quali si rimanda a testi specifici (Sutton, 2000). Esempi di meta-analisi sono abbondanti in letteratura. Si veda ad esempio i due lavori di Bracale et al. (Bracale et al., 2012b, Bracale et al., 2012a), accessibili liberamente in internet<sup>3</sup>. Il secondo dei due studi citati è una network meta-analisi, che rappresenta una delle possibili generalizzazioni delle meta-analisi (Lumley, 2002).

#### 2.4. La valutazione comparata dei costi e delle conseguenze

Nella pratica non capita mai che una nuova tecnologia supera le altre esistenti in tutte le dimensioni dell'analisi. Si dovrà dunque valutare, in coerenza con l'obiettivo dell'analisi, quali dimensioni sono prioritarie. Tale decisione può dipendere fortemente dalla scala dell'analisi stessa. Ad esempio, per un'Azienda Sani-

<sup>3</sup> ePrint University of Nottingham: <http://eprints.nottingham.ac.uk/view/people/Pecchia=3ALeandro=3A=3A.html>

Tab. 1 – Confronto tra le diverse metodologie di analisi costo/conseguenza

	AMC	ACE	ACU	ACB
<i>Risorse</i>	unità monetarie	unità monetarie	unità monetarie	unità monetarie
<i>Conseguenze</i>	assunte uguali	unità fisiche	anni di vita pesati	unità monetarie
<i>Misurando</i>	$\Delta C$	RICE	RICU	RICB
<i>Vantaggi</i>	misura diretta	misura diretta	analisi	analisi
	risultato univoco	misura i risultati	multidimensionale	multidimensionale
	necessarie per le altre	clinici purché	stima indirettamente	stima i risultati (non
		omogenei	i risultati clinici non	solo clinici) non
		risultato univoco	omogenei	omogenei purché
			introducendo il	monetizzabili
			concetto di qualità	
			della vita	
			risultati non univoci	
<i>limiti</i>	non misura gli effetti	analisi	misura indiretta	misura indiretta
	presuppone	monodimensionale	solo aspetti clinici	difficile monetizzare
	omogeneità dei	non esistono scale	non esistono scale	benefici che
	risultati clinici	universali per il RICE	universali per il RICU	riguardano la persona
		presuppone		limiti etici
		omogeneità dei		
		risultati clinici		

taria Locale si potrà ritenere prioritaria la massimizzazione dell'efficienza, mentre in un'Azienda Ospedaliera di elevata specializzazione si potrà ritenere prioritaria l'efficacia, ed in particolare alcuni *outcome*. Come discusso nei paragrafi precedenti, la maggior parte degli studi considera la dimensione economica e quella clinica valutata in termini di efficacia, utilità o beneficio. In letteratura (Nord, 1999) ricorrono quattro metodiche che consentono di analizzare l'efficienza di una tecnologia sanitaria: Analisi di Minimizzazione dei Costi (AMC), Analisi Costo-Efficacia (ACE), Analisi Costo Utilità (ACU), Analisi Costo-Beneficio (ACB). Nella tabella 1 si riassumono le principali caratteristiche, vantaggi e limiti di queste quattro metodiche.

Per scegliere il tipo di analisi più idonea si può seguire l'algoritmo descritto in figura 4.

Valutando i costi e le conseguenze di una nuova tecnologia rispetto ad una ampiamente usata, si possono presentare tre condizioni, come riassunto nella tabella 2 (in cui si fa riferimento all'efficacia come conseguenza ma il discorso è analogo per l'utilità).

[Tab. 2. *Possibili risultati di un analisi costo-efficacia*]

Se una nuova tecnologia (tecnologia A) si dimostra essere meno efficace e più costoso rispetto ad un benchmark (gold standard – tecnologia B), come rappresentato nel primo caso della tabella 2, essa è respinta senza ulteriori analisi, a meno che non vi sia un elevato grado di incertezza nei dati (in tal caso sono necessari ulteriori esperimenti). Al contrario, una tecnologia è un buon candidato per l'adozione, se più efficace e meno costoso (secondo caso della tabella 2). Il terzo caso, in cui il conseguimento di una maggiore efficacia è legato ad un co-

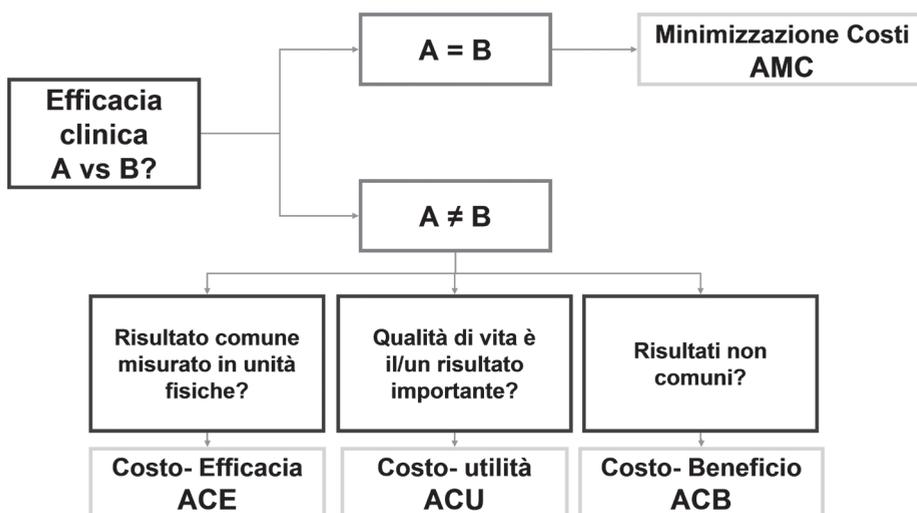


Fig. 4 – Algoritmo per la scelta della Analisi Costo-Conseguenza da effettuare.

