

METODI STATISTICI

Nella maggior parte delle applicazioni, gli indicatori di valutazione degli esiti degli interventi sanitari sono espressi come rapporti in cui il numeratore è rappresentato dal numero dei trattamenti erogati o dal numero di pazienti che hanno sperimentato l'esito in studio, il denominatore dal gruppo di pazienti che riceve quel trattamento o dalla popolazione a rischio. In altri casi gli indicatori di esito possono essere espressi come misure di durata, sopravvivenza o attesa (ad esempio, tempo di attesa tra frattura e intervento del femore).

Se si tratta di confrontare lo stato di salute della popolazione residente in aree geografiche diverse, e guardare l'efficacia dei trattamenti dal punto di vista della funzione di *tutela della salute*, gli indicatori possono essere costruiti utilizzando al denominatore la popolazione residente o specifici sottogruppi della stessa (*population-based denominators*). Nel caso in cui si vogliono descrivere le differenze per soggetto erogatore, il denominatore può essere definito come carico lavorativo, in termini di numero di pazienti, ricoveri, giornate di degenza, numero di visite o altre unità di attività lavorativa (*workload-based denominators*).

Sono riportati di seguito i principali elementi necessari ai fini della definizione analitica, operativa e di misura degli indicatori di esito descritti nel presente rapporto.

Popolazione in studio

Indicatori population-based

La popolazione in studio è costituita o dai residenti nel Lazio al 01/01/2013 (Fonte ISTAT, elaborazioni Agenzia di Sanità Pubblica - Lazio Sanità) o dalla popolazione degli assistiti vivente al 01/01/2013 e affetta da Diabete Mellito o da Broncopneumopatia Cronico-Ostruttiva (BPCO). Gli algoritmi per la definizione delle due popolazioni sono riportati in Appendice.

Indicatori workload-based

La popolazione in studio è costituita dalle dimissioni ospedaliere per la patologia in esame dei residenti nel Lazio. Questa selezione è stata effettuata allo scopo di attribuire a tutti i pazienti la stessa probabilità di trovare ricoveri precedenti l'episodio di cura in esame. La popolazione è costituita dalle dimissioni relative agli anni 2008-2013.

Esiti

Sono state considerate le seguenti misure di esito: mortalità a breve, medio e lungo termine, riammissioni a breve termine, ospedalizzazioni per specifiche condizioni, procedure chirurgiche, complicanze a breve termine a seguito di specifici interventi, tempi di attesa. Le informazioni necessarie per la loro definizione sono state desunte da singoli sistemi informativi o dall'integrazione tra diversi sistemi informativi.

Esposizione

Indicatori population-based

L'esposizione è rappresentata dalle Aziende Sanitarie Locali (ASL) di residenza.

Indicatori workload-based

L'esposizione è rappresentata dalle strutture di trattamento. Queste corrispondono alle strutture di primo accesso al Pronto Soccorso o alle strutture di ricovero.

Misura dell'effetto

La misura di associazione utilizzata è il **Rischio Relativo (RR)**. Il RR è calcolato come rapporto tra due rischi ed esprime l'eccesso (o il difetto) di rischio del gruppo posto al numeratore rispetto al gruppo posto al denominatore. Per esempio, se per la ASL di residenza *A* il RR di mortalità rispetto alla ASL *B* (ASL di riferimento) risulta pari a 2, si può concludere che i residenti nella ASL *A* hanno un rischio di morte due volte superiore rispetto a quello dei residenti nella ASL *B*. Viceversa, se per la ASL di residenza *C* il RR di mortalità è pari a 0.25, i residenti di questa ASL hanno un rischio di morte quattro volte più piccolo di quello osservato nella ASL di riferimento. Per alcuni indicatori la misura di associazione utilizzata è l'**Hazard Ratio (HR)**, interpretabile come un RR istantaneo.

