

## EFFICACIA RELATIVA E DI IMPATTO DI INIZIATIVE NELL'AMBITO DEI SERVIZI ALLA PERSONA DI PUBBLICA UTILITÀ

Pietro Giorgio Lovaglio

### 1. LA VALUTAZIONE ATTRAVERSO RISULTATI OGGETTIVI

La valutazione dell'efficacia degli interventi istituzionali nell'ambito dei servizi alla persona richiede in primo luogo l'individuazione di variabili di performance, misurabili a livello del singolo soggetto coinvolto nell'erogazione del servizio, che siano il più possibile oggettive e connesse agli obiettivi dell'intervento in modo da assicurare elevati livelli di trasparenza (Mead, 1997; 1999; Smith e Meier, 1994; Milward e Provan, 2000; Roderick *et al.*, 2000).

L'opportunità di fondare la valutazione su indicatori di risultato (*outcome*) piuttosto che di qualità di processo deriva dal fatto che i servizi alla persona sono tipicamente *experience good* la cui "qualità" non è una proprietà intrinseca del servizio o del processo, ma si misura in base al risultato, sovente di lungo periodo, che questi hanno sugli individui interessati (Gori e Vittadini, 1999; Dagum e Vittadini, 1996) a differenza di altri tipi di beni e servizi, detti *search goods*, la cui qualità è osservabile direttamente in base alle caratteristiche del bene stesso o del suo processo di produzione.

La stima dei risultati diventa ancora più importante qualora lo scopo dell'analisi consista nella valutazione dell'efficacia di iniziative o progetti avviati da Istituzioni che prevedono l'erogazione di un servizio o un contributo erogato agli utenti, oppure nel caso di valutazione di una serie di agenzie o enti (Agenti) che erogano servizi alla persona, secondo modalità differenti, dettate da differenti assetti organizzativi, o dal diverso contesto demografico-economico del territorio in cui esse operano.

Dal punto di vista operativo gli strumenti fondamentali per la costruzione di un processo di valutazione sono essenzialmente due:

1. la specificazione di un modello che stimi i risultati come misure oggettive di performances o benessere acquisito dagli utenti; in ambito clinico nella maggior parte dei casi i risultati riguardano un determinato aspetto della qualità della vita e sono tipicamente desunti da Item di natura ordinale;

2. la specificazione di un modello statistico che consenta di stimare sia l'efficacia di una serie di Agenti, identificando in maniera precisa quali gestori del

servizio favoriscono migliori performances o assicurano maggiore benessere ai propri utenti (efficacia relativa), sia l'efficacia dell'iniziativa istituzionale *tout court* (efficacia di impatto). La valutazione di efficacia assoluta, o di impatto (Heckman, 1976; Heckman e Hotz, 1989) ha lo scopo di comparare l'effetto dell'intervento con quello che si avrebbe in assenza dello stesso; tale valutazione richiede il confronto con un gruppo di controllo di individui "non trattati" ed è alla base delle metodologie che hanno per oggetto i criteri di confrontabilità degli studi caso-controllo e della *selection bias* (Heckman e Rob, 1986; Heckman e Hotz, 1989; Rosenbaum e Rubin, 1984).

Per quanto riguarda il primo punto, le caratteristiche e le proprietà delle misure ottenute dai modelli psicometrici di Item Response Theory e in particolare della Rasch Analysis (Lovaglio, 2004b) rendono applicabile naturalmente in tale contesto l'uso di tali metodologie, rispetto ad altri approcci legati a metodi di stima di outcome latenti arbitrari, come i modelli strutturali e le tecniche di Optimal Scaling (Lovaglio, 2004a; Vittadini e Lovaglio, 2001).

Il presente lavoro affronta in dettaglio il secondo punto auspicato negli studi di valutazione di impatto e di efficacia relativa, fornendo il razionale per una corretta procedura di valutazione oggettiva al netto dei fattori che possono distorcere i risultati (Paragrafo 2) e introducendo un modello che valuta congiuntamente sia l'efficacia di impatto che relativa (Paragrafo 3); infine nel Paragrafo 4 il modello proposto verrà applicato alla valutazione dell'efficacia di un intervento istituzionale-sperimentale promosso dalla Regione Lombardia nel 2002, consistente in un contributo economico (Buono Socio-Sanitario) a favore delle famiglie che decidono di assistere in casa i propri anziani non autosufficienti, come alternativa al ricovero in Residenze Socio Assistenziali (Rsa).

## 2. LA VALUTAZIONE DELL'EFFICACIA

In letteratura l'outcome più usato nella maggior parte degli studi clinici connessi con l'ospedalizzazione dei pazienti è il tasso di mortalità intraospedaliero entro 30 giorni dalla dimissione (per una rassegna si veda Marshall e Spiegelhalter, 2001). Tuttavia molti autori ritengono che non si possa individuare l'efficacia utilizzando esclusivamente tale outcome (Dubois *et al.*, 1988; Epstein, 1995; Marshall e Spiegelhalter, 2001; Goldstein e Spiegelhalter, 1996; Epstein e Schneider, 1997); inoltre in alcuni ambiti della valutazione di Agenti o iniziative di natura clinica non connesse con il ricovero in ospedale, come l'ospedalizzazione domiciliare o ambiti di cura svolti da strutture territoriali legate a fenomeni di disagio psico-sociale ad esempio, i fattori clinici di interesse coincidono con risultati inerenti la qualità della vita che descrivono la funzionalità fisica o psichica generale o quella inerente ad aspetti particolari della vita del paziente.

Ad esempio la *Functional Independence Measurement* è specifica per le patologie che indicano la riabilitazione (Tesio *et al.*, 1997) oppure la scala SF-12 (Ware, 1993) si usa preferibilmente per malattie cronico-degenerative. Tali risultati sono rilevabili

in modo semplice da indicatori su scala ordinale attraverso opportuni questionari; tuttavia ognuno di loro è specifico per particolari patologie.

Una volta stimati in maniera oggettiva i risultati (che dal punto di vista statistico vengono assunte come variabili latenti sottostanti ad una serie di Item di tipo ordinale) sugli utenti del servizio, per valutare l'effettiva incidenza dell'Agente che eroga il servizio va costruito un adeguato modello basato su indicatori di efficacia, intesi come misure di valore aggiunto, depurati sia da fattori individuali esterni, sia da fattori legati alla tipologia della struttura che eroga il servizio alla persona (Fitz e Gibbon, 1997; Heinrich e Lynn, 2001), in maniera alternativa al semplicistico utilizzo di indicatori univariati basati sulle distribuzioni dei risultati ottenuti. In particolare i modelli che valutano l'efficacia di strutture che erogano servizi alla persona vanno specificati tenendo conto dei seguenti aspetti:

a) poiché si stanno valutando servizi relazionali alla persona, tipicamente di natura *experience goods* (Gori e Vittadini, 1999; Lovaglio, 2001), si devono considerare le caratteristiche personali apparentemente estranee all'erogazione del servizio (Hanushek, 1971; Aitkin e Longford, 1986) poiché è ben noto che esse possono influire in maniera decisiva sui risultati (Goldstein e Spiegelhalter, 1996) a prescindere dall'erogazione del servizio (Stiefel, 1997). Nella valutazione dei corsi di formazione professionale è noto che il sesso, l'età e la scolarità dei formandi contribuisca in larga misura al raggiungimento di obiettivi da parte dell'Agenzia che eroga il servizio (ORML, 1997; ISFOL, 2001; Gori, 2000) così come la gravità all'ingresso o l'età determini l'efficacia dell'intervento ospedaliero o terapeutico (Gibbons e Hedeker, 2000; Goldstein e Spiegelhalter, 1996).