Tab. 2 – Possibili risultati di un analisi costo-efficacia

Confronto tra costi	Confronto tra efficacia	Caso	
$C_A \geq C_B$	$E_A < E_B$	Dominanza	A è più costosa e meno efficace di B. Si continua ad adoperare B
$C_A \leq C_B$	$E_A > E_B$	Cost-Saving	A è meno costosa e più efficace di B. Si introduce la tecnologia A, superando progressivamente B
$C_A \geq C_B$	$E_A > E_B$	Rapporto Incrementale	A è più costosa, ma anche più efficace. Bisogna ricorrere al rapporto incrementale costo-efficacia o costo-utilità

sto aggiuntivo, richiede la valutazione del Rapporto Incrementale Costo-Efficacia (o Utilità) (rispettivamente ICER o ICUR) che esprime quanto costa migliorare l'*outcome* di un'unità. Questa è una situazione frequente principalmente perché una nuova tecnologia prevede costi di progettazione e prototipazione che con produzioni a basso volume risultano più elevati rispetto ad un prodotto già in commercio. Un ICUR (o ICER) sotto una certa soglia può suggerire che la nuova tecnologia è più conveniente, e quindi la sua adozione sia giustificata in quanto i risultati attesi sono ritenute meritevoli di essere i costi sostenuti. In Italia questo rappresenta un punto critico per l'assenza di tabelle di riferimento standard, mentre in UK una soglia normalmente adoperata per i farmaci è quella dei 30.000,00 sterline per QALY. Questo equivale a dire che si ritiene plausibile spendere 30.000,00 sterline per 1 anno di vita completamente guadagnato o equivalentemente per 2 anni di vita vissuti a +50% rispetto al non intervento o intervenendo con un benchmark.

Tab. 3 – Confronto tra farmaci e dispositivi medici

Dispositivo Medico	Farmaco
<b>Ciclo di vita del prodotto</b>	
Breve e con parti in continua evoluzione	Lungo e le componenti le molecole principali rimangono le stesse per anni
<b>Valutazione Clinica</b>	
Difficile realizzazione di uno studio in cieco (non esiste placebo)	Semplice realizzazione di uno studio in cieco usando il placebo
Diversi utenti	Un solo utente finale
Lunga learning curve	Breve learning curve
Efficacia fortemente dipendenti dagli ambienti	Efficacia poco dipendenti dagli ambienti
Difficile standardizzare RCT multicentrici	Facile standardizzare un RCT
<b>Utente</b>	
Efficacia fortemente dipendenti dagli utenti	Efficacia poco dipendenti dagli utenti
Richiede training intensive e continuo	Non richiede apprendimento specifico
Le complicazioni diminuiscono con l'uso	Complicazioni crescono con l'uso frequente
<b>Mercato</b>	
Principalmente piccole/medie imprese	Poche piccole/medie imprese
Poche multinazionali	Prevalentemente multinazionali
Diagnostico, terapeutico e riabilitazione	Terapeutico
<b>Costi</b>	
Overheads significativi	Overheads significativi
Ritorni lenti	Ritorni più veloci
Costi di distribuzione elevati	Costi di distribuzione minori
Costi di manutenzione e installazione	Nessun costo di manutenzione o installazione

### 3. PRINCIPALI LIMITI E CRITICITÀ DELLE METODICHE STANDARD

La maggior parte delle metodiche di HTA sono state concepite e sviluppate per la valutazione di farmaci, piuttosto che di dispositivi medici. Pertanto, i principali metodi impiegati sono adatti a tale scopo. Tuttavia, ci sono alcune differenze tra farmaci e dispositivi medici, che hanno un impatto notevole sulla valutazione da fare (tabella 3).

Inoltre, i metodi di HTA in realtà non sono capaci di dare informazioni sufficienti in fase di ricerca e sviluppo. Ad esempio, in fase di ricerca non esistono RCT sufficientemente ampi o numerosi per consentire una adeguata valutazione delle conseguenze. Invece, è importante avere delle informazioni preliminari in fase molto anticipata di ricerca, preferibilmente in fase di concepimento dell'idea, con una stima affidabile della probabilità di ritorno sugli investimenti o sulle esigenze del mercato specifico cui la tecnologia mira (Ijzerman and Steuten, 2011). Infine, i metodi standard di HTA non consentono di identificare le priorità tra i bisogni dei singoli utenti o di sottogruppi significativi. Ad esempio, i QALY pesano allo stesso modo le diverse dimensioni per qualunque tipologia di paziente. Questo è come dire che la motilità ha lo stesso valore relativo rispetto al dolore provato e che tale valore rimane costante in un anziano o in un giovane atleta.

#### 4. EARLY STAGE HTA (EHTA)

Studi recenti provano a superare alcuni dei limiti descritti nel paragrafo 3 in particolare: nell'ipotesi che dispositivo su cui si stanno effettuando le ricerche funzionerà esattamente come dimostrato dai dati sperimentali (magari derivati per ora da simulazioni al computer), quale beneficio potrà dare al paziente medio rispetto alla tecnologia attualmente considerata di riferimento (e dunque rimborsata dal SSN)? Questa è esattamente la domanda cui l'early stage HTA prova a dare una risposta in maniera scientifica. In questo paragrafo si introducono alcuni di tali metodi.

##### 4.1. *Headroom Analysis*

Affinché una risposta alla domanda appena posta sia realmente utile, essa dovrebbe essere quantificata ed espressa in termini di QALY e successivamente tale quantificazione deve essere confrontata con una soglia di riferimento per essere giudicata costo-efficace o meno, come illustrato in precedenza. Il SSN Inglese, considera in genere appropriata una tariffa di £ 30.000 per QALY. Questo è dunque l'importo aggiuntivo per ogni paziente che l'NHS è disposto a pagare per la nuova tecnologia. Un metodo molto semplificato che prova a dare una risposta a tale domanda è il metodo Headroom (Cosh et al., 2007), che valuta se una tecnologia, nella migliore delle ipotesi possibili, risulti costo-efficace mediante il metodo dell'iCUR. Poiché il calcolo è effettuato sotto le ipotesi più favorevoli, i risultati non danno la garanzia che il dispositivo avrà successo, ma può essere utilizzato per escludere idee che non hanno alcuna possibilità realistica di successo. Il metodo consiste nei seguenti passi (figura 5):

1. si assume che il dispositivo apporti realmente il miglioramento  $\Delta$ QALY stimato (Punto 1);
2. si calcola il costo massimo che l'NHS è disposto a spendere per un tale  $\Delta$ QALY (Punto 2);
3. si sottraggono i costi di produzione del dispositivo (Punto 3);
4. si sottraggono i costi per l'NHS che il dispositivo non riduce (es. somministrazioni);
5. si calcola il margine di guadagno massimo per il prodotto.

