

## Analisi dei dati/*Data analysis*

(GIC - G Ital Cardiol 2006; 7 (Suppl 1-2): 13S-15S)

### Metodologie generali

Le analisi presentate in questo lavoro sono state effettuate stratificando per macroaree geografiche: Nord-Ovest; Nord-Est; Centro; Sud e Isole. Le macroaree sono costituite come riportato di seguito:

- Nord-Ovest: Piemonte, Valle d'Aosta, Lombardia, Liguria;
- Nord-Est: Trentino Alto Adige, Veneto, Friuli Venezia Giulia, Emilia Romagna;
- Centro: Toscana, Lazio, Marche, Umbria;
- Sud e Isole: Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna.

I principali fattori di rischio, riconosciuti in letteratura quali fattori associati indipendentemente alla mortalità a 30 giorni dall'intervento di bypass, sono stati quindi combinati per creare un unico profilo di rischio sul quale valutare l'outcome (esito a 30 giorni) dei singoli Centri di Cardiocirurgia. L'analisi statistica è stata condotta in due fasi: 1) determinare quali tra i fattori di rischio rilevati sono significativamente legati all'esito in studio, 2) determinare come pesare tali fattori di rischio per predire la probabilità di morte dopo l'intervento, per ogni paziente, dato il suo profilo di rischio.

Nella maggior parte dei casi i fattori di rischio sono stati utilizzati come variabili *dummy* (presenza o assenza del fattore); la frazione di eiezione è stata suddivisa in tre categorie: < 30%, 30-49% e ≥ 50%. Per questa variabile è stata, inoltre, introdotta una categoria per i dati mancanti.

Per individuare associazioni significative con l'esito in studio sono stati utilizzati modelli logistici univariati per ognuno dei fattori considerati.

Per valutare l'associazione indipendente tra ciascun fattore di rischio e l'evento in studio è stato utilizzato un modello logistico multivariato. Nel costruire tale modello si è fatto ricorso ad una procedura di tipo *stepwise backward* con probabilità di esclusione pari a 0.20 e probabilità di inclusione pari a 0.10. Per valutare la bontà di adattamento del modello selezionato è stata cal-

colata la statistica  $r^2$ ; la calibrazione tra i decili delle probabilità stimate è stata valutata utilizzando il  $\chi^2$  di Hosmer-Lemeshow, mentre per valutare la capacità predittiva del modello è stata calcolata l'area sotto la curva ROC. La scelta del modello è stata valutata anche attraverso il criterio AIC (Akaike Information Criteria =  $-2 * \log\text{-likelihood} + 2 * \text{numero dei parametri nel modello}$ ) per confrontare l'adattamento dell'algoritmo di correzione scelto rispetto ad altri testati in fase intermedia. L'algoritmo per la correzione del rischio è stato costruito utilizzando solo i dati relativi alle strutture che sono risultate conformi ai criteri di inclusione nel database analitico.

L'analisi dei dati ha previsto la fase di *cross-validation* che consiste nel costruire il modello di predizione su una metà del campione in studio e nel riapplicare tale modello alla restante parte. La popolazione in studio è stata, quindi, suddivisa in modo casuale in due sottocampioni di dimensioni simili: il campione I (n = 17 231) è stato utilizzato per costruire il modello predittivo; il campione II (n = 17 079) è stato utilizzato come database indipendente per validare il modello ottenuto. Infine, la stima dei coefficienti e dei relativi valori di significatività associati ad ogni fattore è stata effettuata utilizzando l'intero dataset a disposizione, in modo da massimizzare la precisione delle stime stesse.

Una volta individuato il miglior algoritmo di aggiustamento del rischio sui dati a disposizione, sono state applicate due diverse procedure di aggiustamento della mortalità rilevata nei singoli Centri: una procedura di tipo indiretto ed una di tipo diretto. Con la prima è possibile stimare un rischio di morte, aggiustato per le covariate introdotte nel modello, che può essere confrontato con il rischio medio rilevato nell'intero campione per stabilire se i singoli Centri o le singole macroaree geografiche presentano una performance significativamente migliore o peggiore rispetto alla media. Con la procedura di aggiustamento di-

retto, una volta scelto un riferimento (*benchmark*), è possibile stimare l'eccesso di rischio dei singoli Centri rispetto a tale riferimento e stabilire quali Centri presentano performance migliore o peggiore.

Per definizione, la performance dei Centri che non appartenevano al *benchmark*, ma che si confrontavano con esso, non può essere significativamente migliore.

