

## PROTOCOLLO D'INTESA

**FINALIZZATO ALL'ADOZIONE DI STRATEGIE CONDIVISE PER LA RILEVAZIONE DI POSSIBILI RISCHI SANITARI RIFERIBILI A CRITICITÀ AMBIENTALI INDIVIDUATE NEI COMUNI COMPRESI NELLA GIURISDIZIONE DELLA PROCURA DELLA REPUBBLICA DI SANTA MARIA CAPUA VETERE**

### Metodi statistici

#### 1. I rapporti standardizzati di incidenza (SIR)

Per l'analisi dell'incidenza di ciascuna sede topografica indagata, sono stati considerati il numero di casi osservati in ciascuna area comunale nel periodo 2010-2013 e confrontati con i rispettivi casi attesi. Per il calcolo dei SIR, il numero dei casi attesi è stato calcolato, separatamente per i maschi e per le femmine, mediante il metodo della standardizzazione indiretta, utilizzando come riferimento i tassi età specifici (per classi di età quinquennali) osservati nella popolazione regionale residente, ad esclusione della popolazione residente nella città di Napoli.

In questo modo, sono stati calcolati, per ciascuna delle sedi topografiche selezionate, separatamente rispetto al genere, negli 80 comuni afferenti alla giurisdizione della Procura della Repubblica di SMCV i rapporti standardizzati di incidenza (SIR) con i rispettivi intervalli di confidenza al 95%.

La stima dei SIR nelle aree scarsamente popolate o in malattie relativamente rare tende ad essere fortemente instabile e difficilmente interpretabile dal punto di vista epidemiologico. Nel modello di Poisson la media e la varianza dei casi osservati nelle aree comunali è posta uguale; tuttavia, tale condizione è spesso violata per la presenza di sovra dispersione che si realizza quando la varianza è maggiore rispetto alla media. Inoltre, un'altra importante assunzione nella analisi spaziale effettuata con i SIR è che le aree comunali siano indipendenti, condizione che non è realizzata soprattutto per lo studio di malattie potenzialmente correlate all'esposizione ambientale.

Per tali motivi l'utilizzo delle stime dei SIR è insufficiente per modellare la distribuzione dei rischi di incidenza nelle analisi spaziali ed è appropriato utilizzare approcci statistici in grado di modellare la sovra dispersione dei dati.

#### 2. Modelli bayesiani gerarchici

In particolare sono stati utilizzati due modelli bayesiani gerarchici:

- Modello Poisson Gamma (PG)

Questo modello assume che il numero dei casi osservati in ciascuna area segua una distribuzione di Poisson. Per ottenere la distribuzione a posteriori dei rischi relativi, la verosimiglianza di Poisson è coniugata con una distribuzione Gamma a priori. Pertanto, i rischi relativi per ciascuna area risultano essere la media pesata dei SIR e la media regionale. In questo modo il modello preserva le stime stabili, basate su molte osservazioni e attrae verso la media generale le stime instabili basate su pochi casi. Tuttavia il modello introduce solo un

fattore di sovradisersione spazialmente non strutturato e la componente spaziale non è inclusa nel modello<sup>1,2</sup>.

- Modello Besag York Mollié (BYM)

In questo modello i rischi relativi comunali vengono modellati tenendo in conto due effetti casuali: una componente di eterogeneità, data dalla sovradisersione non spazialmente strutturata dei dati, e una seconda componente definita di clustering, modellata attraverso una distribuzione di tipo CAR (Conditional Auto Regression) con media e varianza pesata con la media e la varianza delle aree adiacenti<sup>3,4</sup>.

In entrambi i modelli utilizzati sono state scelte distribuzioni a priori non informative per assenza di ipotesi a priori sulla distribuzione dei rischi.

Sebbene i risultati vengano presentati limitatamente all'area oggetto dello studio, è importante specificare che al fine di stimare in modo appropriato gli indicatori di rischio, in modelli all'interno dei quali vengono inclusi anche effetti spazialmente strutturati, sono stati inclusi i dati riferiti a tutti i comuni della Campania, ad esclusione della città di Napoli, sia quelli relativi all'incidenza sia alla deprivazione socio-economica e all'indicatore di pressione ambientale comunale IPC.

Sia per il modello PG sia per il modello BYM sono stati calcolati i rischi relativi (RR) e le probabilità a posteriori (PP) mediante due diversi modelli: un modello grezzo; un modello multivariato con l'inserimento delle covariate IPC e deprivazione socio-economica. Per quest'ultimo modello, al fine di indagare le diverse componenti del rischio, sono stati stimati i RR e le PP: 1) ponendo a zero (media regionale) sia l'IPC sia la deprivazione socio-economica; 2) ponendo a zero (media regionale) solo la deprivazione socio-economica.

Nelle analisi statistiche è stato utilizzato il valore mediano dell'IPC comunale centrato rispetto alla media regionale e il range dei valori degli 80 comuni va da -19,1 a 53,9. Per quanto concerne la variabile deprivazione socio-economica, i dati comunali ISTAT sono stati ricalcolati sulla media regionale, con un range, nell'area oggetto dello studio, che va da -3,5 a 6,2.

Nelle mappe e tabelle allegate vengono presentati i soli risultati ottenuti attraverso il modello PG. Tale scelta è dovuta alla maggiore difficoltà nell'interpretazione epidemiologica delle mappe ottenute con il modello BYM; questi modelli sono fortemente influenzati dalle assunzioni a priori e potrebbero introdurre indebiti trend spaziali in presenza di dati fortemente sparsi, come per alcune sedi oncologiche analizzate.

I risultati sono presentati attraverso mappe e tabelle dei RR e delle PP.

Per i RR sono state scelte 9 classi: 4 per i RR inferiori alla media regionale (<0,60; 0,60-0,79; 0,80-0,89; 0,90-0,94); 4 per i RR superiori alla media regionale (1,05-1,09; 1,10-1,19; 1,20-1,29; >1,29); una che include il rischio relativo pari alla media regionale (0,95-1,04).

Per le PP è stato scelto di rappresentare il livello di credibilità in 4 classi (<0,10; 0,10-0,49; 0,50-0,90; >0,90).

L'associazione tra l'incidenza oncologica e l'IPC e la deprivazione socio-economica, variabili incluse nel modello multivariato, viene espressa nei risultati come variazione percentuale del rischio (mediante l'intervallo di credibilità al 95%) all'incremento unitario delle covariate. Inoltre, viene presentata anche la differenza interquartile (espressa con l'intervallo di credibilità al 95%) del rischio di incidenza sia per l'IPC sia per la deprivazione socio-economica.

## **Bibliografia**

1. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. 1987;43:671–91.
2. Wolpert R, Ickstadt K. Poisson-gamma random field models for spatial statistics. *Biometrika*. 1998;85:251–67.
3. Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika*. 1995; 82:733-46.
4. Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Stat Math*. 1991;43:1–21.