Se tale guadagno rappresenta un margine sufficiente, si può decidere di continuare le ricerche. Un margine che in tali vantaggiose ipotesi non risulti utile ad una azienda, potrebbe stimolare ulteriori ricerche (una maggiore efficacia) o far decidere di investire altrove il proprio budget. Tuttavia avere tali informazioni in fase molto preliminare, è molto utile nella ricerca applicativa.

##### 4.2. *Valutazioni costo-utilità in fase preliminare: i modelli di Markov*

Come introdotto nel paragrafo 2.2.a, un'analisi dei costi consiste nel dare un valore monetario a ciascuna risorsa utilizzata nel processo di cura. Questa analisi

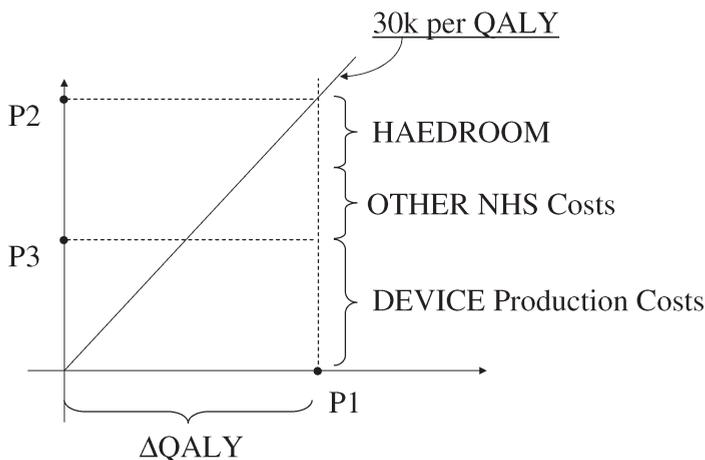


Fig. 5 – Metodo Haedroom.

può diventare più complessa quando le risorse necessarie per ogni paziente possono variare in base all'evoluzione della patologia (ad esempio un peggioramento può richiedere risorse aggiuntive). Infatti, pazienti all'interno della stessa popolazione in esame potrebbero essere in diverse fasi iniziali della patologia (es. lieve, moderata e grave) all'inizio dello studio. Successivamente ogni stato potrebbe subire variazioni: la probabilità di avere esacerbazione, la qualità di esacerbazione (lieve, moderata e grave), la probabilità di passare da uno stato all'altro (ad es. da lieve divenire moderato) nel successivo intervallo temporale. Un tale fenomeno può essere modellato utilizzando modelli di *Markov* (Sonnenberg and Beck, 1993). Questi modelli assumono che un paziente è sempre riconducibile ad uno solo di un numero finito di stati discreti di salute modellati, chiamati stati di Markov (nodi nel modello). Tutti gli eventi sono rappresentati come transizioni da uno stato all'altro (*edge*), con una data probabilità ' $p$ '. Un costo marginale ' $\delta C$ ' è associato ad ogni evento (transizione) ' $k$ ', rappresentando la quantità totale di risorse necessarie per tale evento. Poiché anche alcune conseguenze (efficacia, utilità o benefici) possono variare in un arco temporale, i modelli di Markov possono essere usati per stimare un'analisi incrementale costi-conseguenze (Sonnenberg and Beck, 1993) in popolazioni di pazienti che si suppone in evoluzione.

In questo caso tre fasi sono di solito eseguite:

1. ciascuno stato ' $K$ ' di una patologia (compreso quello iniziale) è associato un costo ( $\delta C_k$ ), che rappresenta la quantità totale di risorse richieste dai pazienti che si trovano in tale stato, e una conseguenza marginale, che può essere una variazione dell'efficacia ( $\delta E_k$ ) o di utilità ( $\delta U_k$ );
2. il passaggio da uno stato  $A$  ad ogni altro  $i$  adiacente (verso cui si può progredire) è associata una probabilità  $p_{Ai}$ ;
3. gli effetti avversi (es. complicazioni) possibili da ogni stato sono modellati come un'uscita del sistema e dunque associati anche essi ad una probabilità  $p_e$  (che può variare per tipologia: lieve o grave) e ad un costo per il trattamento.

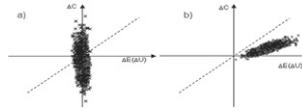


Fig. 6 – Esempio didattico: a) equivalenza tra l'efficacia di due tecnologie (simmetria rispetto all'asse verticale); b) trattamento in esame costo-efficace rispetto al benchmark perché tutti i punti della simulazione sono sotto alla linea di willingness to pay.

4. il sistema così descritto si fa evolvere per un tempo predefinito ed il costo totale e l'efficacia totale (o utilità) sono combinati per calcolare i costi totali (C) e le conseguenze totali (E o U) per i gruppi di pazienti in ogni stato finale.

5. Il modello è quindi valutato dinamicamente (spesso con una simulazione Monte Carlo).

Questo genera una distribuzione di punti Utilità-Costo (U, C)<sub>j</sub>, con  $j = 1 \dots N$  (numero di simulazioni effettuate, come illustrato in figura 6).

Nel caso di un confronto di una nuova tecnologia (A) con un benchmark di riferimento (B), questa procedura si ripete con lo stesso modello di Markov, ma con probabilità, costi e conseguenze derivanti dall'uso delle nuove tecnologie per il calcolo dei costi e delle conseguenze tra le due tecnologie ( $\Delta = A-B$ ). In questo caso, il piano costo-conseguenza ha come asse x le differenze di conseguenze ( $\Delta E$  o  $\Delta U$ ) e l'asse y delle differenze dei costi ( $\Delta C$ ) (Chapman et al., 2000).

