

GIORGIO VITTADINI, PAOLO BERTA
E LUCA GIUSEPPE MERLINO

LA VALUTAZIONE DEGLI ESITI PER SPECIALITÀ: INTRODUZIONE METODOLOGICA

1. *La valutazione dei servizi sanitari*

Nel corso degli ultimi decenni, a livello internazionale si è preso coscienza del fatto che la valutazione della qualità è essenziale per il miglioramento e per la sostenibilità economica dei Sistemi Sanitari [Donabedian 1992; Opit 1993] sia nei termini della valutazione *ex ante*, inerente il processo produttivo attraverso cui si giunge a erogare il servizio, che della valutazione *ex post* che prende in considerazione i risultati ottenuti.

All'origine della valutazione *ex ante* in sanità si colloca il programma «Minimum Standards for Hospitals» proposto dall'associazione dei chirurghi nord-americani nel 1917 con lo scopo di standardizzare la struttura ed il modo di lavorare degli ospedali, nonché valorizzare le istituzioni migliori favorendo l'adeguato riconoscimento davanti alla comunità professionale e motivare le istituzioni con standard inferiori a migliorare la qualità del loro lavoro [Merlino e Vittadini 2008].

Dalla diffusione e sviluppo di questo programma nei decenni successivi si sono creati i presupposti che portarono nel 1951 alla fondazione, promossa congiuntamente da alcune associazioni professionali e dall'associazione statunitense degli ospedali, della Joint Commission on Accreditation of Hospitals (JCAHO), che ancora oggi è una delle più prestigiose istituzioni a livello internazionale in questo campo ed il cui nome è oggi divenuto Joint Commission on Accreditation of Healthcare Organizations (JCAHO).

In questa linea il Servizio Sanitario pubblico e i sistemi assicurativi, soprattutto negli ultimi 10-15 anni, hanno dato vita a programmi di accreditamento istituzionale aventi lo scopo di definire, valutare e regolamentare standard mi-

nimi gestionali, strutturali, tecnologici ed organizzativi come condizione per l'ammissione a rapporti contrattuali delle strutture sanitarie [Merlino e Vittadini 2008].

Tuttavia l'accreditamento istituzionale non è sufficiente da solo per promuovere un reale miglioramento della qualità delle strutture sanitarie.

2. La valutazione «ex post»: tipologie ed esempi nel mondo

Quando si affronta il tema della valutazione *ex post*, tre sono le dimensioni di analisi su cui si è maggiormente orientata la letteratura scientifica: efficienza, efficacia e *patient satisfaction*. Per definirle correttamente e analizzare quali siano le metodologie più adatte ad attuarle, occorre soffermarsi sulla distinzione tra output e outcome sanitario [Gori e Vittadini 1999; Vittadini *et al.* 2003; Pagano e Vittadini 2004]. L'output, legato agli input e al processo produttivo necessari per erogarlo, è misurabile, nel momento stesso in cui è fornito, non rispetto ai pazienti ma alla struttura sanitaria, in termini tecnici (es. numero e tipologia dei pazienti, dei giorni di degenza, delle operazioni chirurgiche, degli esami radiografici) o in termini monetari.

L'insieme degli output, invece, genera un particolare outcome sanitario a livello paziente, come risultato durevole e globale dei servizi sanitari erogati sullo stato di salute del paziente [Lohr 1990; Donabedian 2005].

L'analisi di efficacia [Donabedian 2005] ha a che fare con l'outcome e può essere definita in diversi modi [Gori e Vittadini 1999]: come impatto ovvero come confronto tra i risultati ottenuti dalla prestazione sanitaria e quelli ottenibili in sua assenza; come standard di qualità raggiunto rispetto a un target prestabilito; come rapporto tra domanda soddisfatta e potenziale; come rapporto tra risultati ottenuti e obiettivi prefissi; come confronto tra i risultati ottenuti da una struttura sanitaria rispetto a quelli ottenuti dalle altre strutture sanitarie (efficacia relativa).

In questo quadro così ricco di valutazioni *ex post* occorre chiarire i criteri di scelta di una particolare metodologia di valutazione.