P-value

Il **valore di *p*** rappresenta la probabilità che il rischio relativo osservato differisca da uno (ipotesi nulla) solo per effetto del caso. Un valore di *p* prossimo all'unità indica, ad esempio, che un RR di mortalità non è diverso da 1; in altri termini le mortalità nei due gruppi posti a confronto non sono, dal punto di vista statistico, significativamente diversi. Un valore di *p* basso, al contrario, indica che la differenza osservata tra i due gruppi non è imputabile esclusivamente all'effetto del caso. In particolare, valori di *p* minori o uguali a 0.05 indicano che le mortalità nei due gruppi sono, dal punto di vista statistico, significativamente diversi.

Misura dell'effetto "aggiustato" in presenza di confondimento

Negli studi osservazionali la valutazione comparativa degli esiti deve tenere conto delle possibili disomogeneità esistenti nelle popolazioni studiate, dovute a caratteristiche quali età, genere, gravità della patologia in studio, presenza di comorbidità croniche, etc. Questi fattori possono agire come confondenti dell'associazione tra esito ed esposizione, pertanto nel presente rapporto è stato effettuato un aggiustamento degli indicatori attraverso l'utilizzo di metodi di *risk adjustment*. Questi permettono di studiare le differenze tra strutture o ASL di residenza (espresse in termini di RR), "al netto" del possibile effetto confondente della disomogenea distribuzione delle caratteristiche dei pazienti.

Risk adjustment

La procedura di *risk adjustment* prevede innanzitutto la costruzione di una misura di gravità specifica della popolazione in studio, realizzata attraverso l'applicazione di modelli di regressione multivariata (modelli predittivi) che permettono l'analisi della relazione tra i possibili predittori dell'esito (fattori di rischio) e l'esito considerato.

La misura di associazione utilizzata per l'analisi della relazione tra i predittori e l'esito è l'**Odds Ratio (OR)**. L'OR è calcolato come rapporto tra due odds ed esprime l'eccesso (o il difetto) di rischio del gruppo posto al numeratore rispetto al gruppo posto al denominatore. L'interpretazione è la stessa di quella descritta per il Rischio Relativo, di cui l'OR è una buona approssimazione nel caso di esiti rari. Nel presente rapporto, una volta costruita la misura di gravità, il confronto degli esiti nei gruppi presi in esame, siano essi strutture o popolazioni, è stato effettuato utilizzando il metodo della *standardizzazione diretta*.

Questo metodo permette il confronto dell'esito di ciascuna struttura o ASL di residenza sia con una popolazione di riferimento (per esempio un *pool* di strutture o ASL con i valori più bassi di mortalità aggiustata) che con tutte le altre strutture o ASL in esame.

Per maggiori dettagli consultare il rapporto "[Valutazione comparativa di esito degli interventi sanitari](#)".

Modelli statistici

L'analisi statistica condotta può essere distinta in:

- costruzione del modello predittivo;
- confronto tra strutture e ASL di residenza;
- confronto temporale per struttura e ASL di residenza.

Costruzione del modello predittivo

La costruzione del modello predittivo è stata realizzata attraverso l'analisi della relazione multivariata tra i fattori potenzialmente associati all'esito e l'esito considerato. Lo strumento utilizzato è la regressione multivariata.

Poiché la scelta del modello statistico più adatto all'analisi dei dati è funzione dell'esito in studio, sono stati utilizzati:

1. modelli di regressione logistica, per variabili di esito di tipo dicotomico (presenza *versus* assenza). In questi modelli, la trasformata logaritmica dell'*odds* di esito è espressa come combinazione lineare delle variabili impiegate nel modello predittivo:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + X_{ki}\beta_k$$

2. modelli di sopravvivenza, per esiti espressi come tempo di sopravvivenza. In questi modelli una trasformata dell'esito viene calcolata come funzione lineare delle variabili impiegate nel modello predittivo:

$$\log[\lambda(t_i)] = \log[\lambda_0(t)] + X_{1i}\beta_1 + X_{2i}\beta_2 + X_{ki}\beta_k$$

Nel caso in cui siano verificate le condizioni di proporzionalità dei rischi di sopravvivenza nei diversi livelli di esposizione, è stato utilizzato il modello di Cox. Questo permette la stima degli Hazard Ratio. Nel caso in cui le condizioni di proporzionalità non siano soddisfatte, vengono rappresentate, per i diversi livelli di esposizione, le curve di sopravvivenza aggiustate e i tempi mediani aggiustati.

