



# Programma Nazionale Esiti – PNE

Edizione 2023

## 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/ interventi erogati o dal numero di pazienti che hanno sperimentato l'esito in studio,
- il denominatore è costituito dal gruppo di pazienti che riceve quel trattamento/intervento o dalla popolazione a rischio.

In altri casi gli indicatori di esito possono essere espressi come misure di durata / sopravvivenza / attesa (i.e. tempo di attesa occorso tra frattura del femore e relativo intervento).

Se si tratta di confrontare lo stato di salute della popolazione residente in aree geografiche diverse, e guardare l'efficacia dei trattamenti/interventi 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. Per maggiori dettagli consultare il rapporto di Agabiti et al. "Valutazione comparativa di esito degli interventi sanitari", Epidemiologia e Prevenzione, 2011.

## Popolazione in studio

### *Indicatori population-based*

La popolazione alla base dello studio è costituita dai residenti in Italia al 1 gennaio 2022 (Fonte ISTAT).

### *Indicatori workload-based*

La popolazione alla base dello studio è costituita dalle dimissioni ospedaliere per la patologia/intervento in esame, dei residenti in Italia, per i quali è stata verificata la corrispondenza anagrafica da Anagrafe Tributaria.



Questa selezione è stata effettuata allo scopo di attribuire a tutti i pazienti la stessa probabilità di trovare ricoveri precedenti e successivi l'episodio di cura in esame. La popolazione è costituita dalle dimissioni relative agli anni 2015-2022.

## Esito

Sono state considerate le seguenti misure di esito: mortalità a breve 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 dal flusso SDO e dalla Banca dati dell'Anagrafe Tributaria.

## Esposizione

### *Indicatori population-based*

L'esposizione è rappresentata dall'area di residenza: ASL o regione di residenza

### *Indicatori workload-based*

L'esposizione è rappresentata dalle strutture di ricovero.

## Misura dell'effetto

La misura di associazione utilizzata è il Rischio Relativo (RR). Il RR è calcolato come rapporto tra due tassi/proporzioni ed esprime l'eccesso (o il difetto) di rischio del gruppo posto al numeratore rispetto al gruppo posto al denominatore. Per esempio, se per l'ASL di residenza A il RR di mortalità rispetto all'ASL B (area di riferimento) risulta pari a 2, si può concludere che i residenti nella ASL A hanno un rischio di morte due volte superiore 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 RR istantanee



## 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 i tassi di 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  $p$  molto bassi (ad esempio minori o uguali a 0,05) indicano che i tassi di 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 e/o aree territoriali (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/area di residenza sia con una popolazione di riferimento (per esempio un pool di strutture/aree con i valori più bassi di mortalità aggiustata) che con tutte le altre strutture/aree in esame.



## Modelli statistici

L'analisi statistica condotta può essere distinta in:

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

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

$$\log [\pi_i/(1-\pi_i)] = \beta_0 + X_{1i} \beta_1 + X_{2i} \beta_2 + X_{ki} \beta_k$$

2. modelli di sopravvivenza, per esiti espressi come tempo di sopravvivenza. In questi modelli una trasformatadell'esitoviene 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 grezzi ed aggiustati.

3. modelli di regressione di Poisson, per indicatori population-based. In questi modelli, la trasformata logaritmica del tasso (t) dell'esito è espressa come combinazione lineare delle variabili impiegate nel modello predittivo:

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



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

Infine, per alcuni indicatori, i tassi sono calcolati solo in forma grezza: per tali indicatori si calcola un intervallo di confidenza utilizzando il metodo proposto da Wilson (1927). In tal caso i tassi sono visti come una realizzazione di un processo aleatorio che ha come possibile concretizzazione il verificarsi o meno dell’evento.

### **Numerosità minima per struttura**

La determinazione della numerosità minima degli eventi per struttura è stata nella misura tale da consentire 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  $\pi$  osservato nella popolazione complessiva. La potenza  $(1 - \beta)$  del test è stata fissata al 70%, l’errore  $\alpha$  di I Tipo al 5%. Inoltre, la numerosità minima garantisce che quelle diagnosi e/o procedure siano attività proprie della struttura e non ricoveri occasionali.

### **Confronto tra strutture e aree di residenza**

I rischi di ciascuna struttura e di ciascuna area sono stati confrontati col rischio “overall” della popolazione generale, e con i rischi nell’anno precedente. Le strutture con un volume di attività inferiore alla numerosità minima precedentemente definita 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 aree presentassero la stessa distribuzione della “popolazione generale” per età, sesso, gravità della patologia/intervento 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.