In ambito clinico, ad esempio, la variazione dello stato di salute alla dimissione non è identica per ogni stato iniziale (valore basale) dell'utente (Fleiss, 1986; Rice e Leyland, 1996); generalmente esiste correlazione tra miglioramenti dovuti al trattamento e valore basale; da questo può derivare che i pazienti inizialmente più gravi siano quelli che mostrano il maggior miglioramento a prescindere dal tipo di intervento effettuato da parte dell'Agente: questo suggerirebbe di considerare come informative per i vari Agenti-trattamenti  $j$  non semplicemente la media (valutata per i pazienti nel  $j$ -esimo trattamento) dei miglioramenti ( $m_j = x_j - z_j$ ) tra stato di benessere finale ( $x_j$ ) e stato iniziale ( $z_j$ ), ma il miglioramento medio al netto dello stato iniziale ( $e_j$ ), ottenuto attraverso il modello di regressione (Fleiss, 1986):

$$e_j = m_j - \beta^* z_j \quad j = 1, \dots, k \quad (1)$$

dove  $\beta^*$  è il coefficiente stimato dalla regressione di  $x_j$  su  $z_j$ , in modo tale che  $e_j$  misuri l'effettivo risultato di efficacia dell'Agente-trattamento  $j$  al netto dello stato di ingresso dei pazienti.

Al di là della semplicistica proposta dell'equazione (1), si ribadisce quindi che in processi di valutazione, specie in ambito clinico, tra le caratteristiche individuali va specificata dunque nel modello una misura dello stato di benessere prima dell'erogazione del servizio, in modo che la valutazione degli Agenti ne sia depurata; se così non fosse non solo si distorcerebbero i confronti tra gli Agenti, ma essi sarebbero giustificati a selezionare i propri utenti solo tra coloro che sono po-

tenzialmente migliorabili o nelle condizioni di partenza migliori, inducendo un problema di *selezione avversa* (Moramarco, 1999).

b) Al fine di confrontare gli Agenti, a parità di condizioni, bisogna considerare sia variabili di contesto sia variabili valutate su singolo Agente in modo da eliminare fattori potenzialmente distorsivi; ad esempio nella valutazione dell'efficacia di una serie di enti di formazione professionale sarebbe opportuno tenere conto della situazione del mercato del lavoro nella provincia in cui essi operano (Gori, 2000), in ambito sanitario-assistenziale della struttura demografica del territorio (Goldstein e Spiegelhalter, 1996) oltre a variabili che identificano la dimensione e le risorse (umane, strutturali e finanziarie) degli Agenti sanitari;

c) in numerosi studi di efficacia relativa (Rice e Leyland, 1996; ORML, 1997; ISFOL, 2001) l'evidenza empirica mostra che l'efficacia degli Agenti dipende dal meccanismo di accesso/selezione degli utenti a tali strutture (es. Ospedali / Enti di Formazione Professionali) solitamente dipendente dalla *tipologia normativa-istituzionale* di ogni Agente (es. Ospedali pubblici, privati, Istituti di ricerca, Case di cura, Centri di riabilitazione / Enti di formazione professionale per la formazione superiore, formazione post laurea, formazione continua, oppure Enti di carattere pubblico o privato); occorre quindi eliminare questo fattore che determina il problema della *selection bias* (Heckman e Rob, 1986; Heckman e Hotz, 1989) inserendo nel modello opportune interazioni (Cox e McCullagh, 1982) tra il valore basale e la tipologia istituzionale degli Agenti.

Alla luce di queste considerazioni, il significato di efficacia relativa (effetto) di un Agente si fa coincidere unanimemente, almeno nella letteratura anglosassone, con il concetto di "valore aggiunto" che l'Agente apporta ai propri "clienti", misurato dalle variazioni dello stato di benessere (rispetto allo stato iniziale) al netto di fattori che possono distorcere i risultati (variabili individuali e variabili legate all'Agente).

Dal punto di vista metodologico il candidato naturale alla stima del valore aggiunto è il modello di regressione a struttura multilivello (Goldstein e Spiegelhalter, 1996) come alternativa ai modelli di regressione classici, esemplificati ad esempio nell'equazione (1), poichè la struttura gerarchica del sistema (utenti clusterizzati all'interno degli Agenti) induce una correlazione significativa tra gli individui (non selezionati casualmente tra gli Agenti secondo i consueti disegni sperimentali) trattati dallo stesso Agente (intra-class correlation), facendo venir meno l'assunzione di indipendenza tra le osservazioni, tipica del modello lineare classico.

I modelli multilivello, ampiamente utilizzati nell'ambito di valutazioni delle performance, specialmente in campo educativo e sanitario (Muthen e Speckart, 1985; Rice e Leyland, 1996; Turner *et al.*, 2000), consentono di tenere conto della struttura gerarchica del processo, riuscendo a depurare gli effetti dell'Agente da variabili esterne (sia a livello di utente, di Agente e di contesto socio-economico-ambientale) e a far comprendere il ruolo dei fattori a livello di Agente (tipo di struttura, organizzazione, risorse) che determinano la variabilità delle performance tra Agenti, individuando le unità più o meno efficaci tra quelle incluse nel campione di Agenti (efficacia relativa).

### 3. UN MODELLO PER LA STIMA DELL'EFFICACIA RELATIVA E DI IMPATTO

In ambito di valutazione, un modello multilivello consiste nella specificazione di un modello di regressione, valutato a livello degli utenti (I livello), in cui gli effetti inerenti l'efficacia media degli Agenti (II livello) sono catturate dalle intercette del modello e supposte di tipo casuale, variabili al variare dell'Agente.

L'equazione di I livello viene specificata nella (2)

$$y_{ij} = \beta_{0j} + \sum_s \beta_{sj} x_{sij} + \gamma_j z_{ij} + r_{ij} \quad r_{ij} \sim N(0, \sigma^2) \quad i=1, \dots, n_j \quad j=1, \dots, k \quad s=1, \dots, S \quad (2)$$

dove  $y_{ij}$  è la determinazione dell'outcome valutato sull' $i$ -esimo utente ( $i=1, \dots, n_j$ ) cui è erogato il servizio dall'Agente  $j$ -esimo ( $j=1, \dots, k$ ) dopo l'erogazione del servizio;  $z_{ij}$  è la determinazione del valore basale sull'utente  $i$  trattato dall'Agente  $j$ ,  $x_{sij}$  è la determinazione della variabile esplicativa  $s$ -esima ( $s=1, \dots, S$ ) appartenente all'insieme delle variabili esplicative individuali rilevata sull'utente  $i$ -esimo, trattato dall'Agente  $j$  ( $\beta_{sj}$  è l' $s$ -esimo coefficiente di regressione parziali),  $r_{ij}$  i residui casuali relativi al soggetto  $i$  trattato dall'Agente  $j$ , supposti I.I.D. distribuiti normalmente con valore atteso nullo e con varianza costante ( $\sigma^2$ ).