3. modelli di regressione di Poisson, per indicatori *populaton-based*. In questi modelli, la trasformata logaritmica del tasso dell'esito (τ) è espressa come combinazione lineare delle variabili impiegate nel modello predittivo:

$$\log(\tau_i) = \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + X_{ki}\beta_k$$

Di tutti i fattori di rischio identificati sulla base delle conoscenze disponibili in letteratura, sono stati inclusi nel modello predittivo i seguenti:

- fattori di rischio "a priori", ossia fattori da includere nel modello indipendentemente dalla loro associazione con l'esito: essi includono l'età, il genere ed eventuali fattori di gravità della patologia in studio;
- fattori selezionati attraverso procedure automatiche di tipo *stepwise*: essi includono la presenza/assenza di comorbidità (specifiche per esito considerato) desunte dal ricovero indice e dai ricoveri ospedalieri dei due anni precedenti.

Numerosità minima per struttura

Per ciascun indicatore è stata calcolata la numerosità minima per struttura che consenta di segnalare come statisticamente significativo un predeterminato effetto atteso, espresso come rapporto tra il rischio p_i osservato nella i -esima struttura e il rischio π osservato nella popolazione complessiva. La potenza $(1-\beta)$ del test è stata fissata al 70%, l'errore α di I Tipo al 5%.

$$\text{Numerosità minima per struttura} = (Z_\alpha + Z_\beta)^2 / [2 \times \arcsin(\sqrt{\pi}) - 2 \times \arcsin(\sqrt{p_i})]^2$$

Le ipotesi sull'effetto atteso, calcolato in termini di rischio relativo (RR), sono state differenziate in relazione alla frequenza percentuale dell'esito in studio nella popolazione complessiva:

- per esiti con frequenza inferiore all'1%, è stato considerato un RR pari a 4;
- per esiti con frequenza compresa tra l'1% e il 5%, è stato considerato un RR pari a 3;
- per esiti con frequenza maggiore del 5%, è stato considerato un RR pari a 2.

Confronto tra strutture e ASL di residenza

I rischi di ciascuna struttura e di ciascuna ASL sono stati confrontati col rischio "overall" della popolazione regionale. Le strutture con un volume di attività inferiore ad una soglia definita a priori vengono incluse in un'unica categoria, analizzata complessivamente. I rischi sono stati aggiustati (standardizzazione diretta) mediante l'utilizzo dei modelli predittivi precedentemente definiti, stimati senza intercetta, in cui oltre alle variabili selezionate come componenti della misura di gravità, sono state incluse n variabili *dummy* (variabili che assumono il valore 1 per i soggetti appartenenti al gruppo considerato e il valore 0 per tutti gli altri soggetti) che rappresentano gli n gruppi messi a confronto. I parametri stimati dal modello sono stati utilizzati per calcolare i *rischi attesi* nel caso in cui tutte le strutture e tutte le ASL presentassero la stessa distribuzione della popolazione generale per età, genere, gravità della patologia e comorbidità. I rischi attesi così ottenuti sono stati successivamente corretti attraverso un fattore moltiplicativo che tiene conto della natura non lineare dei modelli utilizzati.

Presentazione dei rischi e dei tassi aggiustati

I rischi, espressi in percentuale, vengono presentati in forma tabellare e rappresentati graficamente.