### **Confronto con la media della popolazione osservata**

Con la metodologia di aggiustamento indiretto, il modello statistico individuato come "migliore" viene utilizzato per stimare le probabilità di morte dei singoli pazienti in studio. Sommando tali probabilità per struttura si ottiene il numero dei decessi attesi. Dividendo il numero di decessi osservati per il numero dei decessi attesi e moltiplicando tale rapporto per il tasso di mortalità medio dell'intero campione si ottiene la mortalità aggiustata, tecnicamente definita come *risk-adjusted mortality rate*. Questo indicatore rappresenta, per ogni Centro, la mortalità che avrebbe avuto l'intera popolazione se avesse avuto lo stesso rapporto decessi osservati/decessi attesi di quel Centro. Un *risk-adjusted mortality rate* significativamente inferiore rispetto al tasso di mortalità medio indica che la performance di quella determinata struttura è migliore rispetto a quella media rilevata nell'intero campione (*low outliers*); al contrario, un *risk-adjusted mortality rate* significativamente più elevato individua una performance peggiore (*high outliers*).

Pertanto, con questa procedura, è possibile stimare un rischio di morte aggiustato per le covariate introdotte nel modello. Questo tipo di aggiustamento permette di stabilire se i singoli Centri presentano una situazione significativamente migliore o peggiore rispetto alla media dell'intero campione ma non consente di confrontare tra di loro i singoli Centri in studio (ogni Centro è confrontabile solo con la media, ma non con gli altri).

Sulla base della significatività statistica dei confronti, per la presentazione dei risultati, è stato costruito uno score strutturato in tre categorie: *low outliers*, non diversi dalla media, *high outliers*.

### **Confronto con il pool di riferimento**

Con la procedura di aggiustamento diretto il modello statistico selezionato in precedenza è stato utilizzato per valutare l'eccesso di rischio delle strutture partecipanti rispetto ad un riferimento interno. In questo caso il *benchmark* scelto è costituito dai Centri che presentavano contemporaneamente le seguenti caratteristiche:

- la più bassa mortalità aggiustata sulla base del modello di *risk adjustment* utilizzato;
- qualità dei dati buona (sulla base del monitoraggio clinico effettuato);

- oltre 500 interventi di bypass aortocoronarico isolato registrati nel periodo di osservazione;
- oltre 18 mesi di osservazione.

Questo tipo di aggiustamento consente il confronto diretto delle performance dei singoli Centri, oltre che con il riferimento utilizzato, anche tra loro, si può quindi stabilire un ordinamento dei gruppi messi a confronto rispetto all'esito in studio. È stato quindi possibile costruire un grafico che consente il confronto diretto tra le performance dei singoli Centri di Cardiocirurgia.

### **Analisi per macroarea geografica di residenza**

È stata effettuata un'analisi della mortalità a 30 giorni dopo intervento di bypass aortocoronarico isolato per macroarea geografica di residenza. Tale analisi consente di valutare l'esistenza di una differenza, in termini di esito, tra i pazienti che si sottopongono all'intervento nella propria area geografica di residenza e coloro che lo effettuano in un'area diversa. Sono stati calcolati gli odds ratio grezzi e aggiustati della variabile di esposizione (intervento nella macroarea di residenza) utilizzando come riferimento tutti i pazienti residenti in una determinata area geografica e sottoposti ad un intervento di bypass isolato nella stessa area di residenza. Anche per questo tipo di analisi è stato utilizzato il modello di aggiustamento del rischio descritto in precedenza. Per permettere una migliore comparabilità dei risultati, per tutte le macroaree, è stato utilizzato lo stesso modello.

### **Ulteriori analisi**

Per i Centri 40, 64 e 67 è stata rilevata un'eccessiva presenza di dati mancanti su molte variabili risultate importanti predittori del rischio di morte a 30 giorni dall'intervento. È stato calcolato il range di variazione delle stime nei due casi estremi in cui i *missing* fossero tutti no o tutti sì. Nelle tabelle è stata riportata la stima che si sarebbe ottenuta forzando la non presenza dei fattori di rischio nel caso di dati mancanti.

Inoltre, è stata effettuata un'analisi descrittiva dell'andamento della mortalità per giorni dall'intervento confrontando la situazione dei *low outliers* con quella degli *high outliers*.

È stato, infine, valutato l'andamento temporale della mortalità nei singoli Centri confrontando la situazione al 2003-2004 con quella al 2002. Per ogni Centro è stato calcolato l'eccesso di rischio del periodo 2003-2004 rispetto al 2002 (odds ratio = 1), utilizzando per la correzione delle stime il modello di aggiustamento validato sull'intero campione. La differenza tra i due periodi analizzati è stata considerata significativa qualora il valore di *p* fosse < 0.05. Questo tipo di analisi è stata effettuata solo per i Centri che avevano raccolto dati sufficienti (almeno 100 interventi di bypass aortocoronarico isolato) in entrambi i periodi di osservazione.

## Livello di significatività

Il livello scelto per definire una differenza statisticamente significativa è stato  $p < 0.05$ . Pertanto nella raffigurazione grafica dei Centri rispetto alla media della popolazione e rispetto al pool di riferimento, le barre che descrivono gli intervalli di confidenza sono state costruite tenendo conto di questo criterio.