Per quanto detto ai fini della valutazione, il parametro  $\gamma_j$ , che misura il legame tra stato di benessere all'ingresso e stato di benessere finale o di lungo periodo, viene supposto non fisso, ma di tipo casuale (con  $u_j$  il vettore di residui di II livello) variabile al variare della tipologia istituzionale dell'Agente, identificata dalla variabile categoriale  $D_j$

$$\gamma_j = \pi D_j + u_j \quad u_j \sim N(0, \theta_z) \quad (3)$$

dove  $\pi$  è il parametro che valuta l'effetto dell'interazione tra tipologia di Agente e valore basale e  $u_j$  il vettore di errori casuali supposti I.I.D distribuiti normalmente con valore atteso nullo e con varianza costante ( $\theta_z$ ).

Allo stesso modo il vettore delle intercette ( $\beta_{0j}$ ), che individua l'effetto medio sull'outcome  $y$  dell'Agente  $j$ -esimo e contiene i parametri di efficacia globale degli Agenti, derivante dalle caratteristiche e dalle risorse degli Agenti, dipende da variabili a livello di Agente e di contesto, come specificato nella (4):

$$\beta_{0j} = \mu + \sum_r \phi_{rj} f_{rj} + m_j \quad m_j \sim N(0, \theta_m) \quad r=1, \dots, R \quad (4)$$

in cui,  $\mu$  la media generale,  $f_{rj}$  è il regressore  $r$ -esimo ( $r=1, \dots, R$ ) relativo al  $j$ -esimo Agente (covariate di II livello) con  $\phi_{rj}$  parametro di regressione ed  $m_j$  il vettore di errori casuali, supposti I.I.D distribuiti normalmente con valore atteso nullo e con varianza costante ( $\theta_m$ ).

Congiuntamente gli errori di II livello ( $m_j, u_j$ ) hanno distribuzione normale bivariata con valori attesi nulli, con varianze costanti rispettivamente pari a ( $\theta_m, \theta_z$ ) e covarianza  $\theta_{mz}$ , mentre  $r_{ij}$  e ( $m_j, u_j$ ) sono assunti indipendenti.

Sostituendo la (3) e la (4) nella (2), la forma compatta del modello

$$y_{ij} = \mu + \sum_s \beta_{sj} x_{sj} + \sum_r \phi_{rj} f_{rj} + \pi D_j z_{ij} + m_j + z_{ij} u_j + r_{ij} \quad (5)$$

A                      B                      C                      E

informa che  $m_j$  (vettore dei residui di II livello), essendo depurata dalle caratteristiche individuali, di contesto e da quelle legate all'Agente erogatore, costituisce l'efficacia derivata dalla "capacità manageriale" dell'Agente o efficacia relativa (E), al netto di tutti gli effetti di disturbo sopra menzionati: le caratteristiche individuali degli utenti (A), quelle relative agli Agenti o ai fattori di contesto (B) e dai fattori di accesso alle strutture erogatrici (C).

Nella valutazione dell'efficacia di iniziative o progetti istituzionali è di altrettanto interesse l'efficacia assoluta, o di impatto (l'effetto dell'intervento con quello che si avrebbe in assenza dello stesso); le misure di efficacia  $m_j$  nella (5) come residui di II livello sono soltanto valutazioni relative (Gori e Vittadini, 1999) e non assolute, nel senso che un Agente potrebbe risultare più o meno efficace di altri, senza che ciò faccia emergere se le performances raggiunte dai suoi utenti siano maggiori o minori del risultato che sarebbe stato possibile ottenere senza la presenza dell'intervento.

In tale contesto l'efficacia di impatto viene condotta confrontando i risultati degli utenti cui è stato erogato il servizio o un trattamento da parte di un particolare Agente (gruppo di trattati) con un gruppo di utenti non trattati, ma simili per caratteristiche (gruppo di controllo).

Poiché generalmente esiste correlazione tra (assegnazione del) trattamento e variabile risposta, appare chiaro che la stima d'impatto dipenda in maniera significativa dal processo di selezione del gruppo di controllo; infatti, negli studi classici di efficacia di impatto (Heckman e Rob, 1986; Heckman e Hotz, 1989) l'eventuale presenza di disomogeneità tra i due gruppi (gruppo di trattati e gruppo di controllo) dovuta a fattori inosservabili potrebbe condurre a problemi di distorsione da selezione, o *selection bias* dovuta alla non randomizzazione dei trattamenti assegnati ai soggetti.

A tale scopo i soggetti appartenenti al gruppo di controllo vengono selezionati attraverso metodologie che sulla base di covariate a livello individuale ( $z_{ij}$ ) bilanciano i due gruppi, rendendo omogenee le distribuzioni delle  $z_{ij}$  all'interno degli strati tra i due gruppi; una di queste metodologie, definita *Propensity score*, (Rosenbaum e Rubin, 1984) si basa sulla stima della probabilità che ogni soggetto in analisi appartenga al gruppo dei trattati, condizionatamente ad un vettore di covariate  $z_{ij}$ , non necessariamente le stesse di quelle inserite nel modello di valutazione di efficacia di impatto.

Il *propensity score* è utile poichè in qualche modo riassume in una sola dimensione le fonti di disomogeneità dei due gruppi, poichè condizionatamente al propensity score vi è indipendenza tra le covariate e la variabile trattamento.

Rispetto alla stima dei parametri del modello di efficacia, alcuni approcci propongono la specificazione di un parametro di sensitività che riflette il grado di non casualità delle osservazioni (Copash e Li, 1997; Maddala, 1983; Rosenbaum,

1995; Copas e Eguchi, 2001): in altri termini tali metodi valutano la distorsione della *selection bias* dovuta a fonti latenti (*selection on unobservables*, Heckman e Hotz, 1989) attraverso analisi di sensitività, ma non in contesti di dati di tipo gerarchico; per una rassegna dei metodi tradizionali valutati in una logica multilevel si veda Bellio e Gori (2000).

Per quanto detto, valutare l'efficacia di impatto congiuntamente all'efficacia relativa si realizza inserendo all'interno della (5) un parametro ( $\varepsilon$ ) che misura l'impatto dell'intervento rispetto al non intervento, come evidenziato nella (6):

$$y_{ij} = \beta_{0j} + \sum_s \beta_{sj} x_{sij} + \gamma_j z_{ij} + \varepsilon I_{ij} + r_{ij} \quad j=0, \dots, k \quad i=1, \dots, n_j$$

$$\beta_{0j} = \mu + \sum_r \phi_{rj} f_{rj} + I_{ij} m_j \quad \gamma_j = \pi D_j + I_{ij} u_j \quad (6)$$

In tale contesto l'analisi di efficacia relativa si pone solo tra strutture erogatrici del servizio, mentre quella di impatto tra i due gruppi di pazienti (a cui è stato erogato il servizio e il gruppo di controllo): ai fini della modellazione statistica ciò equivale ad inserire, in corrispondenza degli errori di II livello una variabile dummy  $I_{ij}$  che vale 0 per gli individui  $i$  del gruppo di controllo (non trattati da nessun Agente cui corrisponde l'indice  $j=0$ ) e 1 per gli individui trattati (da uno dei possibili  $k$  Agenti cui corrispondono indici  $j=1, \dots, k$ ).