Nello sviluppo di un nuovo dispositivo o di un'innovazione tecnologica può essere difficile ottenere dati che consentano di stratificare i pazienti in base al rischio e agli interventi passati o di considerare l'intero range di *outcome* che si possono verificare. Tuttavia, può essere sufficiente in una fase preliminare considerare gli *outcome* principali e limitare la stratificazione dei pazienti a pochi gruppi. Uno degli approcci analitici adoperati si basa sull'uso dei modelli di Markov con un numero ridotto di stati e transizioni in relazione alla disponibilità dei dati. Questo approccio è stato applicato da Dong e Buxton (Dong and Buxton, 2006) per confrontare una tecnologia innovativa (*computer-assisted*) e la procedura standard di sostituzione totale del ginocchio (TKR). Il numero di strati di pazienti che hanno avuto una TKR primaria in base al rischio è limitato a tre gruppi: assenza di complicazioni, complicazioni minori, complicazioni serie. I pazienti con complicazioni possono richiedere un re-intervento o altri trattamenti. Le probabilità di transizione tra gli stati, le varie complicazioni addizionali o la morte da inserire nel modello furono ricavate dalla letteratura o stimate ad hoc. Il confronto fu effettuato con un modello a 9 stati dimostrando che la valutazione early stage offriva dati sufficienti a molto prima che ambi RCT e dunque meta-analisi fossero disponibili in letteratura. In uno studio successivo (Craven and Morgan, 2011) un modello a 4 stati è sta-

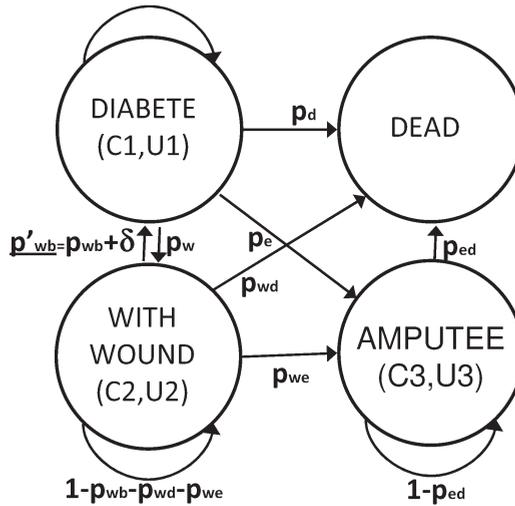


Fig. 7 – Modello di valutazione early stage via Markov Model.

to usato per un'analisi *what-if* di un dispositivo per promuovere la guarigione delle ulcere nel piede diabetico. Il modello includeva quattro stati: assenza di ferita, ferita (ulcera), amputazione, e morte. Questa è una semplificazione di un modello più sofisticato che prevedeva una stratificazione in alto e basso rischio e considerava più tipi ulcerazioni e amputazioni con probabilità diverse. Il valore del dispositivo è stato esaminato assumendo che la probabilità di transizione tra 2 stati (ritorno da ulcerato ad assenza di ferita) fosse migliorata dall'utilizzo del dispositivo in esame, come illustrato in figura 7.

#### 4.3. AHP per l'analisi dei bisogni e la valutazione comparata

Come introdotto nel paragrafo 2.1.a, diversi metodi sono stati proposti per l'analisi dei bisogni per l'HTA, tra cui *conjoint analysis* (CA) (Bridges et al., 2012), *discrete choice experiments* (de Bekker-Grob et al., 2012), *best-worst scaling* (Gallego et al., 2011), e di recente lo *Analytic Hierarchy Process* (AHP) (Pecchia et al., 2011b). Secondo Scholl et al. (Scholl et al., 2005), il metodo AHP si è dimostrato più adatto rispetto a CA per decisioni complesse che coinvolgono molti fattori. Mulye (Mulye, 1998) ha suggerito che AHP è più efficace di CA quando più di 6 attributi sono considerati. Ijzerman et al. (Ijzerman et al., 2008) ha concluso che AHP, confrontato con CA, risulta in questionari più facili da implementare, più brevi e flessibili, sebbene incongruenze possano generarsi e altri metodi possano avvantaggiarsi di un approccio più olistico. In questo paragrafo si descrive brevemente tale metodo e la sua applicazione per eHTA.

AHP è una metodo di *decision-making* per la risoluzione di problemi multifattoriali e multidimensionali. Questo metodo è particolarmente efficace nel quan-

tificare le opinioni degli operatori in base alle loro esperienze personali e di modellare un quadro coerente decisione. Esso prevede la definizione di una gerarchia di elementi da classificare presentando questionari in cui ogni esperto, attraverso confronti a coppia, esprime un giudizio circa l'importanza relativa degli elementi. Ponendo delle domande ridondanti è possibile valutare la coerenza del rispondente. Applicando questo metodo iterativo è possibile stimare: l'importanza relativa di ogni bisogno nella sua categoria, l'importanza relativa di ciascuna categoria, l'importanza assoluta di ogni bisogno rispetto a tutti gli altri. Ulteriori dettagli sul AHP si possono trovare nei riferimenti in bibliografia (Pecchia et al., 2009c, Pecchia et al., 2009a, Pecchia et al., 2011b). In questo capitolo, ci limiteremo a descrivere sinteticamente le quattro fasi seguenti:

1. Identificazione dei bisogni che i nuovi dispositivi biomedici mirano a soddisfare.
2. Progettazione di un albero di bisogni con nodi (categorie) e foglie (bisogni).
3. Sviluppo e presentazione di un questionario per rilevare le esigenze degli utenti.
4. Stima numerica dei pesi.
5. Valutazione comparata di più tecnologie

L'identificazione dei bisogni (1° Fase) può essere condotta opportunamente mediante un *focus group*, che coinvolga esperti del dominio (ad esempio, medici specialisti, ingegneri biomedici o clinici), potenziali utilizzatori del dispositivo (personale medico, pazienti) ed 1 facilitatore, esperto di AHP.

Nella seconda fase, i bisogni precedentemente individuati vengono riorganizzati in una struttura ad albero. In genere, l'esperto di AHP assolve il ruolo di facilitatore, proponendo un'ipotesi di albero, da rivedere nel *focus group* al fine di verificarne la consistenza e completezza. Tre categorie principali sono incluse nella maggior parte dei modelli di albero e sono: bisogni clinici, bisogni economici, bisogni tecnici.