3. Metodi per valutare l'efficacia relativa

Tra i diversi tipi di efficacia ricordati quello più adeguato per confrontare l'efficacia di diverse strutture sanitarie nell'erogare un medesimo servizio è quello di efficacia relativa. Per operare un confronto corretto, non affetto da distorsione, occorre tenere conto di alcuni aspetti fondamentali [Gori e Vittadini 1999; Pagano e Vittadini 2004; Vittadini *et al.* 2004]:

a) l'asimmetria informativa legata alla non conoscenza della gravità degli assistiti: se la si ignorasse si rischierebbe non solo di confrontare strutture sanitarie in modo distorto (perché un outcome può essere giudicato più o meno buono a secondo della maggiore o minore gravità iniziale del paziente) ma anche di provocare fenomeni di selezione avversa, eticamente inaccettabile, quando le strutture sanitarie, per avere una migliore valutazione, scelgono di curare preferibilmente o esclusivamente i pazienti meno gravi [Goldstein e Spiegelhalter 1996]. Occorre, perciò, effettuare confronti a parità di condizioni (o come spesso si dice *ceteris paribus*), basati sul valore dell'outcome sanitario del paziente «corretto» mediante metodologie denominate di risk adjustment [Goldstein e Spiegelhalter 1996] in modo da tener conto del suo «case mix», vale a dire delle sue caratteristiche cliniche e socio-demografiche;

b) la personalizzazione degli outcome per ogni paziente: si devono utilizzare metodologie statistiche che costruiscano l'informazione inerente all'efficacia media della struttura sanitaria a partire dai microdati inerenti outcome sanitari «corretti» dei singoli pazienti;

c) la struttura gerarchica dei dati: ogni paziente è ricoverato in un determinato reparto appartenente a un dato ospedale [Vittadini 2010; Berta *et al.* 2012; Berta *et al.* 2013; Vittadini *et al.* 2003].

La metodologia più semplice di risk adjustment è la cosiddetta «standardizzazione» [Zaslavsky 2001]. Considerando la popolazione di pazienti, suddivisi per patologia in g gruppi ($w = 1, \dots, g$), si costruisce per ciascun ospedale j l'outcome «standardizzato» y_j come somma ponderata degli outcome y_{wj} relativi ai diversi gruppi, con pesi pari alla proporzione di utenti π_{wj} di ogni gruppo (standardizzazione diretta) o ottenuti da una teorica popolazione standard (standardizzazione indiretta)¹:

$$y_j = \sum_w \pi_{wj} y_{wj} \quad (1)$$

Tuttavia la standardizzazione non tiene conto dei problemi di risk adjustment, non parte dagli outcome personalizzati per paziente, non considera la struttura gerarchica dei dati.

Un approccio più adeguato è rappresentato dai modelli lineari e logistici, utilizzati dalle principali agenzie di benchmark sanitario (CIHI [2009]; JCAHO²; NHS³; PNE⁴; AHRQ⁵).

Con outcome dicotomici, dati $X_1, \dots, X_m, \dots, X_q$ indicatori che descrivono il case mix di caratteristiche del paziente, si utilizza il modello logistico:

$$\ln((p_{ij} / (1 - p_{ij}))) = \alpha + \sum_m \beta_m x_{ijm} + RS_{ij} \quad (2)$$

ove (2) rappresenta l'odd, logaritmo del rapporto tra la probabilità che si verifichi l'outcome favorevole p_{ij} e la

¹ Ad esempio nell'ambito del programma di benchmark della Commission for Health Improvement (CHI) del National Health Service [NHS 2004] si propone la standardizzazione indiretta di alcuni outcome clinici.

² <http://www.jointcommission.org/>.

³ <https://www.england.nhs.uk>.

⁴ <http://95.110.213.190/PNEed14/>.

⁵ I quality indicators misurati dall'AHRQ si basano sul rapporto tra eventi osservati ed eventi attesi. Questi ultimi sono calcolati in termini di risk adjustment applicando modelli di regressione corretti per sesso, età, DRG e comorbidità (<http://www.ahrq.gov>).

probabilità che si verifichi l'outcome avverso $(1 - p_{ij})$. RS_{ij} indica il cosiddetto risk score associato all' i -esimo paziente della j -esima struttura. La misura dell'efficacia relativa della j -esima struttura sanitaria è data dalla media degli odd (2) per tutti gli n_j pazienti curati da essa.