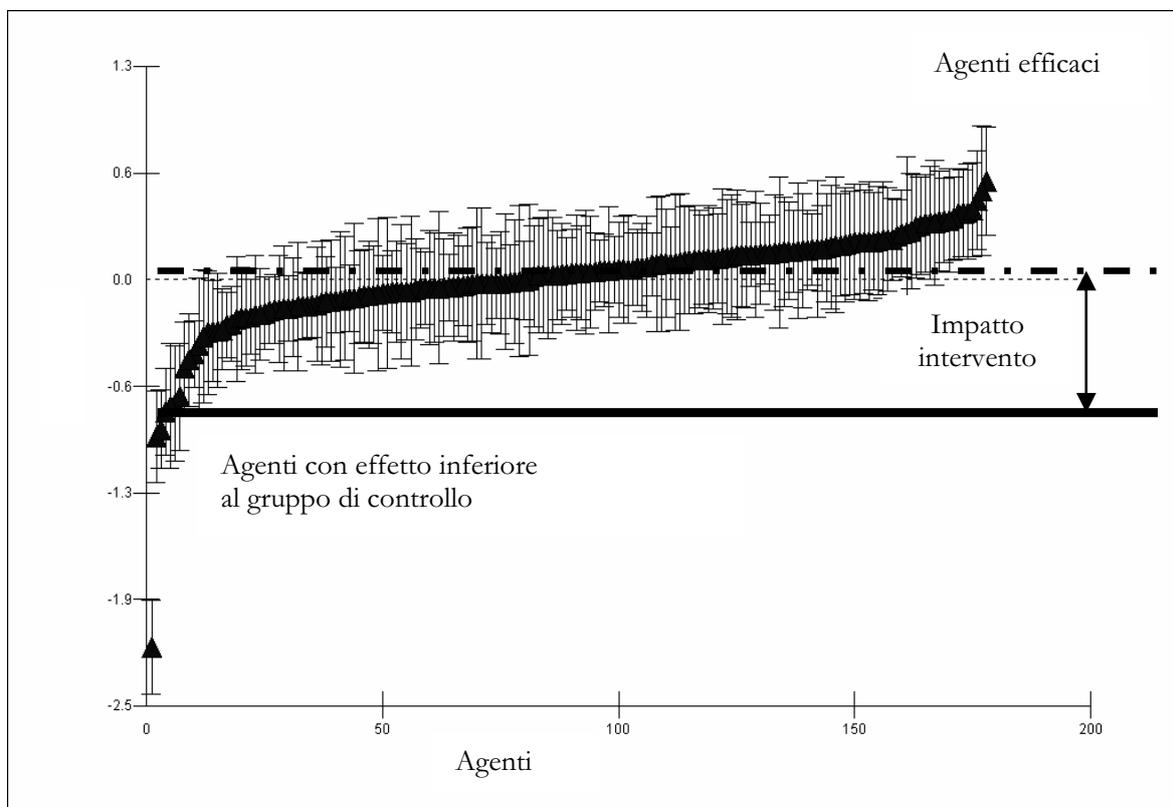


Figura 1 – Effetto degli Agenti in ambito relativo e di efficacia di impatto.

In tal modo, seguendo il modello di Langford e Lewis (1998), è possibile valutare non solo l'efficacia dell'intervento, attraverso la stima di  $\epsilon$ , ma anche quali Agenti risultano efficaci in senso relativo (rispetto agli altri Agenti) e in senso assoluto (rispetto al non intervento), attraverso la stima degli errori di II livello  $m_j$ , ottenuta da stimatori Empirical Bayes o "shrinkage estimators" (Goldstein e Spiegelhalter, 1996).

I risultati che si ottengono sono sintetizzabili attraverso il grafico rappresentato in figura 1, in cui oltre ai consueti intervalli di confidenza della stima degli effetti degli Agenti appare l'ulteriore informazione costituita dall'impatto medio dell'intervento (linea tratto-punto) rispetto al non intervento (linea continua). Dalla figura emerge che, in questo tipo di analisi, è possibile individuare gli Agenti efficaci (sulla parte destra della figura) e Agenti per i quali l'erogazione del servizio risulta avere un effetto inferiore od uguale a quello del gruppo di controllo (sulla parte sinistra della figura).

#### 4. UNA APPLICAZIONE: LA VALUTAZIONE DEL BUONO SOCIO SANITARIO DELLA REGIONE LOMBARZIA

Il modello di efficacia relativa e di impatto proposto nei paragrafi precedenti viene applicato alla valutazione di un intervento promosso dalla Regione Lombardia per l'anno 2002 consistente nell'erogazione sperimentale di un contributo economico denominato Buono Socio-Sanitario (Bss) a favore di quegli anziani non autosufficienti che decidono di essere assistiti in famiglia, in alternativa al ricovero in casa di cura (Rsa).

Poichè l'allontanamento dell'anziano dal nucleo familiare con conseguente ricovero in Rsa sembra determinato in primo luogo dal progressivo peggioramento dello stato cognitivo (Tsuij *et al.*, 1995; Tesio *et al.*, 1996), la Regione Lombardia, visto il lento invecchiamento della popolazione, cerca strumenti volti a favorire l'accudimento dell'anziano in ambiente familiare rispetto al ricovero in Rsa, per evidenti ragioni economiche, oltre che di natura politica legate a favorire la libertà di scelta dei cittadini. In tale ottica l'efficacia dell'intervento del Bss sarebbe compatibile con una situazione di minor peggioramento cognitivo (nel tempo) degli assistiti in famiglia, rispetto al campione di anziani residenti in Rsa o quantomeno con una situazione di sostanziale equivalenza, giustificando in tal modo altre iniziative analoghe.

Per valutare dunque se il Bss comporti dei benefici sullo stato di salute mentale (autosufficienza cognitiva) degli anziani beneficiari rispetto a quelli ricoverati in Rsa, viene specificato un modello multilivello che permette di stimare congiuntamente:

1. l'efficacia dell'iniziativa Bss valutata dal confronto dello stato di autosufficienza cognitiva acquisito dai beneficiari dopo l'erogazione (al netto del valore basale), rispetto agli anziani residenti in Rsa (efficacia dell'impatto);
2. l'efficacia relativa degli Agenti che hanno prestato servizi socio-assistenziali domiciliari sui beneficiari del Bss, distinguendo in particolar modo tra forme di

assistenza domiciliare informale (entourage familiare) e assistenza domiciliare istituzionale o accreditata (erogata da Agenti sottoscrittori di patto per l'accreditamento con la Regione Lombardia).

Per rendere equo il confronto tra assistenza familiare e "assistenza accreditata" nel modello di efficacia relativa verranno selezionati solo quegli anziani che hanno acquistato servizi sanitari di tipo socio-assistenziale (aiuti nella gestione del menage familiare: igiene personale, supervisione, vigilanza, aiuto nei pasti, etc.) che da sola rappresenta il 60% della spesa totale degli anziani beneficiari del Bss. Non verranno selezionati anziani che hanno impiegato i finanziamenti in spese di tipo sanitario perchè non è ragionevole supporre che un anziano che necessita di cure medico-infermieristiche possa essere accudito in famiglia senza ricorso ad assistenza accreditata o specializzata.