La terza fase si basa sullo sviluppo di questionari che permettano a ciascun rispondente di confrontare l'importanza relativa di ciascun bisogno rispetto agli altri della stessa categoria. Per ciascuna coppia di bisogni ( $i, j$ ), ad ogni rispondente è posta una domanda di questo tipo: «In base alla sua esperienza, considerato l'obiettivo specifico della valutazione, come considera il bisogno  $i$  rispetto al bisogno  $j$ ?». L'esperto risponde scegliendo uno tra i giudizi proposti. Al riguardo, esistono diverse scale che definiscono i giudizi e attribuiscono a ciascun giudizio un valore numerico. Nel metodo AHP è molto diffusa la scala naturale proposta dallo stesso inventore del metodo Prof. Saaty (Saaty, 1977), in cui un valore numerico è attribuito ai giudizi positivi, ad es.: 3 per «più importante», 5 per «molto più importante». Il valore 1 è attribuito al giudizio «ugualmente importante». I valori reciproci sono attribuiti ai giudizi rimanenti:  $1/3$  per «meno importante»,  $1/5$  se «molto meno importante». La scala naturale di Saaty prevede nove possibili giudizi, (1- ugualmente importante, 9-estremamente più importante), tuttavia, sono ampiamente utilizzate varianti con un numero minore (sempre dispari) di giudizi, perché agevola la compilazione del questionario da parte di rispondenti che non hanno familiarità con il metodo. Inoltre, è stato constatato che la gran parte dei rispondenti utilizza non più di tre giudizi nel confrontare fino a quattro elementi e che mediante procedure matematiche (Saaty and Vargas, 2001) i risultati ottenuti

con una scala a 5 giudizi sono equivalenti a quelli ottenuti usando la scala naturale di Saaty (a 9 giudizi).

Il processo è ripetuto per ciascuna categoria e il questionario deve essere sviluppato per ridurre al minimo il *bias*. Ad esempio, se il rispondente scrive da sinistra a destra e dall'alto al basso, è inconsciamente portato a giudicare gli elementi in alto a sinistra come più importanti di quelli in basso a destra, per questo è opportuno che ciascun elemento sia presentato lo stesso numero di volte a destra e a sinistra, in alto e in basso.

La quarta fase prevede una serie di calcoli numerici. Per ogni categoria, viene popolata una matrice numerica dei giudizi  $A_{n \times n}$ , dove  $n$  è il numero dei bisogni nella categoria stessa ed in cui:

1. il generico elemento ( $a_{ij}$ ) rappresenta il rapporto tra l'importanza relativa del bisogno  $i$  ( $N_i$ ) rispetto a quella del bisogno  $j$  ( $N_j$ );
2. l'elemento  $a_{ji}$  è il reciproco dell'elemento  $a_{ij}$ , assumendo la reciprocità dei giudizi (se  $N_i$  è 3 volte più importante di  $N_j$ , allora  $N_j$  dovrebbe essere 1/3 di  $N_i$ );
3. gli elementi sulla diagonale principale ( $a_{ii}$ ) sono uguali ad 1 ( $N_i$  è di uguale importanza rispetto a se stesso);
4. la matrice  $A$  è assunta come transitiva, ovvero

$$a_{ij} = a_{ik} * a_{kj} \quad \text{per } i, j, k \text{ qualsiasi } (1 \leq i, j, k \leq n)$$

Che segue dalla definizione di  $a_{ij}$ :

$$a_{ij} = \frac{N_i}{N_j} = \frac{N_i}{N_k} * \frac{N_k}{N_j} = a_{ik} * a_{kj} \quad (4)$$

La proprietà transitiva riflette l'idea che se il bisogno  $i$  è considerato due volte più importante di  $j$  ( $N_i = a_{ij} * N_j$ ), e  $j$  è considerato tre volte più importante di  $k$  ( $N_j = a_{jk} * N_k$ ), allora  $i$  dovrebbe essere giudicato 6 volte ( $2 \times 3$ ) più importante di  $k$  ( $N_i = a_{ik} * N_k$ , con  $a_{ik} = a_{ij} * a_{jk}$ ). È facilmente dimostrabile (Saaty, 1977) che in una matrice con le predette proprietà ogni colonna è proporzionale alle altre e un solo autovalore reale ( $\lambda$ ) esiste, che è uguale a  $n$ . L'autovettore associato è ancora proporzionale ad ogni colonna e, normalizzato, rappresenta l'importanza relativa di ciascun bisogno confrontato con ciascuno degli altri nella stessa categoria. L'importanza relativa del bisogno  $i$  all'interno della categoria  $m$  viene in genere indicato in letteratura come *Local Weight* ( $LW_i^m$ ). Può capitare che i giudizi non siano pienamente consistenti. Ad esempio un esperto può giudicare il bisogno  $i$  due volte più importante di  $j$  ( $N_i = a_{ij} * N_j$ ), e  $j$  tre volte più importante di  $k$  ( $N_j = a_{jk} * N_k$ ), e poi giudica  $i$  cinque (non sei) volte più importante di  $k$  ( $N_i = a_{ik} * N_k$ , con  $a_{ik} < a_{ij} * a_{jk}$ ). Questo esempio rientra in un livello di incoerenza normalmente considerato tollerabile in letteratura. In parte perché i giudizi umani non sono completamente consistenti ed in parte perché la scala naturale limitata non consente sempre il rispetto della proprietà transitiva. Diverso sarebbe stato se il rispondente avesse giudicato  $i$  meno importante di  $k$ , in palese contraddizione con i primi 2 giudizi dati. Nei casi, in cui i giudizi non sono pienamente consistenti, le colonne della matrice non sono proporzionali e la matrice ha più autovettori e nessuno è proporzionale a tutte le colonne. In questo caso, l'autovettore principale, che è quello corrispon-