Il modello logistico si basa su outcome personalizzati, corretti da variabili di risk adjustment, ma non tiene conto della struttura gerarchica dei dati sanitari e ciò comporta una serie di problemi metodologici. La struttura gerarchica dei dati fa sì che non valga più l'ipotesi, ritenuta valida nel modello logistico (2), secondo cui le singole osservazioni sono indipendentemente distribuite, in quanto esiste dipendenza distributiva fra le osservazioni inerenti pazienti che si curano nello stesso ospedale [Hox 1995].

Inoltre, poiché negli usuali modelli lineari e logistici si ignora la struttura gerarchica dei dati, si incorre: nell'*ecological fallacy* [Robinson 1950], se si considerano solo dati aggregati a livello di ospedale [Hox 1995]; nell'*atomistic fallacy*, se si effettua l'analisi a partire da dati osservati a livello di pazienti. In tal caso, quanto maggiore è la correlazione tra le unità entro gli ospedali, tanto più elevata è la distorsione nella stima degli errori standard dei parametri [Snijders e Bosker 2011; Leyland e Groenewegen, 2003].

Perciò, numerosi autori [Dubois *et al.* 1987; Jencks *et al.* 1988; Thomas *et al.* 1994; Normand *et al.* 1995; Epstein 1995; Schneider e Epstein 1996; Morris e Christiansen 1996; Goldstein e Spiegelhalter 1996; Rice e Leyland 1996; Leyland e Boddy, 1998; Marshall e Spiegelhalter 2001; HCUP 2007] hanno proposto di utilizzare il modello multilevel [Hox 1995; Goldstein 2003], modello di covarianza a effetti misti per dati gerarchici, per la valutazione dell'efficacia degli ospedali⁶. Si ha quindi per l' i -esimo paziente curato nel j -esimo ospedale:

$$y_{ij} = \sum_m \beta_m x_{ijm} + u_j + e_{ij} \quad (i = 1, \dots, n; j = 1, \dots, p; m = 1, \dots, q) \quad (3)$$

⁶ Ash *et al.* [2012], in collaborazione con Centers for Medicare and Medicaid Services, raccomanda l'uso di modelli Multilevel per la valutazione di outcome quali mortalità e ricoveri ripetuti.

ove y_{ij} è l'outcome relativo all' i -esimo paziente curato nella j -esima struttura sanitaria; x_{ijm} è la m -esima variabile esplicativa di risk adjustment relativa all' i -esimo paziente curato nella j -esima struttura sanitaria e β_m è il parametro ad essa connesso; e_{ij} è il residuo casuale, u_j è una variabile casuale che descrive l'efficacia relativa del j -esimo agente.

Si ipotizza distribuzione normale dei parametri casuali e_{ij} e u_j e di indipendenza e quindi incorrelazione fra tali distribuzioni di e_{ij} e u_j :

$$e_{ij} \sim N(0, \sigma_j^{2*}); u_j \sim N(\sigma_j^{2^*}); \text{cor}(e_{ij}; u_j) = 0 \quad (4)$$

Si ipotizza inoltre incorrelazione tra variabili di risk adjustment e parametri di efficacia in modo che l'efficacia degli ospedali dipenda esclusivamente dai parametri di efficacia u_j (da verificare attraverso opportuni test di Hausman [Snijders e Bosker 2011])⁷.

La varianza degli e_{ij} nel j -esimo ospedale σ_j^{2*} rappresenta la varianza residua dell'outcome corretto per le variabili di risk adjustment, mentre $\sigma_j^{2^*}$ è la varianza spiegata, ove per (4):

$$\sigma_j^2 = \sigma_j^{2*} + \sigma_j^{2^*} \quad (5)$$

Date le medie delle varianze residue σ^{2*} e spiegate σ^{2^*} per tutti i p ospedali:

$$\sigma^2 = \sigma^{2*} + \sigma^{2^*} \quad (6)$$

il loro rapporto denominato coefficiente intraclasse, ci dice quale è la quota della varianza totale dell'outcome corretto per le variabili di risk adjustment attribuibile all'efficacia degli ospedali:

⁷ Un ulteriore pregio del modello multilevel sta nel fatto che grazie all'introduzione degli *shrinkage residuals* un sottocampione di osservazioni relativo ai pazienti di un ospedale pesa sulle stime complessive in proporzione alla sua numerosità rispetto alla numerosità della popolazione [Goldstein 2003].