#### 4.1 Campionamento e rappresentatività del campione

La scheda di rilevazione è stata somministrata ad un campione rappresentativo di anziani beneficiari dell'intervento istituzionale, prima dell'erogazione del Bss tra Ottobre e Dicembre 2001, selezionando casualmente un campione stratificato (in modo proporzionale al bacino dei beneficiari distribuiti nelle ASL Lombarde) di 670 anziani beneficiari.

In tale applicazione non è stato possibile applicare la metodologia del Propensity Score per selezionare il gruppo di controllo di anziani residenti in Rsa per l'indisponibilità di un archivio elettronico dei residenti in Rsa lombarde al momento della scelta del disegno campionario.

Il campionamento degli anziani "appaiati" in Rsa è avvenuto selezionando casualmente 330 ricoverati in Rsa dalla popolazione eligibile costituita da anziani ricoverati in Rsa lombarde al 1 Ottobre 2001 da almeno di 6 mesi che avevano sia gli stessi requisiti normativi per l'assegnazione del Bss (invalidità civile al 100% con assegno di accompagnamento, reddito familiare sotto una certa soglia) sia caratteristiche anagrafiche analoghe (età superiore agli 85 anni, presenza di figli viventi).

I due campioni sono stati dimensionati complessivamente sul 10% della popolazione corrispondente, composta da 6780 beneficiari del Bss e da 3228 ricoverati in Rsa "appaiati" per caratteristiche alla popolazione di beneficiari del Bss.

Dopo un lasso di tempo di 4-8 mesi (tra il Febbraio e il Maggio 2002) si è verificato un basso tasso di abbandono tra i due istanti temporali: alla seconda rilevazione nei due campioni figuravano 663 assegnatari del Bss e 322 ricoverati in Rsa.

#### 4.2 La specificazione del modello per la stima dell'efficacia

Come misura di risultato si è utilizzata la stima dello stato di dipendenza cognitiva ottenuta attraverso il modello *Rasch Partial Credit* (Lovaglio, 2004b), partendo dai punteggi degli Item cognitivi della scala *Functional Independent Measure* (FIM, Granger *et al.*, 1993), un questionario di valutazione dello stato di dipendenza motoria e cognitiva accolto come standard internazionale dall'U.S. National Institute

on Disability and Rehabilitation Research che censisce 13 attività della vita quotidiana legata alle funzioni sfinterico-motorie e 5 legate a funzioni cognitive (Comprensione, Espressione, Rapporto con altri, Soluzione problemi, Memoria), attraverso Item a risposta ordinale con punteggi variabili da 1 (completa dipendenza dagli altri) e 7 (completa autosufficienza).

Dalle caratteristiche individuali vengono selezionate, sulla base di statistiche di adattamento come l'Akaike Information Criterion e la Logverosimiglianza (Goldstein e Spiegelhalter, 1986), lo stato mentale cognitivo valutato prima dell'intervento ( $z$ ), lo stato di autosufficienza sfinterico-motorio ( $mot$ ), misurato con il modello Rasch sulle altre 13 attività censite dalla scala FIM, lo stato di instabilità clinica ( $inst$ ), valutato attraverso la Scala Pulses (Granger *et al.*, 1979) che prevede 4 gradi di instabilità clinica (1=stabile, 4=altamente instabile), il tempo medio ( $min$ ), in minuti, di assistenza giornaliera dedicato ad attività socio-assistenziali (alimentazione, lavaggio, vestizione, supervisione, spostamenti).

Per l'efficacia di impatto la variabile dummy  $I_{ij}$  vale 0 se l'anziano è residente in Rsa e 1 se l'anziano gode del Bss e viene accudito in casa: per questi ultimi soggetti i valori dell'indice  $j$ , che descrivono l'Agente che ha prestato assistenza domiciliare, variano da 1 a  $k$ .

La Tipologia istituzionale  $t$  dell'Agente che eroga assistenza domiciliare ( $D_t$ ) è stata definita sulla base dell'impiego dei finanziamenti ricevuti: essa coincide con l'entourage familiare ( $D_t=0$ ), con Istituzioni sottoscrittrici di Patto per l'Accreditamento con la Regione Lombardia ( $D_t=1$ ) per coloro che hanno dichiarato di aver acquistato assistenza da Agenti accreditati e con Agenti non accreditati ( $D_t=2$ ).

Rispetto all'analisi di efficacia relativa si sono confrontate le performances sullo stato di scadimento cognitivo tra i due periodi per i beneficiari che hanno investito il Bss nell'entourage familiare (Agente con codice 0), sulle Istituzioni non accreditate (Agente con codice 1) o su uno dei 4 Agenti accreditati (Agenti con codici 30, 31, 38, 45 per motivi di privacy) che risultavano dal database regionale.

Il modello multilevel finale viene specificato nella (7)

$$y_{ij} = \beta_{0j} + \gamma_j(z)_{ij} + \beta_2(inst)_{ij} + \beta_3(mot)_{ij} + \beta_4(min)_{ij} + I_{ij}\varepsilon + r_{ij} \quad j=0, \dots, k \quad i=1, \dots, n_j$$

$$\beta_{0j} = \mu + I_{ij} m_j \quad \gamma_j = \pi D_j + I_{ij} u_j \quad (7)$$

Per mancanza di informazioni disponibili non vengono inserite ulteriori variabili (costi, dimensione, personale) relative agli Agenti che hanno erogato assistenza.

Stimando il modello senza variabili esplicative, inserendo la variabile Agente come unico regressore (casuale) è emersa una correlazione intraclassa pari a 0.18, giustificando in tal modo un'analisi multilivello; poichè circa il 18% della variabilità dello stato di dipendenza cognitivo dipende dal particolare tipo di Agente che ha preso in carico l'anziano, si conferma che il processo di valutazione non può prescindere dalla struttura gerarchica (multilivello) dei dati.

In tavola 1 sono mostrate le stime delle componenti di varianza del modello, in tavola 2 le stime dei parametri fissi e in tavola 3 le stime dei parametri casuali (efficacia relativa), con i relativi estremi degli Intervalli di confidenza (IC) al 95%.

TAVOLA 1

*Stime delle componenti di varianza*

Varianze	Stime	St. Error	Z	Pr Z
$\theta_m$	1.104	1.141	0.97	0.1665
$\theta_z$	4.816	1.12	4.3	<.0001
$\sigma^2$	10.7587	0.6253	17.21	<.0001

La tavola 1 mostra che la componente di variabilità valutata tra i diversi Agenti ( $\theta_m$ ) non è più significativa, lasciando intendere che, al netto delle variabili esplicative inserite nel modello, i valori medi di dipendenza cognitiva degli anziani in carico ai diversi Agenti hanno una componente di variabilità pressochè nulla; diversamente risultano ancora fortemente variabili sia i punteggi cognitivi dei pazienti all'interno degli Agenti ( $\sigma^2$ ) sia i coefficienti che valutano la relazione tra lo stato cognitivo iniziale e quello finale tra i vari Agenti ( $\theta_z$ ).

Le stime delle covarianze (non mostrate in tabella) non risultano significative.