dente all'autovalore massimo in modulo ( $\lambda_{\max}$ ) viene scelto e le sue componenti normalizzate rappresentano la relativa importanza di ogni bisogno. Tale procedura, si dimostra valida fino ad un certo limite di incoerenza, che è ben definito in letteratura (Saaty and Vargas, 2001). L'inconsistenza viene sistematicamente stimata ponendo domande ridondanti. Ad esempio, considerando tre bisogni ( $i, j, k$ ), vengono poste le domande per il confronto  $i-j$  a  $j-k$ , e anche per il confronto ridondante  $i-k$ . La risposta all'ultima domanda viene confrontata con quella attesa dalle prime due domande in base alla proprietà transitiva. La differenza tra la risposta reale e quella attesa è un indice di inconsistenza. Più formalmente, l'inconsistenza è stimata misurando la differenza tra l'autovalore massimo e  $n$ , normalizzata ad  $n$  stesso. L'inconsistenza può essere dovuta in molti casi a distrazione o perdita di interesse, e richiedere che il questionario sia somministrato nuovamente. Tuttavia, una certa inconsistenza è attesa, tenendo conto dei limiti della scala usata e del giudizio umano e per questo in letteratura sono stati proposti valori di soglia (Pecchia et al., 2010) e scale alternative (Saaty, 1982).

Applicando lo stesso algoritmo iterativamente alle categorie di bisogni si possono valutare le importanze relative di ciascuna categoria. L'importanza relativa di ogni categoria è, in genere, indicata in letteratura come *Categorical Weight* ( $CW^m$ ).

Infine, l'importanza relativa del bisogno  $i$  rispetto agli altri (anche quelli non appartenenti alla stessa categoria) è indicata come *global-importance* o *Global-Weight* ( $GW_i$ ) ed è calcolato moltiplicando il *Local Weight* per il *Categorical Weight*, come riportato nella formula seguente:

$$GW_i = LW_i^k * CW^k \quad (5)$$

Mettendo insieme tali pesi, si ottiene il vettore dei pesi globali ( $\underline{GW}$ ), un vettore di dimensione  $m$ , dove  $m$  è proprio il numero di bisogni individuati, che esprime l'importanza di tutti i bisogni. Uno dei maggiori vantaggi di tale metodo è proprio quello di misurare l'importanza relativa di due bisogni appartenenti alla stessa categoria, e consentire poi l'estensione del confronto a bisogno completamente diversi. Ad esempio, tale metodo consente di stimare mediante questionari l'importanza relativa di diversi outcome clinici (es.: mobilità verso dolore) e poi mettere a confronto uno di questi outcome clinici con un bisogno appartenente ad una categoria non clinica (es.: ergonomia del dispositivo). Questo metodo è stato applicato con successo per rilevare i bisogni degli utenti in studi precedenti: per l'identificazione del miglior modello di cura per lo scompenso cardiaco cronico (Pecchia et al., 2009a), per la scelta di un contratto di manutenzione in base alla localizzazione dell'ospedale ed alle prestazioni offerte (Pecchia et al., 2009c), per l'identificazione dei fattori di rischio per le cadute negli anziani (Pecchia et al., 2011b), per l'acquisto di un tomografo computerizzato (Pecchia et al., Submitted 2012), la definizione di una scala della qualità della vita specifica per il benessere dell'anziano (Pecchia et al., 2011a).

Infine, si può applicare lo stesso metodo descritto per confrontare il dispositivo in fase di sviluppo con le tecnologie esistenti. A tal fine, la domanda da porsi è: «quanto la tecnologia A è efficace rispetto alla tecnologia B nel soddisfare il bisogno  $i$ ?». Usando gli stessi pesi descritti in precedenza, si otterrà un vettore  $\underline{E}$  (di  $m$  elementi, dove  $m$  è il numero complessivo dei bisogni individuati) che descrive

l'efficacia della tecnologia in fase di sviluppo, rispetto a tutte le altre. Calcolando il prodotto scalare di tale vettore per il vettore GW introdotto in precedenza, si otterrà un giudizio sintetico che rappresenta la performance globale del dispositivo in esame rispetto alle altre tecnologie esistenti. Esempi di tale applicazioni possono trovarsi in letteratura applicati a tecnologie in varie fasi di sviluppo (Pecchia et al., 2007, Pecchia et al., 2009c, Liberatore and Nydick, 2008, Sloane et al., 2003, Ijzerman and Steuten, 2011).

## 5. RINGRAZIAMENTI

LP e MC ringraziano i partner del Programma MATCH (EPSRC Grant EP/F063822/1), nell'ambito dei quali sono stati sviluppati i metodi ed i tool presentati nel paragrafo 4.

## 6. CONCLUSIONE

In questo capitolo, dopo aver brevemente richiamato alcuni concetti chiave come quello di efficacia, efficienza e performance in sanità, sono state introdotte alcune metodiche di HTA che definiscono uno standard de facto in quanto rappresentano un modello di riferimento per le pubblicazioni su riviste scientifiche di settore. Tali metodiche sono molto utili e la loro diffusione capillare è quanto mai auspicabile. Tuttavia esse sono state storicamente sviluppate per valutare farmaci e presentano dei limiti di applicabilità quando le si utilizza per valutare nuove tecnologie biomedicali, in particolare se si vuole stimare la loro potenziale sostenibilità in fase di sviluppo, ricerca o addirittura di concepimento (*early stage*).

Negli ultimi anni sono stati sviluppati dei metodi di valutazione concepiti proprio per l'*early stage* HTA di tecnologie biomediche, che provano a dare una risposta alla seguente domanda: «nell'ipotesi che dispositivo su cui si stanno effettuando le ricerche funzionerà esattamente come dimostrato dai dati sperimentali, quale beneficio potrà dare al paziente rispetto alle tecnologie attualmente esistenti sul mercato?». In questo capitolo sono stati presentati tre di questi metodi. In tale fase, che è di particolare interesse per l'ingegneria biomedica, le valutazioni sono rese particolarmente difficili per diverse ragioni, tra cui la non disponibilità di dati clinici adeguati. Sarebbe, invece, sempre utile avere una valutazione, seppur preliminare, per decidere in tempo se e come investire risorse. Infatti, in fase di ricerca e sviluppo, con piccole modifiche si possono ottenere grandi risultati in termini di sostenibilità della tecnologia stessa.