Distribuzione dei confondenti con effetto di aggiustamento

Le variabili confondenti che concorrono ad una variazione relativa di almeno il 10% tra le stime del rischio grezzo e quello aggiustato sono state identificate mediante l'applicazione della seguente metodologia:

1. per ciascun indicatore, a partire dalle variabili incluse nei modelli predittivi, viene applicata una tecnica tipo *change-in-estimate* al fine di identificare i fattori confondenti che concorrono alla variazione delle stime. Il metodo definisce tali confondenti attraverso una procedura iterativa. Al primo *step* viene selezionato un modello "poco parsimonioso" che include, oltre all'esposizione e all'esito, tutti i fattori inclusi nel modello di *risk adjustment*.

Successivamente, vengono esclusi dal modello tutti quei fattori la cui presenza non determina una variazione della stima del coefficiente di interesse (associazione tra esposizione ed esito). La variazione della stima normalmente considerata apprezzabile, e quindi tale da consigliare la non esclusione del confondente, è compresa, in relazione al fenomeno studiato, tra il 10% e il 20%, ma la scelta rimane comunque arbitraria. In questo studio è applicata una procedura che consente una selezione automatica dei confondenti nel caso di esposizioni su più livelli; vengono selezionate tutte quelle variabili che determinano, in almeno il 10% delle strutture esaminate, una variazione delle stime pari ad almeno il 10%;

2. per ciascuna delle strutture analizzate, è presentata in forma tabellare la distribuzione delle variabili a maggiore frequenza individuate al punto precedente.

3. Questa metodologia è applicata solo agli indicatori calcolati nel treemap.

Selezione del benchmark regionale

Il *benchmark* è stato definito sulla base dei seguenti passi:

1. inclusione nel modello predittivo di un numero di *dummy* pari alle strutture o ASL di residenza, individuate per la patologia in esame. Le strutture con un volume di attività inferiore ad una soglia definita a priori vengono incluse in un'unica categoria, analizzata complessivamente e che non entra nella procedura successiva di selezione del benchmark;
2. nella fase preliminare, inclusione nel gruppo iniziale di riferimento delle strutture o ASL di residenza con risultato più favorevole rispetto agli esiti considerati;
3. ordinamento delle misure di associazione (RR o HR) aggiustate, stimate per ciascuna struttura o ASL rispetto al gruppo iniziale di riferimento, ed identificazione della struttura o ASL con risultato migliore;
4. inclusione nel gruppo iniziale di riferimento della struttura o ASL identificata nel passo precedente, ed inclusione nel modello predittivo di un numero di *dummy* pari a tutte le altre strutture o ASL di residenza; ordinamento delle misure di associazione aggiustate, stimate per ciascuna struttura o ASL rispetto a quella di riferimento, ed identificazione di una ulteriore struttura o ASL con risultato migliore.
5. la procedura di inclusione nel benchmark si arresta quando la misura di associazione della struttura o ASL identificata nel passo precedente evidenzia una variazione di rischio aggiustato superiore al 10% e un *p-value* inferiore a 0.20.

Confronto temporale per struttura ASL di residenza

L'analisi dell'andamento temporale 2008-2013 è stata realizzata includendo, in un unico modello di aggiustamento, il termine di interazione tra esposizione (struttura ospedaliera o ASL di residenza) e anno di analisi. Questo assicura una completa confrontabilità tra le stime ottenute per anni differenti. Pertanto, il *trend* temporale è aggiustato per tutti i potenziali fattori confondenti individuati dai modelli predittivi. L'analisi dell'andamento temporale viene presentata per tutte le ASL di residenza e per tutte le strutture con un volume di attività superiore alla soglia prefissata nel 2013.