Nella tavola 2, che mostra la stima degli effetti fissi del modello, l'impatto dell'intervento, valutata dal parametro  $\varepsilon$ , risulta positivo e statisticamente significativo per livelli di significatività più elevati di quelli standard ( $p=0.09$ ), inducendo dunque una certa cautela sulla conclusione che lo stato cognitivo degli anziani beneficiari del Bss sia superiore rispetto al gruppo di anziani in Rsa; appare più conservativa la tesi che non esiste sostanziale differenza tra lo scadimento cognitivo tra beneficiari del Bss ed anziani in Rsa, contrariamente a quanto affermato in letteratura per giustificare il trasferimento degli anziani in Rsa (Tsuij *et al.*, 1995).

TAVOLA 2

*Stime degli effetti fissi (in parentesi la descrizione dei parametri)*

Parametri Fissi	Stima	St. Error	t	Pr >  t
$\mu$	-2.1129	0.9895	-2.14	0.0858
$\gamma_{10}$	0.5855	0.0427	13.69	<.0001
$\beta_2$ (Instabilità clinica: 1 vs 4)	1.9940	0.8796	2.27	0.0238
$\beta_2$ (Instabilità clinica: 2 vs 4)	1.7279	0.8888	1.94	0.0524
$\beta_2$ (Instabilità clinica: 3 vs 4)	1.5752	0.9713	1.62	0.1054
$\beta_3$ (Fim motoria)	0.1075	0.0097	11.05	<.0001
$\beta_4$ (Minuti di assistenza)	-0.8802	0.0943	-2.38	0.0072
$\varepsilon$ (Bss vs RSA)	0.4753	0.2804	1.69	0.0905
$\pi_0$ (Tipo Agente: Famiglia)	0.9548	0.1089	8.76	<.0001
$\pi_1$ (Tipo Agente: Accreditati)	0.8040	0.1851	4.34	<.0001
$\pi_2$ (Tipo Agente: No Accreditati)	0.8087	0.1876	4.31	<.0001

In particolare risultano altamente significative la Fim motoria ( $t=11.05$ ) e soprattutto il valore basale cognitivo ( $t=13.69$ ), mentre emerge un sostanziale miglioramento cognitivo per chi ha una bassa instabilità clinica, rispetto a chi ha una forte instabilità clinica (contrasto 1 vs 4,  $t=2.27$ ).

Infine le interazioni tra tipo di Agente e stato cognitivo iniziale ( $\pi$ ) mostrano che la famiglia assicura tassi di miglioramento più elevati rispetto alle altre tipologie di Agenti, poiché si fa carico tipicamente di anziani meno compromessi dal punto di vista cognitivo: tale considerazione costituisce un importante fattore da tenere in considerazione per la valutazione di efficacia relativa.

Rispetto all'efficacia relativa, la tavola 3 mostra che gli effetti degli Agenti né sono statisticamente differenti dall'efficacia media dell'intervento rispetto al non intervento, né esistono sostanziali differenze tra le performances di efficacia tra i 6 Agenti, rispetto all'efficacia media dell'intervento, come evidenziato dagli intervalli di confidenza dei residui  $m_j$  di II livello dell'equazione (7): infatti, in una logica cautelativa, si ritiene l'Agente  $\gamma$  più efficace dell'Agente  $j$  se l'estremo inferiore dell'intervallo di confidenza del residuo di II livello per  $\gamma$  è superiore all'estremo superiore dell'intervallo di confidenza dell'Agente  $j$ .

TAVOLA 3

*Stime degli effetti casuali (efficacia degli Agenti)*

Agente	Stima	St. error	t	Pr >  t	IC 95%
0	-2.2295	1.1403	-1.96	0.0511	(-4.4697 , 0.0106)
1	-2.1034	1.2214	-1.72	0.0856	(-4.5029 , 0.2960)
30	1.5656	1.6780	0.93	0.3513	(-1.7311 , 4.8622)
31	2.6042	1.5891	1.64	0.1019	(-0.5177 , 5.7260)
38	1.1978	1.4267	0.84	0.4016	(-1.6051 , 4.0006)
45	-1.0346	1.3182	-0.78	0.4329	(-3.6243 , 1.5552)

La conclusione che la famiglia non risulti più efficace degli altri Agenti sembra in contraddizione con il fatto che i tassi di miglioramento per i pazienti curati in famiglia siano maggiori rispetto alle altre forme di assistenza, come visto precedentemente dalla significatività di  $\pi_0$ ; tale apparente contraddizione deriva in larga parte dal fatto che i pazienti accuditi in famiglia sono meno compromessi dal punto di vista cognitivo, rispetto ai pazienti in carico ad erogatori accreditati; secondo la logica di valutazione *coeteris paribus* il modello multilivello riesce quindi ad eliminare questo fattore che inficia i confronti di efficacia tra gli Agenti, stimando l'efficacia della famiglia in maniera non più significativa.

Per concludere quindi non vi sono sostanziali differenze nel tipo e nel grado di autosufficienza, e quindi nemmeno nel tempo medio assistenziale per le attività della vita quotidiana, visto l'alto legame tra punteggi Fim e carico assistenziale (Granger *et al.*, 1993), fra anziani assistiti in famiglia e anziani in Rsa: nell'insieme la formula del Bss è risultata appropriata, intervenendo a remunerare nuclei familiari che forniscono una forma di assistenza sì generica, ma costante e continua che altrimenti nessun altro intervento di carattere sociale potrebbe garantire.

## 5. CONCLUSIONI E PROSPETTIVE

Rispetto ai modelli che valutano l'efficacia di impatto sono auspicabili metodologie che selezionino attentamente il gruppo di controllo attraverso metodologie che siano sensibili a possibili errori da selezione dovuti a distorsioni di appaiamento tra i due campioni, soprattutto in tipiche situazioni di dati osservazionali e piccoli campioni: una via alternativa al *Propensity score* è quella basata sulla specificazione di modelli di regressione con parametri che riflettono il grado di non casualità delle osservazioni (Maddala, 1983).

Rispetto invece alla valutazione di efficacia relativa sono auspicabili metodologie che, sotto il profilo interpretativo, consentano di valutare l'efficacia di più Agenti erogatori sulla base della simultanea analisi di più risultati che descrivono aspetti differenti, ma collegati dello stato di benessere; l'alternativa di costruire graduatorie basate sul confronto di ogni indicatore di risultato preso singolarmente non sembra la strada più efficiente dal punto di vista statistico a causa della distorsione e la perdita di efficienza degli stimatori (Heagerty e Zeger, 1996).

Le proposte multivariate a tale scopo quali i modelli di regressione marginali multivariati con variabili miste (Liang *et al.*, 1992; Carey *et al.*, 1993; Heagerty e Zeger, 1996) o gli approcci di Item Response Theory nell'ambito dei modelli Rasch in ottica Multilevel per analizzare in un approccio unificante la stima di outcomes latenti con dati gerarchici (Gibbons e Hedeker, 2000) non sembrano una valida alternativa, causa l'impossibilità di introdurre parametri di interazione tra le covariate a differenti livelli, ponendo dubbi sull'interpretabilità dei parametri di efficacia.

Per tener conto di più outcome recentemente sono stati proposti modelli ad equazione SURE in ottica multilevel (Vittadini, 1999) che prevedono un modello di regressione ad effetti random per ogni outcome e la specificazione di legami stocastici tra differenti equazioni (correlazione contemporanea tra più outcome per ogni individuo).