Negli ultimi anni la comunità internazionale di ingegneria biomedica comincia a mostrare maggiore interesse all'HTA, cominciando a percepire l'importanza di tali metodiche anche per la ricerca. Ciò è dovuto anche all'accorciarsi della distanza temporale tra ricerca, sviluppo e commercializzazione di un dispositivo.

## Bibliografia

- ANTHONY, R.N. & YOUNG, D.W. 2003. *Management control in nonprofit organizations*, Boston, McGraw-Hill/Irwin.
- BRACALE, U., MELILLO, P., PIGNATA, G., DI SALVO, E., ROVANI, M., MEROLA, G. & PECCHIA, L. 2012a. Which is the best laparoscopic approach for inguinal hernia repair: TEP or TAPP? A systematic review of the literature with a network meta-analysis. *Surg Endosc*.
- BRACALE, U., ROVANI, M., BRACALE, M., PIGNATA, G., CORCIONE, F. & PECCHIA, L. 2012b. Totally laparoscopic gastrectomy for gastric cancer: meta-analysis of short-term outcomes. *Minim Invasive Ther Allied Technol*, 21, 150-60.
- BRIDGES, J.F., MOHAMED, A.F., FINNERN, H.W., WOEHL, A. & HAUBER, A.B. 2012. Patients' preferences for treatment outcomes for advanced non-small cell lung cancer: A conjoint analysis. *Lung Cancer*, 77, 224-31.
- CHAPMAN, R.H., STONE, P.W., SANDBERG, E.A., BELL, C. & NEUMANN, P.J. 2000. A comprehensive league table of cost-utility ratios and a sub-table of "panel-worthy" studies. *Med Decis Making*, 20, 451-67.
- COSH, E., GIRLING, A., LILFORD, R., MCATEER, H. & YOUNG, T. 2007. Investing in new medical technologies: A decision framework. *J Commer Biotechnol*, 13, 263-271.
- CRAVEN, M.P. & MORGAN, S.P. 2011. Early Stage Economic Evaluation with a Small Medical Device Start-Up Company Using a Markov Model. *Journal of Medical Devices*, 5, 027516-1.
- DE BEKKER-GROB, E.W., RYAN, M. & GERARD, K. 2012. Discrete choice experiments in health economics: a review of the literature. *Health Economics*, 21, 145-172.
- DONG, H. & BUXTON, M. 2006. Early assessment of the likely cost-effectiveness of a new technology: A Markov model with probabilistic sensitivity analysis of computer-assisted total knee replacement. *Int J Technol Assess Health Care*, 22, 191-202.
- FURST, D.E., SCHIFF, M.H., FLEISCHMANN, R.M., STRAND, V., BIRBARA, C.A., COMPAGNONE, D., FISCHKOFF, S.A. & CHARTASH, E.K. 2003. Adalimumab, a fully human anti tumor necrosis factor-alpha monoclonal antibody, and concomitant standard antirheumatic therapy for the treatment of rheumatoid arthritis: results of STAR (Safety Trial of Adalimumab in Rheumatoid Arthritis). *J Rheumatol*, 30, 2563-71.
- GALLEGO, G., BRIDGES, J.F., FLYNN, T. & BLAUVELT, B.M. 2011. Predicting the Future Impact of Emerging Technologies on Hepatocellular Carcinoma (Hcc): Measuring Stakeholders Preferences with Best-Worst Scaling. *Value in Health*, 14, A176-A176.
- GUYATT, G.H., SACKETT, D.L., SINCLAIR, J.C., HAYWARD, R., COOK, D.J. & COOK, R.J. 1995. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA*, 274, 1800-4.
- HEITMAN, E. 1998. Ethical issues in technology assessment. Conceptual categories and procedural considerations. *Int J Technol Assess Health Care*, 14, 544-66.

- IFMBE. *HTA division webpage* [Online]. Available: [http://www.ifmbe.org/index.php?option=com\\_content&view=category&id=53:health-care-technology-assessment&Itemid=167&layout=default](http://www.ifmbe.org/index.php?option=com_content&view=category&id=53:health-care-technology-assessment&Itemid=167&layout=default) [Accessed 20/08/2012].
- IJZERMAN, M.J. & STEUTEN, L.M. 2011. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Appl Health Econ Health Policy*, 9, 331-47.
- IJZERMAN, M.J., VAN TIL, J.A. & SNOEK, G.J. 2008. Comparison of two multi-criteria decision techniques for eliciting treatment preferences in people with neurological disorders. *Patient*, 1, 265-72.
- JARM, T., MIKLAVCIC, D., PALLIKARAKIS, N., BLIZNAKOV, Z., MAGJAREVIC, R., LACKOVIC, I., PECCHIA, L., STAGNI, R., JOBBAGY, A. & BARBENEL, J. 2012. Proposal for Generic Biomedical Engineering Programs Based on European Experience.
- 5th European Conference of the International Federation for Medical and Biological Engineering. In: JOBBÁGY, Á. & MAGJAREVIC, R. (eds.). Springer Berlin Heidelberg.
- KEYSTONE, E.C., KAVANAUGH, A.F., SHARP, J.T., TANNENBAUM, H., HUA, Y., TEOH, L.S., FISCHKOFF, S.A. & CHARTASH, E.K. 2004. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum*, 50, 1400-11.
- LIBERATORE, M.J. & NYDICK, R.L. 2008. The analytic hierarchy process in medical and health care decision making: A literature review. *European Journal of Operational Research*, 189, 194-207.
- LUMLEY, T. 2002. Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*, 21, 2313-24.
- MULYE, R. 1998. An empirical comparison of three variants of the AHP and two variants of conjoint analysis. *Journal of Behavioral Decision Making*, 11, 263-280.
- NIHR. *Health Technology Assessment programme* [Online]. Available: <http://www.hta.ac.uk/> [Accessed 20/08/2012].
- NORD, E. 1999. Towards cost-value analysis in health care? *Health Care Anal*, 7, 167-75.
- OMS. *OMS Medical Devices HTA program* [Online]. Available: [http://www.who.int/medical\\_devices/assessment/en/](http://www.who.int/medical_devices/assessment/en/) [Accessed 20/08/2012].
- PALLIKARAKIS, N., BLIZNAKOV, Z., MIKLAVCIC, D., JARM, T., MAGJAREVIC, R., LACKOVIC, I., PECCHIA, L., STAGNI, R., JOBBAGY, A. & BARBENEL, J. Year. Promoting harmonization of BME education in Europe: The CRH-BME Tempus project. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, Aug. 30 2011-Sept. 3 2011 2011. 6522-6525.
- PECCHIA, L., ACAMPORA, F., ACAMPORA, S. & BRACALE, M. 2007. A Multi Scale Methodology for Technology Assessment. A case study on Spine Surgery. *11th Mediterranean Conference on Medical and Biological Engineering and Computing 2007, Vols 1 and 2*, 16, 762-765.
- PECCHIA, L., BATH, P., PENDLETON, N., JACKSON, S., CLARKE, C., BRIGGS, P.,