$$\rho = \sigma^2 / \sigma^2 \quad (7)$$

Tale metodo per confrontare l'efficacia relativa degli ospedali permette di «probabilizzare» il loro ranking [Goldstein e Spiegelhalter 1996]. Infatti, in questo modo, in luogo di basarsi solo su medie campionarie dei parametri di efficacia si paragonano loro intervalli di confidenza entro cui (con un livello di probabilità del 90%, 95%, 99%) giacciono i valori attesi della distribuzione della loro popolazione. Inoltre, se l'ampiezza dell'intervallo di confidenza di un ospedale è elevata segnala una forte variabilità dell'efficacia al suo interno e quindi pone delle cautele sulla effettiva validità informativa dell'efficacia media riscontrata nell'ospedale. Si possono visualizzare i ranking di ospedali attraverso il grafico in figura 1, in questo caso relativo alla mortalità corretta.

La proposta più «classica» di Goldstein e Spiegelhalter [1996] è basata su un modello statistico che analizza l'efficacia relativa di 87 chirurghi operanti nello Stato di New York rispetto ad un piccolo campione di pazienti su cui viene effettuato un bypass coronarico. L'outcome è la mortalità (dicotomica), le variabili esplicative sono definite solo a livello di pazienti. Il modello Multilevel per la valutazione dell'efficacia è stato in seguito proposto in versione più complessa, ad esempio introducendo variabili di risk adjustment di secondo livello (es. ospedali) [Normand *et al.* 1997], parametri di interazione tra diverse variabili esplicative [Leyland e Boddy 1998], versioni di tipo longitudinale [Rowan *et al.* 1993].

Numerosissimi sono gli esempi di applicazione del Multilevel model per la valutazione di strutture sanitarie come notato da una ricerca di Houchens e colleghi [Houchens *et al.* 2007] che parla di 650 articoli pubblicati già fino a quell'anno e da svariati altri autori [e.g. Rice e Leyland 1996; Duncan 1998; Leyland e Goldstein 2001; Goldstein *et al.* 2002]. Soprattutto l'HCUP [2007] ha suggerito e promosso l'uso del Multilevel Model per la valutazione delle strutture sanitarie e molti autori l'hanno concreta-

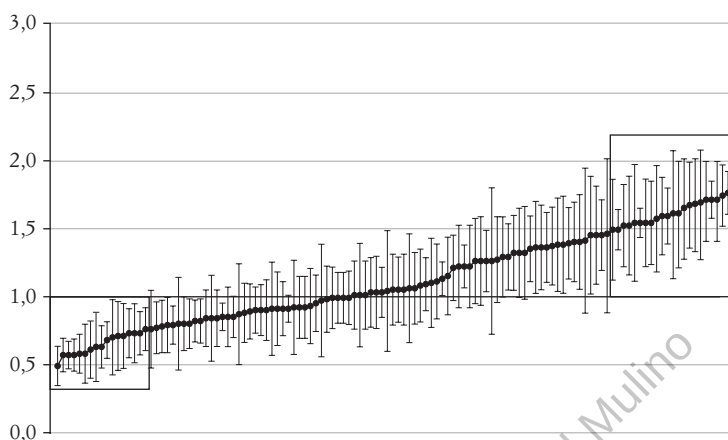


FIG. 1. Esempio di ranking degli ospedali. Mortalità totale: intervalli di confidenza per ospedale.

mente applicato a dati HCUP [e.g. Localio *et al.* 1997; Amaravadi *et al.* 2000; Spiegelhalter *et al.* 2002; Carey 2002; Scott *et al.* 2004]⁸.

Anche il più recente *White Paper* [Ash 2012] promosso dal «Centers for Medicare and Medicaid Services» in collaborazione con le principali società scientifiche statistiche americane, ha riconosciuto come l'approccio metodologico basato sui modelli Multilevel sia da preferire quando si effettua una valutazione degli outcome ospedalieri.