Modelli a effetti fissi versus modelli a effetti casuali

La scelta tra modelli “a effetti fissi” e modelli “a effetti casuali” è oggetto di grande dibattito. Nel *Programma Regionale per la Valutazione degli Esiti degli Interventi Sanitari* (PREValE) tutti i modelli utilizzati per il confronto tra strutture ospedaliere o ASL di residenza sono “a effetti fissi”. I modelli “a effetti casuali”, anche noti come modelli multilivello, sono insostituibili per stimare le componenti della varianza (variabilità entro e variabilità tra le strutture) ma producono stime distorte se vengono utilizzati per confrontare tra loro ospedali o ASL di residenza. Infatti, applicando le tecniche multilivello, i rischi di ciascun ospedale vengono contratti verso la media generale (effetto *shrinkage*), pertanto le strutture dalle performance “migliori” sarebbero presentate troppo negativamente mentre quelle dalle performance “peggiori” sarebbero presentate troppo positivamente. Inoltre, l'effetto *shrinkage* dipende dalla numerosità campionaria, di conseguenza la distorsione non è uguale per tutte le strutture: tale “differenzialità” introdurrebbe notevoli alterazioni nelle posizioni reciproche tra gli ospedali.

Appendice

1. Selezione della coorte di pazienti diabetici

Sono considerati come casi di diabete tutti i soggetti con età maggiore o uguale a 35¹ anni, vivi, residenti e assistiti nella Regione al 31 dicembre dell'anno di stima (2012), che soddisfano i criteri di inclusione di seguito riportati:

- dal Sistema informativo ospedaliero, inclusivo della mobilità passiva, si selezionano i dimessi, in regime ordinario, con diagnosi principale o secondaria di diabete (ICD-IX-CM: 250.XX) nell'anno di stima o nei cinque anni precedenti. Per la mobilità la ricerca è ristretta ai cinque anni precedenti la stima;
- dall'archivio delle Prescrizioni Farmaceutiche (Farmaceutica territoriale e farmaci ad erogazione diretta) si selezionano i soggetti con un consumo cronico (definito come consumo di almeno 2 prescrizioni in un anno) di farmaci antidiabetici (ATC=A10) nell'anno di stima o in almeno uno dei tre anni precedenti l'anno di stima;
- dall'archivio delle esenzioni si selezionano i soggetti con un codice di esenzione per diabete (013) attivato precedentemente al periodo in cui si effettua la stima.

2. Selezione della coorte di pazienti con BPCO

L'identificazione della popolazione affetta da BPCO² viene effettuata utilizzando un algoritmo basato sui dati relativi ai ricoveri ospedalieri e alle prestazioni farmaceutiche.

Sono considerati come casi di BPCO tutti i soggetti con età maggiore o uguale a 45 anni, vivi, residenti e assistiti nella Regione al 31 dicembre dell'anno di stima (2012), che soddisfano i criteri di inclusione di seguito riportati:

- dal Sistema informativo ospedaliero, inclusivo della mobilità passiva, si selezionano i dimessi, in regime ordinario, con diagnosi principale o secondaria di BPCO (ICD-9-CM: 490-492, 494, 496) nell'anno in stima o nei cinque anni precedenti;
- dall'archivio delle Prescrizioni Farmaceutiche (Farmaceutica territoriale ed erogazione diretta), si selezionano i soggetti con consumo di farmaci respiratori (R03) nell'anno di stima caratterizzati da almeno una delle seguenti combinazioni di quantità e durata:

più di 120 giorni tra la prima prescrizione e l'ultima, almeno 5 confezioni;

○

intervallo tra prima e ultima prescrizione variabile tra 30 e 120 giorni e numero di prescrizioni variabile tra 3 e 10, una sola classe terapeutica (ATC a quattro cifre);

○

intervallo tra prima e ultima prescrizione variabile tra 120 e 210 giorni e numero di prescrizioni variabile tra 3 e 4, una sola classe terapeutica (ATC a quattro cifre).

¹ Vengono esclusi i pazienti con meno di 35 anni in quanto si tratta, nella quasi totalità, di Diabete di Tipo1 (insulino-dipendente), caratterizzato da bassa numerosità e gravità clinica tale da portare questi malati ad essere seguiti da centri specialistici che garantiscono uniformità di cure, due condizioni che producono stime instabili e scarsa variabilità.