Un'ulteriore importante questione è la giustificazione empirica dell'uso di modelli a effetti casuali rispetto ai modelli con effetti fissi soprattutto alla luce della rigidità delle ipotesi sugli errori (normalità distributiva e indipendenza tra errori di I e di II livello) e la sovrabbondanza di parametri al crescere della numerosità dei gruppi di II livello che provoca la distorsione e l'inefficienza delle stime che possono affliggere i modelli multilivello (Lancaster, 2000).

Inoltre uno dei maggiori problemi che affliggono i modelli che valutano l'evoluzione delle performances (risultati) rispetto ad uno stato iniziale è dovuto alla forte correlazione esistente tra stato finale e stato iniziale che riduce la potenziale capacità esplicativa di possibili covariate, (incluso anche l'effetto dell'Agente e dunque la valutazione dell'impatto) sul valore aggiunto (Allison, 1990).

Nella letteratura anglosassone, in ambito educativo (Thum, 2001; Rowan *et al.*, 2002) è in corso il dibattito su quale modello statistico sia più adeguato in una logica di valutazione; recentemente alcune tendenze stanno soppiantando i modelli di "valore aggiunto" (che cercano di spiegare i risultati attuali in funzione dei risultati precedenti), a favore di approcci basati sui modelli di crescita o *growth curves*, (Rogosa *et al.*, 1982): tale approccio, che richiede comunque osservazioni per ciascun utente all'interno degli Agenti in almeno tre occasioni temporali, consentirebbe di formalizzare in maniera più oggettiva il concetto di valore aggiunto (Hanushek, 2002), oltre al fatto di descrivere simultaneamente, sia il livello iniziale (stato cognitivo iniziale), sia la relativa variazione nel tempo, attraverso la specificazione di due diverse equazioni, ponendo le misure di follow-up e i valori basali sullo stesso piano.

## RIFERIMENTI BIBLIOGRAFICI

- M. AITKIN, N. LONGFORD (1986), *Statistical modelling issue in school effectiveness studies*, "Journal of the Royal Statistical Society", A, 149, pp. 1-43.
- P. ALLISON, (1990), *Change scores as dependent variables in regression analysis*, in C. CLOGG (eds.) *Sociological Methodology*, Jossey-Bass, San Francisco, pp. 93-114.
- R. BELLIO, E. GORI, (2000), *La valutazione dell'efficacia dei corsi di formazione professionale mediante modelli multilivello in presenza di possibile selezione non casuale*, AIEL, Ancona, Settembre 2000.
- V. CAREY, S.L. ZEGER, P. DIGGLE, (1993), *Modelling multivariate binary data with alternating logistic regressions*, "Biometrika", 80, pp. 517-526.
- J.-B. COPAS, H.G. LI, (1997), *Inference for non-random samples*, "Journal of Royal Statistical Society", B, 59, pp. 55-95.
- J.-B. COPAS, H.G. LI, S. EGUCHI, (2001), *Local sensitivity approximations for selectivity bias*, "Journal of the Royal Statistical Society" Series B, 63, pp. 871-895.
- D.-R. COX, P. MCCULLAGH, (1982), *Some Aspects of Analysis of Covariance*, "Biometrics", 38, pp. 541-561.
- CRISP (2001), *Osservatorio Regionale sulla qualità dei servizi sanitari*, www.crisp.org.
- C. DAGUM, G. VITTADINI, (1996), *Human Capital Measurement and Distribution*, Proceedings of the 156<sup>th</sup>, Meeting of the American Statistical Association, Business and Economic Statistics Section.
- R.-W. DUBOIS, R.-H BROOK, W.-H ROGERS, (1988), *Adjusted hospital death rates: A potential screen for quality of medical care*, "American Journal of Public Health", 77, pp. 1162-1177.
- A. EPSTEIN, (1995), *Performance reports on quality-prototypes, problems and prospects*. "New England Journal of Medicine", 333, pp. 57-61.
- A. EPSTEIN, F.-C. SCHNEIDER, (1997), *Influence of cardiac-surgery performance reports on referral practices and access to care*. *New England Journal of Medicine* 335, 251-266.
- T. FITZ, C.T. GIBBON, (1997), *The Value Added National Project*, Final Report, CEM e University of Durham.
- J. FLEISS (1986), *The study and analysis of clinical experiments*, Wiley, New York.
- R.-D. GIBBONS, D. HEDEKER, (2000), *Applications of mixed-effect models in biostatistics*, "Sankhya", 62, pp. 70-103.
- H. GOLDSTEIN, D.J. SPIEGELHALTER, (1996), *League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance*, "Journal of the Royal Statistical Society", A, 159, pp. 385-443.
- E. GORI, (2000), *La valutazione dei corsi di formazione professionale in Lombardia*, "Lavoro e Relazioni Industriali", 1, pp. 95-109.
- E. GORI, G. VITTADINI, (1999), *La valutazione dell'efficienza ed efficacia dei servizi alla persona. Impostazione e metodi*, in E. GORI, G. VITTADINI, (eds.), *Qualità e valutazione nei servizi di pubblica utilità*, ETAS, Milano.
- C.-V. GRANGER, B.-B. HAMILTON, J.-M. LINACRE, A.W. HEINEMANN, B.-D. WRIGHT, (1993), *Performance profiles of the Functional Independence Measure*, "Archives of Physical".
- E.-A. HANUSHEK, (1971), *Teacher characteristics and gains in student achievement: estimations using micro data*, "American Economic Review", 61, pp. 280-288.
- E.-A. HANUSHEK (2002), *Publicly Provided Education*, in A. AUERBACH and M. FELDSTEIN (Eds.), "Handbook of Public Economics", Amsterdam: North-Holland, pp. 2047-2143.
- P.-J. HEAGERTY S.L. ZEGER, (1996), *Marginal regression models for clustered ordinal measurements*, "Journal of the American Statistical Association", 91, pp. 1024-1036.
- J. HECKMAN, (1976), *The common structure of statistical models of truncation, sample selection and limi-*