4. Critiche ai metodi di «risk adjustment» e nuove proposte

Tuttavia, l'uso degli outcome per la costruzione di graduatorie delle strutture sanitarie è stato fortemente

⁸ La Care Quality Commission (CQC, <http://www.cqc.org.uk>) inglese valuta gli esiti attraverso diversi indicatori tra i quali la mortalità o le infezioni ospedaliere, applicando modelli di tipo Poisson o Negative Binomial nel caso di overdispersion.

criticato⁹. Tale critica suggerisce però, più che un abbandono delle tecniche finora proposte, un affinamento delle modalità secondo cui sono applicate per valutare l'efficacia delle strutture ospedaliere.

a) Ampiezza campionaria: se la scelta di utilizzare piccoli campioni di pazienti affetti da particolari patologie sembra giustificata per studi di nicchia di tipo epidemiologico, per studi inerenti l'efficacia di strutture sanitarie occorre utilizzare informazioni che riguardano un numero molto più elevato di pazienti e molte più patologie [Leyland e Boddy 1998]. I dati amministrativi, essendo generalmente su base informatizzata, hanno il vantaggio di poter utilizzare informazioni senza dover implementare nuovi flussi, evitando quindi incrementi di costo. Inoltre permettono di raccogliere con sistematicità grandi quantità di dati inerenti tutte le patologie e tutti i pazienti o su campioni di grande ampiezza e sono più facili da maneggiare e più sistematici nel tempo e nello spazio [Damberg *et al.* 1998]. I risultati di una valutazione di efficacia sono così molto più robusti.

b) L'outcome più utilizzato dalle agenzie di benchmark è il tasso di mortalità intraospedaliero entro 30 giorni dalla dimissione [Thomas *et al.* 1994; Leyland e Boddy 1998; Normand *et al.* 1997; Goldstein e Spiegelhalter 1996; Marshall e Spiegelhalter 2001]. Tuttavia, per diverse ragioni, la mortalità non è sufficiente a descrivere l'efficacia di un ospedale. Prima di tutto, non tutte le patologie comportano necessariamente il rischio di mortalità in generale (es. malattie croniche, ortopediche) o nel breve periodo (malattie oncologiche nella prima fase della malattia)¹⁰. Secondariamente, anche per malattie gravi il tasso di mortalità può non essere da solo un in-

⁹ «Outcome is neither a sensitive nor a specific marker for quality of care» [Lilford *et al.* 2004, 1148-1149].

¹⁰ «The relationship of death rates to quality of care remains controversial and unproven. Unless findings are adjusted for patient characteristics, conclusions about quality based on an evaluation of [patient] outcomes may be erroneous.» [Iezzoni *et al.* 1996, 1379].

dicatore di efficacia (ad esempio, un ospedale può sistematicamente ricorrere a trasferimenti dei malati gravi)¹¹. In terzo luogo può essere molto difficile ricavare opportune variabili di risk adjustment ad essa legate [Iezzoni 1997; Marshall e Spiegelhalter 2001; CIHI 2009]¹². Come costruire quindi insiemi di outcome capaci di descrivere meglio la qualità delle cure? La letteratura suggerisce di utilizzare: outcome clinici con opportuni indicatori di *case-mix* desumibili da cartelle cliniche [Iezzoni *et al.* 1996]; *proxy* degli outcome clinici che descrivono la frequenza di trattamenti o risultati classificabili come *best practices*, come fanno ad esempio molte agenzie internazionali di *benchmark* [CIHI 2009; AHRQ 2012]; questionari sulla qualità della vita inerenti le condizioni di salute del paziente nel suo complesso, come per esempio le scale SF-12 e SF-36 [Ware e Sherbourne 1992; Apolone *et al.* 1997], o aspetti particolari, ad esempio il suo stato fisico-mentale dopo cure riabilitative, come la scala FIM [Tesio 1993]; «eventi sentinella» imprevisti che comportino per il paziente la morte o danni seri di tipo fisico o psicologico, o elevati rischi di morte e gravi danni.

Non è facile adottare tali outcome in quanto per alcuni le informazioni sono usualmente conservate in cartelle cliniche o registri appositi compilati a mano mentre altri richiedono costose rilevazioni *ad hoc* per essere rilevati [Damberg *et al.* 1998].

Tuttavia, se si opta per le ragioni dette per dati desumibili da archivi amministrativi, si possono utilizzare outcome di contesto che, pur non descrivendo direttamente i risultati dell'assistenza sul paziente, registrano fattori e condizioni che sono determinanti, in senso positivo e/o

¹¹ «Mortality may depend more on the mix of patients reaching hospital in the first place, rather than the quality of care given once admitted. In hospital mortality in particular may be prone to bias and manipulation» [Goldstein e Spiegelhalter 1996, 399].