Sebbene tale scelta sia ampiamente condivisa nella letteratura internazionale, esiste almeno un aspetto che merita di essere sottolineato. I Centri con un maggiore volume di interventi presenteranno molto più frequentemente differenze statisticamente significative semplicemente perché le loro stime saranno più stabili ed i loro intervalli di confidenza più stretti. La situazione inversa si verificherà, invece, per i Centri a basso volume di interventi per i quali sarà più difficile supportare statisticamente eventuali differenze riscontrate.

Va sottolineato, inoltre, che la scelta di considerare significativa o meno una particolare differenza osservata dovrebbe tenere conto dall'impatto e delle conseguenze che tale scelta potrebbe determinare nel contesto sociale in cui si opera.

## Analisi di sensibilità

Per valutare l'effetto della perdita al follow-up di alcuni pazienti, sulla stima del rischio in ciascun Centro, è stata eseguita un'analisi di sensibilità. Tale analisi è stata effettuata considerando alternativamente le due ipotesi più estreme del range di variabilità delle stime:

- tutti i persi al follow-up vivi a 30 giorni dall'intervento (approccio conservativo);
- tutti i persi al follow-up deceduti a 30 giorni dall'intervento.

Tutte le analisi sono state effettuate utilizzando il pacchetto statistico Stata versione 8.1.

\* \* \*

The analyses presented in this paper were conducted for the whole sample and by region: North-East, North-West, Center, South and Islands.

Main recognized risk factors associated with mortality at 30 days after coronary artery bypass graft intervention were combined to obtain a common case-mix of patients, valid for estimating the outcome (30-day mortality) in each Cardiac Surgery Center. Most of the risk factors were used as dummy variables; ejection fraction was considered as a categorical variable (< 30%, 30-49%, ≥ 50%, missing).

Multivariate logistic regression was performed to evaluate the independent association between each risk factor and 30-day mortality risk. A backward stepwise procedure was used, with an exclusion probability of 0.20 and an inclusion probability of 0.10.

The  $r^2$  statistic was calculated to evaluate the model's goodness of fit; the Hosmer-Lemeshow  $\chi^2$  was applied to evaluate the calibration between deciles of estimated probabilities; the area

under the receiver-operating characteristic curve was calculated to describe the predictive capability of the model. To select the final adjusting model among those tested in the intermediate phase, the Akaike Information Criteria was used.

Only data from Centers that met all the inclusion criteria were included and were used to build the risk-adjustment algorithm.

A cross validation procedure was applied, randomly assigning patients to two equal-size samples: sample I was used to build the predictive model ( $n = 17\,231$ ); sample II was used as an independent database for model validation ( $n = 17\,079$ ). In order to provide more precise parameter estimates, the entire dataset was used to compute definitive coefficients and their  $p$  values.

After having determined the best algorithm based on available data, we applied both direct and indirect risk-adjustment procedures.

The latter estimates mortality risk after adjusting for potential confounders introduced in the model. This risk can be compared with the average mortality of the whole sample in order to establish if individual centers or regions present significantly different conditions when compared to the average. This adjustment procedure does not permit a direct comparison between Centers or regions (each center/area can only be compared with the average).

By applying the best predictive algorithm back to each Center (area) data set the expected number of deaths of that Center (area) can be estimated. The risk-adjusted mortality rate is then calculated by dividing the observed by the expected number of deaths in each Center or area, and by multiplying this ratio by the average mortality rate of the whole sample. This indicator is the best estimate of what the whole population mortality rate would be if the population had the same observed/expected death ratio of that Center (area). A risk-adjusted mortality rate significantly ( $p < 0.05$ ) lower than the average mortality rate indicates that the health care provider performs better than the average of the whole sample (low outliers); on the contrary, a risk-adjusted mortality rate significantly higher shows a worse performance (high outliers).

The direct risk-adjustment procedure can estimate the risk excess of each Center compared with the chosen benchmark, and Centers that perform better/worse can be specified. In this analysis, the reference – benchmark – consisted of a pool of Centers, among those included in the analytical database, which presented the following characteristics: a) the lowest risk-adjusted mortality rates; b) good quality of data; c) more than 500 procedures performed; d) more than 18 months of observation.

We compared the performance of each Center with all the others using the direct standardization procedure.

A 30-day mortality analysis by region of residence was performed. This approach allowed us to compare outcomes of patients who underwent coronary artery bypass graft interventions in their region of residence and outcomes of those who underwent the same intervention outside. Crude and adjusted odds ratios for the exposure variable (surgery performed in the area of residence) were calculated. All patients who had a coronary artery bypass graft in their area of residence were used as the reference category.

The same risk-adjustment model described above was used to perform this analysis.

Distribution of deaths according to number of days after the date of intervention and emergency condition was analyzed.

The temporal trend of mortality between the periods 2003-2004 and 2002 was estimated. The odds ratio for the 2003-2004 (2002 was the reference period) was computed and considered statistically significant if  $p < 0.05$ .

Finally, a sensitivity analysis was performed to analyze the "lost to follow-up" effect on the risk estimate in each Center.

All statistical procedures were performed using Stata 8.1.