- MCINNES, L., ANGELOVA, M. & M., B. 2011a. The use of analytic hierarchy process for the prioritization of factors affecting wellbeing in elderly. *International Symposium on Analytic Hierarchy Process (ISAHP)*. Sorrento, Naples, Italy.
- PECCHIA, L., BATH, P.A., PENDLETON, N. & BRACALE, M. 2010. Web-based system for assessing risk factors for falls in community-dwelling elderly people using the analytic hierarchy process. *International Journal of the Analytic Hierarchy Process*, 2.
- PECCHIA, L., BATH, P.A., PENDLETON, N. & BRACALE, M. 2011b. Analytic Hierarchy Process (AHP) for examining healthcare professionals' assessments of risk factors. The relative importance of risk factors for falls in community-dwelling older people. *Methods Inf Med*, 50, 435-44.
- PECCHIA, L., BRACALE, U. & BRACALE, M. 2009a. Health Technology Assessment of Home Monitoring for the Continuity of Care of patient suffering from congestive heart failure. In: DÖSSEL, O. & SCHLEGEL, W.C. (eds.) *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*. Springer Berlin Heidelberg.
- PECCHIA, L., BRACALE, U., MELILLO, P., SANSONE, M. & BRACALE, M. 2009b. AHP for Health Technology Assessment. A case study: prioritizing care approaches for patients suffering from chronic heart failure. *International Symposium on AHP (ISAHP)*. Pittsburgh, Pennsylvania, USA.
- PECCHIA, L., MARTIN, J.L., RAGOZZINO, A., VANZANELLA, C., SCOGNAMIGLIO, A., MIRARCHI, L. & MORGAN, S. Submitted 2012. User needs elicitation via analytic hierarchy process (AHP). A case study on a CT scanner. *BMC Medical Informatics and Decision Making*.
- PECCHIA, L., MIRARCHI, L., DONIACOVO, R., MARSICO, V. & BRACALE, M. 2009c. Health Technology Assessment for a Service Contract: a new method for decisional tools. *World Congress on Medical Physics and Biomedical Engineering, Vol 25, Pt 12, 25*, 105-108.
- REISER, S.J. 1988. A perspective on ethical issues in technology assessment. *Health Policy*, 9, 297-300.
- SAATY, T. 1982. How to Structure and Make Choices in Complex Problems. *Human Systems Management*, 3, 255-261.
- SAATY, T.L. 1977. A scaling method for priorities in hierarchical structures *Journal of Mathematical Psychology*, 15, 8.
- SAATY, T.L. & VARGAS, L.G. 2001. *Models, methods, concepts & applications of the analytic hierarchy process*, Boston, Kluwer Academic Publishers.
- SACCHINI, D., VIRDIS, A., REFOLO, P., PENNACCHINI, M. & DE PAULA, I.C. 2009. Health technology assessment (HTA): ethical aspects. *Med Health Care Philos*, 12, 453-7.
- SCHOLL, A., MANTHEY, L., HELM, R. & STEINER, M. 2005. Solving multiattribute design problems with analytic hierarchy process and conjoint analysis: An empirical comparison. *European Journal of Operational Research*, 164, 760-777.
- SLOANE, E.B., LIBERATORE, M.J., NYDICK, R.L., LUO, W.H. & CHUNG, Q.B. 2003. Using the analytic hierarchy process as a clinical engineering tool to facilitate an iterative, multidisciplinary, microeconomic health technology assessment. *Computers & Operations Research*, 30, 1447-1465.

- SONNENBERG, F.A. & BECK, J.R. 1993. Markov models in medical decision making: a practical guide. *Med Decis Making*, 13, 322-38.
- SUTTON, A. J. 2000. *Methods for meta-analysis in medical research*, Chichester ; New York, J. Wiley.
- VAN DE PUTTE, L.B., ATKINS, C., MALAISE, M., SANY, J., RUSSELL, A.S., VAN RIEL, P.L., SETTAS, L., BIJLSMA, J.W., TODESCO, S., DOUGADOS, M., NASH, P., EMERY, P., WALTER, N., KAUL, M., FISCHKOFF, S. & KUPPER, H. 2004. Efficacy and safety of adalimumab as monotherapy in patients with rheumatoid arthritis for whom previous disease modifying antirheumatic drug treatment has failed. *Ann Rheum Dis*, 63, 508-16.
- VAN DE PUTTE, L.B., RAU, R., BREEDVELD, F.C., KALDEN, J.R., MALAISE, M.G., VAN RIEL, P.L., SCHATTENKIRCHNER, M., EMERY, P., BURMESTER, G.R., ZELDNER, H., MOUTSOPOULOS, H.M., BECK, K. & KUPPER, H. 2003. Efficacy and safety of the fully human anti-tumour necrosis factor alpha monoclonal antibody adalimumab (D2E7) in DMARD refractory patients with rheumatoid arthritis: a 12 week, phase II study. *Ann Rheum Dis*, 62, 1168-77.
- WEINBLATT, M.E., KEYSTONE, E.C., FURST, D.E., MORELAND, L.W., WEISMAN, M.H., BIRBARA, C.A., TEOH, L.A., FISCHKOFF, S.A. & CHARTASH, E.K. 2003. Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. *Arthritis Rheum*, 48, 35-45.