¹² «Risk adjustment methods do not show whether the unexplained difference in mortality rates results from differences in effectiveness of care or unmeasured differences in patient risk at the time of admission» [Jencks *et al.* 1988, 3611].

negativo, per la riuscita della cura [Pagano e Vittadini 2004]. Tali outcome di contesto sono perciò *proxy* dell'efficacia dell'assistenza ospedaliera erogata. Esempi di tali outcome sono le dimissioni volontarie, i trasferimenti tra strutture, i ricoveri ripetuti e i ritorni in sala operatoria. In connessione con essi si possono utilizzare indicatori di case mix di tipo generale (età, sesso, tipo di ricovero e dimissioni) o specifici per patologie.

c) Valutazione a livello di ospedale, reparto o DRG: una delle critiche esplicite di Lilford *et al.* [2004] concerne le variabili di risk adjustment¹³. Tale critica appare particolarmente valida se si pensa di valutare la qualità a livello di ospedale come spesso avviene. Costruire graduatorie specifiche per reparto o addirittura per DRG [Berta *et al.* 2013] permette di scegliere indicatori di risk adjustment meno generici e più legati a specifiche patologie.

d) Sistema di presentazione dei risultati: date le critiche sopra esposte ai ranking delle strutture sanitarie, in modo più prudente e conservativo ci si può limitare a suddividerle in differenti aree. Le *best practices*, le *worse practices* e gli ospedali posti intorno alla media. Come si vede nella figura 1 ciò può essere fatto definendo le tre aree in base ai quartili (o altri tipi di quantili) della distribuzione degli intervalli di confidenza dei parametri casuali u_i [Gori e Vittadini 1999; Pagano e Vittadini 2004]. È ciò che viene fatto, ad esempio, dall'agenzia canadese CIHI¹⁴.

5. La valutazione dell'efficacia in sanità: l'esperienza in Regione Lombardia

Il Sistema di Valutazione sopra descritto è stato sperimentato e utilizzato dal 2004 ad oggi per valutare gli

¹³ «The sensitivity of an institution's position in league tables to the method of risk adjustment used suggests that comparisons of outcomes are unlikely to tell us about the quality of care» [Lilford *et al.* 2004, 1148-1149].

¹⁴ <https://www.cihi.ca/en>.

ospedali lombardi [Vittadini 2010; Berta *et al.* 2012; Berta *et al.* 2013]. Dovendo, per quanto prima esposto, scegliere una base dati di tipo amministrativo si è optato per la scheda di dimissione ospedaliera. Le peculiarità che rendono la scheda di dimissione ospedaliera uno strumento essenziale per la valutazione della qualità ospedaliera riguardano l'immediata reperibilità, la facilità interpretativa delle informazioni contenute, il contenimento dei costi di rilevazione, la presenza delle informazioni cliniche dei pazienti e la disponibilità su base informatica. Afferma, infatti, il d.m. 28 dicembre 1991 a riguardo delle SDO, che possono essere di «supporto ai processi di valutazione, programmazione, gestione e controllo dell'attività ospedaliera, nonché rilevazione sistematica di carattere epidemiologico».

Tutto ciò rimane condizionato alla verifica della rigidità con cui i dati sono raccolti e certamente una variabilità rispetto alla qualità dei dati potrebbe essere una fonte di distorsione in termini di valutazione della variabilità tra Regioni dell'efficacia.

La scheda di dimissione ospedaliera non fornisce tutte le informazioni necessarie per implementare un adeguato e completo Sistema di Valutazione. Ad esempio, a riguardo della mortalità a 30 giorni, outcome utilizzato a livello internazionale, è necessario provvedere ad un *matching* con altre fonti di dati regionali che ne consentano il calcolo.

Dalle SDO si sono desunti gli outcome di contesto e relative variabili di risk adjustment riportati nelle tabelle 1 e 2.

La scelta di non includere covariate ospedaliere nel modello multilevel è rafforzata dalla recente letteratura in materia scientifica [vedi in particolare Ash *et al.* 2012], poiché implicherebbe che i risultati di efficacia siano letti non solo al netto della gravità degli assistiti ma anche delle diverse risorse finanziarie e strutturali dei singoli ospedali.

Come si è detto, è preferibile operare la valutazione, piuttosto che al livello degli ospedali nel loro complesso, a livello di singolo reparto ospedaliero.