² Prodotto del gruppo di lavoro del CINECA del Consorzio Mario Negri.

Bibliografia

- Laura Amato, Paola Colais, Marina Davoli, Eliana Ferroni, Danilo Fusco, Silvia Minozzi, Fulvio Moirano, Paolo Sciattella, Simona Vecchi, Martina Ventura, Carlo Alberto Perucci. Volumi di attività ed esiti delle cure: prove scientifiche dalla letteratura e dalle valutazioni empiriche in Italia. *Epidemiologia e Prevenzione* 2013; 37(2-3) suppl (1):1-100
- Agabiti N, Davoli M, Fusco D, Stafoggia M e Perucci CA. Valutazione degli esiti degli interventi sanitari. *Epidemiologia e Prevenzione* 2011; *Epidemiologia e Prevenzione* 2011; 35(3-4) suppl (2):1-80
- Arcà M, Fusco D, Barone AP e Perucci CA. Introduzione ai metodi di riskadjustment nella valutazione comparativa dell'outcome. *Epidemiologia e Prevenzione* 2006; 30(4-5) suppl (1):1-48
- Arcà M, Fusco D, Barone AP, Perucci CA. Risk adjustment and outcome research. Part I. *Journal of Cardiovascular Medicine (Hagerstown)*. 2006; 7(9):682-90
- Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Quality and Safety in Health Care* 2003; 12:458-464
- Coory M, Duckett S, Sketcher-Baker K. Using control charts to monitor quality of hospital care with administrative data. *International Journal for Quality in Health Care* 2008; 20(1): 31-39
- Fantini MP, Stivanello E, Frammartino B, Barone AP, Fusco D, Dallolio L, Cacciari, Perucci CA. Risk adjustment for inter-hospital comparison of primary cesarean section rates: need, validity and parsimony. *BMC Health Services Research* 2006; 6(1):100
- Ferraris VA, Ferraris SP. Risk Stratification and Comorbidity. In: Cohn LH, Edmunds LH Jr, eds. *Cardiac Surgery in the Adult*. New York: McGraw-Hill 2003:187-224
- Greenland S, Brumback B. An overview of relations among causal modelling methods. *International Journal of Epidemiology* 2002; 31:1030-1037
- Iezzoni LI. *Risk Adjustment for measuring healthcare outcomes*. Health Administration Press 2nd ed., 1997
- Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004; 363(9415):1147-54
- Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet* 1997; 350 (9085): 1128-30
- Stafoggia M, Lallo A, Fusco D, Barone AP, D'Ovidio M, Sorge C, Perucci CA. Spie charts, target plots, and radar plots for displaying comparative outcomes of health care. *J Clin Epidemiol*. 2011; 64(7):770-8
- McNamee R. Confounding and confounders. *Occupational and Environmental Medicine* 2003; 60: 227-23
- Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Lippincott-Williams & Wil 3rd ed., 2008
- Sherlaw-Johnson C, Gallivan S, Treasure T, Nashef SA. Computer tools to assist the monitoring of outcomes in surgery. *European Journal of Cardiothoracic Surgery* 2004; 26(5):1032-6
- Spiegelhalter D, Grigg O, Kinsman R, Treasure T. Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *International Journal for Quality in Health Care* 2003; 15:7-13
- Winkel P, Zhang NF. Risk-Adjusted Control Charts. In: *Statistical Development of Quality in Medicine*. John Wiley & Sons, Ltd, 2007: 149-163
- Wilcosky TC, Chambless LE. A comparison of direct adjustment and regression adjustment of epidemiologic measures. *Journal of chronic diseases* 1985; 38 (10): 849-856
- Woodall WH. The Use of Control Charts in Health-Care and Public-Health Surveillance. *Journal of Quality Technology* 2006; 38 (21): 89-104
- Zou G. A modified poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* 2004; 159: 702-6