Va specificato che i modelli utilizzati non forniscono i  $p$ -value dei rischi relativi nel caso in cui non si verifichi alcun evento (ad esempio un rischio di morte è pari a zero) o quando si verifichino solo eventi favorevoli (ad esempio la quota di interventi di femore entro le 48 ore sia pari al 100%).

### **Indicatori su base pluriennale**

Per ragioni legate alla potenza statistica, alcuni indicatori sono stati calcolati su base biennale, triennale o quadriennale.



**Presentazione rischi aggiustati** I rischi, espressi in percentuale, vengono presentati in forma tabellare e rappresentati graficamente tramite un diagramma a barre.

**Effetto aggiustamento** L'effetto dell'aggiustamento viene misurato in termini di differenza assoluta tra rischi grezzi ed aggiustati. Maggiore è l'effetto dell'aggiustamento, tanto più il rischio aggiustato si discosta da quello grezzo e quindi tanto più ampia sarà la differenza. Una differenza negativa esprime, nel caso di esiti sfavorevoli (es. mortalità), un miglioramento delle stime dovuto all'aggiustamento; viceversa, nel caso di esiti favorevoli.

### **Selezione delle variabili potenzialmente confondenti**

Le variabili confondenti che concorrono ad una variazione significativa tra le stime del rischio grezzo e quello aggiustato sono state identificate mediante l'applicazione della seguente metodologia: per ciascun indicatore, a partire da tutte le variabili incluse nei modelli predittivi, viene applicata una procedura iterativa *stepwise*, che seleziona il numero ottimale di covariate, attraverso un metodo indiretto basato sulla funzione di verosimiglianza.

### **Confronto temporale per struttura e aree di residenza**

L'analisi dell'andamento temporale 2015-2022 è stata realizzata includendo, in un unico modello di aggiustamento, il termine di interazione tra esposizione (struttura ospedaliera o area 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 aree di residenza e per tutte le strutture con un volume di attività superiore alla soglia prefissata nel 2022. Nella lettura dei dati va tenuto presente che, quando vengono calcolati i trend, la popolazione di riferimento è dell'intero periodo di osservazione dei trend e dunque non è la stessa rispetto a quella che si considera negli indicatori annuali; ne consegue che i valori degli indicatori aggiustati relativi all'ultimo anno di un trend, differiscono da quelli calcolati considerando esclusivamente l'ultimo anno.

### **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 Nazionale Esiti tutti i modelli utilizzati per il confronto tra strutture ospedaliere o aree di residenza sono "a effetti fissi". I modelli "a effetti casuali", anche noti come modelli multilivello tendono a produrre stime distorte se vengono utilizzati per confrontare tra loro ospedali o aree 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. I modelli multilivello sono tuttavia insostituibili per stimare le componenti della varianza (variabilità entro e variabilità tra le strutture) e sono utilizzati all’interno del PNE per misurare la variabilità, intra-regionale, tra ospedali o aree per alcuni indicatori. I *modelli multilevel* utilizzati includono le stesse comorbidità utilizzate per il calcolo dell’indicatore aggiustato e considerano, come secondo livello di analisi, o l’ospedale di erogazione della prestazione, o l’area di residenza dei pazienti. Il parametro preso in considerazione per la misurazione della variabilità è il *median odds ratio* (MOR).

## La stima della variabilità: il Median Odds Ratio

Il *Median Odds Ratio* (MOR) misura la variabilità tra gruppi, come le *strutture ospedaliere* o le *aree di residenza*. Questa misura è sempre maggiore o uguale ad 1. Se il MOR è uguale ad 1, non c’è variabilità tra i gruppi. Se c’è variabilità tra i gruppi, il MOR sarà elevato. Il MOR può anche essere interpretato come l’incremento «atteso» di rischio di un individuo che si sposta casualmente da un’area all’altra, nell’ipotesi in cui lo spostamento avvenga sempre verso aree con un rischio più elevato. Questa misura ha numerose proprietà statistiche che la rendono preferibile rispetto ad altri indici di variabilità. Infatti, il MOR è un indicatore di variabilità «corretto» che tiene in considerazione diversi parametri, quali: la numerosità all’interno dei gruppi, il numero di esiti all’interno dei gruppi, le differenze «iniziali» tra i singoli gruppi e la media generale (ad esempio, la media nazionale), la distribuzione dei fattori socio-demografici e clinici nei diversi gruppi (effetto composizione).



## Bibliografia

- 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 risk adjustment 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
- Hox, J. *Multilevel Analysis: techniques and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates, 2002
- 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
- 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





*Ministero della Salute*



*Agenzia Nazionale per i Servizi Sanitari Regionali*

- 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
- Wilson, E. B. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association. 1927; 22 (158): 209–212.
- 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