- ted dependent variables and a simple estimator of such models, "Annals of Economic and Social Measurement", 5, pp. 475-492.
- J. HECKMAN, V.-J. HOTZ (1989), *Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training* (with discussion), "Journal of the American Statistical Association", 84, pp. 862-874.
- J. HECKMAN, R. ROB, (1986), *Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes* (with discussion), in H. WAINER (eds.) *Drawing Inferences from Self-Selected Samples*, pp. 57-113, Springer, New York.
- C.-J. HEINRICH, E.-L. LYNN, (2001), *Governance Matters: The Influence of Program Structure and Management on Job Training Partnership Act (JTPA) Performance*, Irving B. Harris Graduate School of Public Policy Studies, University of Chicago.
- ISFOL, (2001), *Formazione e Lavoro: Effetti del Fondo Sociale Europeo sull'occupabilità in Italia*, Struttura di Valutazione del Fondo Sociale Europeo, Roma.
- T. LANCASTER, (2000) *The incidental parameter problem since 1948*, "Journal of Econometric", 95 pp. 391-413.
- I.-H. LANGFORD, T. LEWIS, (1998), *Outliers in multilevel data* (with discussion), "Journal of the Royal Statistical Society" Series A, 161, pp. 121-160.
- K.-Y. LIANG, S.-L. ZEGER, B. QAQISH, (1992), *Multivariate regression analyses for categorical data*, "Journal of the Royal Statistical Society", 54, pp. 3-40.
- P.-G. LOVAGLIO, (2001), *La stima di outcome latenti*, "Atti del Convegno Intermedio della Società Italiana di Statistica, Processi e Metodi Statistici di Valutazione", Cisu, Roma, pp. 393-396.
- P.-G. LOVAGLIO, (2004a), *The estimate of customer satisfaction in a reduced rank regression framework*, "Total Quality Management", 1, pp. 33-44.
- P.-G. LOVAGLIO, (2004b), *La stima di outcome clinici con la Rasch Analysis*, "Statistica", in press.
- G.-S MADDALA, (1983), *Limited dependent and qualitative variables in econometrics*, Cambridge Univ. Press, New York.
- E.-C MARSHALL, D.-J. SPIEGELHALTER, (2001), *Institutional Performance*, in Goldstein H., Leyland A.H. (eds.), *Multilevel Modelling of Health Statistics*, Wiley, New York, pp. 127-142.
- L.-M. MEAD, (1997), *Optimizing JOBS: Evaluation Versus Administration*, "Public Administration Review", 57 pp. 113-123.
- L.-M. MEAD, (1999), *Performance Analysis*, Unpublished manuscript, New York University.
- H.-B. MILWARD, K. PROVAN, (2000), *How Networks are Governed*, in E.-L. LYNN, C.-J. HEINRICH, (eds.), *Governance and Performance: Models, Methods and Results*, Washington, DC: Georgetown University Press.
- V. MORAMARCO, (1999), *Qualità ed efficienza economica*, in Gori E., Vittadini G. (Eds.) *Qualità e valutazione nei servizi di pubblica utilità*, ETAS, Milano.
- B. MUTHEN, G. SPECKART, (1985), *Latent variable probit ANCOVA: Treatment effects in the California Civil Addict Program*, "British Journal of Mathematical and Statistical Psychology", 38, pp. 161-170.
- ORML, (1997), *Monitoraggio e valutazione ex post dell'attività di formazione professionale. Anno formativo 1993/94*, Flash Lavoro, Quaderni 46, Regione Toscana.
- N. RICE, A. LEYLAND, (1996), *Multilevel models: applications to health data*, "Journal of health services research and policy", 1, pp. 154-164.
- M. RODERICK, B. JACOB, A. BRICK, (2000), *Evaluating Chicago's Efforts to End Social Promotion*, in E.-L. LYNN, C.-J. HEINRICH, (eds.), *Governance and Performance: Models, Methods and Results*, Washington, DC: Georgetown University Press.
- D.-R. ROGOSA, D. BRAND, M. ZIMOWSKI, (1982), *A growth curve approach to the measurement of change*, "Psychological Bulletin", 90, pp. 726-748.

- P. ROSENBAUM (1995) *Observational Studies*, Springer, New York.
- P. ROSENBAUM, D.-B. RUBIN, (1983) *The central role of the propensity score in observational studies*, "Biometrika", 70, pp. 41-55.
- P. ROSENBAUM, D.-B. RUBIN, (1984), *Reducing bias in observational studies using the subclassification on the propensity score*, "Journal of the American Statistical Association", 79, pp. 516-524.
- B. ROWAN, R. CORRENTI, R.-J. MILLER, (2002), *What Large-Scale, Survey Research Tells Us About Teacher Effects On Student Achievement: Insights from the Prospects Study of Elementary Schools*, "CPRE Research Report Series" 051, Consortium for Policy Research in Education.
- K.-B. SMITH, K.-J. MEIER, (1994), *Politics, Bureaucrats and Schools*, "Public Administration Review", 54, pp. 551-558.
- L. STIEFEL, (1997), *Measurement of Output Quality in US non-profit organization*, "Atti del Convegno SIS, La statistica per le imprese", 1, pp. 355-370.
- L. TESIO, F. FRANCHIGNONI, L. PERUCCA, G. PORTA, (1996), *The influence of age on length of stay, functional independence and discharge destination of rehabilitation in patients in Italy*, "Disability Rehabilitation", 18, pp. 502-508.
- L. TESIO, L. PERUCCA, M.-A. BATTAGLIA, F. FRANCHIGNONI, (1997), *Quality assessment of the FIM (Functional Independence Measure) Ratings Through Rasch Analysis*, "Europa medicophysica", 33, pp. 55-57.
- Y.-M. THUM, (2002), *Measuring Progress towards a Goal: Estimating Teacher Productivity using a Multivariate Multilevel Model for Value-Added Analysis*, "Sociological Methods & Research", 32, pp. 153-207.
- I. TSUIJ, S. WHALEN, T.-E. FINUCANE, (1995), *Predictors of nursing home placement in community base long term care*, "Journal of American Geriatric Society", 43, pp. 761-766.
- R.-M. TURNER, R.-Z. OMAR, M. YANG, H. GOLDSTEIN, S.-G. THOMPSON, (2000), *Multilevel models for meta-analysis of clinical trials with binary outcomes*, "Statistics in Medicine", 19, pp. 3417-3432.
- G. VITTADINI, (1999), *On the use of Multivariate Regression models in the context of Multilevel Analysis*, convegno "Classification and Data Analysis Group" Riunione scientifica Società Italiana di Statistica, Roma.
- G. VITTADINI, P.-G. LOVAGLIO, (2001), *The estimate of latent variables in a structural model an alternative approach to PLS*, 2ND International Symposium on PLS and related methods, pp. 423-434, Cisia-Ceresta, Montreuil, France.
- J.E. WARE, (1993), *Sf-36 Health Survey. Manual and interpretation Guide*, The Health Institute, New England Medical Center.

## RIASSUNTO

### *Efficacia relativa e di impatto di iniziative nell'ambito dei servizi alla persona di pubblica utilità*

Il presente lavoro affronta in dettaglio la specificazione di un modello che valuta congiuntamente l'efficacia di iniziative o progetti Istituzionali in ambito clinico-epidemiologico che prevedono l'erogazione di un servizio alla persona e l'efficacia di una serie di agenzie (Agenti) che erogano lo stesso servizio secondo modalità differenti, fornendo il razionale per una corretta procedura di valutazione oggettiva, al netto dei fattori che possono distorcere i risultati; infine si procederà a valutare l'efficacia di un intervento sperimentale promosso dalla Regione Lombardia nel 2002, consistente in un contributo economico a favore degli anziani non autosufficienti assistiti in famiglia, come alternativa al ricovero in Residenze Socio Assistenziali (Rsa).

## SUMMARY

### *Relative and impact evaluation of clinic and epidemiologic framework*

The present paper addresses in detail the specification of a model that values the effectiveness of institutional plans within clinical or epidemiologist projects previewing the supply of a service to the person and jointly the effectiveness of agencies (Agents) that distribute the same service with different modalities, supplying the rationale for a rigorous procedure of objective evaluation.

Finally the model proposed is applied to a public plan promoted from the Regione Lombardia in 2002, consisting in an economic contribution to family that decide to assist in house the not self-sufficient old, like alternative to the shelter in Residences.