TAB. 1. *Outcome*

Descrizione
Mortalità totale (mortalità intraospedaliera + mortalità entro 30 gg dalla dimissione)
Dimissioni volontarie
Trasferimenti tra strutture
Ricoveri ripetuti per lo stesso MDC entro l'anno
Ritorni in sala operatoria avvenuti durante lo stesso ricovero

TAB. 2. *Variabili case-mix paziente*

Descrizione
Sesso
Età paziente
Diagnosi sentinella (condizioni caratterizzate da un livello elevato di urgenza ai fini del trattamento)
Peso del DRG (rappresenta la «costosità» relativa del caso trattato rispetto al costo medio del sistema)
Diagnosi di tipo cardiovascolare
Diagnosi oncologica
Comorbidità (misurata attraverso l'indice di Elixhauser [Elixhauser <i>et al.</i> 1998])

Questo consente di indirizzare la valutazione ad una divisione ospedaliera con una maggiore omogeneità di pazienti trattati e «coadiuvare» in questo modo la funzione di risk adjustment del modello. Infatti, con questa scelta si riduce l'eterogeneità dei pazienti trattati, che risultano così più simili e i modelli risentono meno di eventuali eterogeneità non controllabili a causa di covariate non osservate.

Per ciò che concerne l'analisi per reparto si sono scelti i seguenti i nove reparti tra i più significativi in termini di dimensione, volume e di tipologia di attività svolta: Chirurgia generale, Medicina, Cardiologia, Cardiochirurgia, Neurologia, Neurochirurgia, Urologia, Oncologia, Ortopedia.

L'applicazione dei modelli multilevel, considerando ognuno dei cinque outcome per ogni reparto, permette di ottenere 45 diversi modelli con relativi risultati espressi in termini di intervalli di confidenza degli outcome corretti

dalle variabili di risk adjustment, rappresentabili tutti nei termini della figura 1.

Procedendo in modo analogo a quanto descritto in precedenza, si giunge a suddividere gli ospedali per ogni outcome in tre aree: best practices (sopra la media regionale); worse practices (sotto la media regionale); reparti intorno alla media regionale.

Questo passaggio risulta fondamentale per un altro obiettivo: sintetizzare la valutazione per un utilizzo gestionale. Questo obiettivo può essere raggiunto attraverso una rappresentazione grafica sintetica.

La rappresentazione sintetica è finalizzata ad una lettura gestionale della performance, che in questo modo non si riduce ad un mero confronto tra ranking, ma diventa la descrizione complessiva della qualità di un ospedale. In questo modo il manager ospedaliero ha una sintesi dell'attività del proprio ospedale e può riconoscere le aree di best practices da quelle critiche.

Il manager regionale può invece confrontare diversi grafici sintetici e avere un quadro complessivo degli ospedali con buona performance e di quelli che necessitano un tempestivo miglioramento.

Un sistema sulla valutazione *ex post* dell'efficacia relativa, articolato su reparti e svariati ranking, posizionando gli ospedali in aree di best, medium, worse practice ha innanzitutto lo scopo di favorire un auto-miglioramento delle strutture sanitarie. In questo modo si ricavano indicazioni operative, più immediatamente comprensibili ed utilizzabili dal mondo professionale.

6. Conclusioni

Nel 2015, il Laboratorio MES della Scuola Superiore Sant'Anna di Pisa e il Centro di Ricerca CRISP dell'Università Milano-Bicocca hanno condiviso l'opportunità di sperimentare questa metodologia per la valutazione degli outcome di efficacia per gli ospedali appartenenti ai Sistemi Sanitari delle Regioni del Network. La sperimen-

tazione portata avanti mirava ad estendere il confronto alle numerose strutture ospedaliere presenti nei contesti regionali, con un particolare focus iniziale sulle Aziende Ospedaliero-Universitarie. Il capitolo successivo descrive dunque come la collaborazione tra Laboratorio MES e CRISP-Bicocca abbia portato il Sistema di Valutazione del Network delle Regioni a sperimentare la valutazione sistematica dell'efficacia adottata in Lombardia, adattandone in parte alcune specifiche metodologiche a seguito del confronto tra le Regioni facenti parte del Network.

Copyright © 2016 by Società editrice il Mulino

La valutazione della performance delle Aziende Ospedaliere-Universitarie.
Finalità, metodi, e risultati a confronto, a cura di Sabina Nuti e Tommaso Grillo Ruggieri,
Ed. Il Mulino, 2016, ISBN 9788815265944

Copyright © 2016 by Società editrice il Mulino