

**IL DISEGNO E LA METODOLOGIA DEGLI STUDI EPIDEMIOLOGICI**

## INTRODUZIONE

Lo studio di coorte è comunemente utilizzato in epidemiologia per l'analisi degli effetti di salute di un potenziale fattore di rischio. Questo approccio, a differenza di altri, permette di stimare il tempo esatto in cui ogni soggetto è stato a rischio di sviluppare l'evento in studio (tempo-persona). La sequenza temporale dello studio viene stabilita rispetto alla collocazione nel tempo dell'osservatore. Lo studio di coorte è definito prospettico, se l'informazione sull'esposizione è registrata all'inizio dello studio ed il *follow-up* è condotto nel futuro o retrospettivo, se la coorte viene definita nel passato ed il tempo a rischio è intercorso prima dell'inizio dello studio.

Un aspetto cruciale nella pianificazione di qualsiasi studio epidemiologico riguarda l'individuazione di fonti dei dati adatte all'identificazione della popolazione che ha dato origine ai casi (popolazione sorgente) e dell'esito in studio. Uno studio di coorte retrospettivo può essere svolto anche utilizzando i dati raccolti dai Sistemi Informativi correnti, grazie all'integrazione di archivi di popolazione e archivi di dati sanitari. L'indubbio vantaggio in termini di contenimento dei costi può comportare un grande dispendio in termini di gestione dei dati perché questi archivi nascono con finalità diverse da quelle di uno studio epidemiologico. E', perciò, importante che questi archivi siano completi (per limitare la perdita selettiva di informazioni), validi (per limitare la presenza di errori sistematici), di qualità omogenea per tutto il periodo in studio ed accurati (per evitare errori di informazione). Il disegno dello studio di coorte tende, per costruzione, a ridurre i problemi relativi al *bias* di informazione e di selezione, ma la scarsa qualità della fonte informativa potrebbe tuttavia inficiare la validità dell'analisi. Fonti informative largamente disponibili sono le anagrafi comunali che, nel Lazio dal 1996, sono disponibili in formato digitale e forniscono informazioni relative alla popolazione residente ed i suoi movimenti nel tempo.

La misura dell'esposizione è particolarmente delicata in epidemiologia ambientale poiché la variabile di esposizione individuale è spesso solo un'approssimazione della esposizione vera<sup>1</sup>. Quanto più è imprecisa sarà questa misura tanto più vi sarà il rischio di una misclassificazione dell'esposizione che può portare ad una attenuazione dei rischi relativi stimati. La definizione dell'esposizione, in molti studi, si basa su modelli di dispersione degli inquinanti emessi dalla fonte di inquinamento presa in esame nell'analisi<sup>2-3</sup>. In quest'ottica risulta indispensabile l'integrazione dei dati di

popolazione con sistemi d'informazione geografica (GIS – Geographic Information System). Nell'ambito dell'epidemiologia ambientale è sempre più diffuso l'utilizzo di dati di popolazione georiferiti per la selezione degli esposti sulla base della distanza della loro residenza da fonti di inquinamento presenti nel territorio quali industrie<sup>4</sup>, antenne radio e tv<sup>5</sup>, inceneritori<sup>6</sup>, discariche<sup>7</sup> e strade<sup>8-10</sup>.

Uno degli obiettivi di ERAS Lazio è la valutazione, attraverso studi di coorte di popolazione, dello stato di salute delle persone residenti in prossimità degli impianti per il trattamento dei rifiuti e dei lavoratori addetti al ciclo dei rifiuti. In questo ambito sono stati redatti i seguenti rapporti:

- Valutazione epidemiologica dello stato di salute della popolazione residente nei pressi delle discariche per i rifiuti urbani del Lazio;
- Valutazione epidemiologica dello stato di salute della popolazione residente nell'area di Malagrotta a Roma;
- Valutazione epidemiologica dello stato di salute della popolazione residente nei pressi dei termovalorizzatori del Lazio;
- Valutazione epidemiologica dello stato di salute della popolazione residente nei pressi degli impianti di trattamento meccanico-biologico del Lazio;
- La salute riproduttiva delle donne residenti nei pressi dei termovalorizzatori e delle discariche del Lazio;
- Valutazione epidemiologica dello stato di salute dei lavoratori addetti alla raccolta, trasporto e smaltimento dei rifiuti urbani a Roma.

La metodologia utilizzata in queste analisi è riassunta nella Figura 1. Le coorti di popolazione arruolate hanno utilizzato come fonte dei dati gli archivi anagrafici dei comuni interessati e i Sistemi Informativi Sanitari Regionali per la definizione degli esiti sanitari. L'esposizione ad impianti per il trattamento dei rifiuti (discariche e termovalorizzatori) è stata definita sulla base di due misure: la distanza dal perimetro degli impianti e la concentrazione al suolo dell'inquinante considerato come traccante del singolo impianto, definita sulla base di modelli di dispersione. Tutte le coorti sono state georeferenziate (attribuzione delle coordinate geografiche agli indirizzi di residenza dei soggetti inclusi nello studio), non solo per attribuire ad ognuno la misura dell'esposizione, ma anche per descrivere al meglio le caratteristiche individuali (livello socio-economico) e la realtà circostante (eventuali altri fattori di pressione ambientale presenti nel territorio).

In questo rapporto vengono riassunti i metodi utilizzati per l'analisi epidemiologica nell'ambito del Progetto ERAS Lazio:

- lo sviluppo di una metodologia affidabile per la costruzione di coorti di popolazione da fonte anagrafica e il *follow-up* utilizzando i Sistemi Informativi Sanitari;
- la georeferenziazione dei soggetti inclusi nello studio (attribuzione delle coordinate geografiche agli indirizzi di residenza delle coorti);
- la definizione dell'esposizione attraverso l'uso di modelli di dispersione;
- l'analisi statistica e le misure di associazione utilizzate per valutare la relazione tra esposizione e mortalità/morbosità (Modello di Cox, Hazard ratio).

## **METODI**

### *Costruzione di coorti di popolazione da fonte anagrafica e accertamento degli esiti sanitari*

Agli uffici anagrafici dei comuni in studio sono stati richiesti i dati della popolazione residente al 1 Gennaio 1996 e di tutti coloro che sono entrati nel comune (per nascita o immigrazione) o usciti dal comune (per morte o emigrazione) nel periodo 1996-2008. Per ogni persona sono state richieste le informazioni anagrafiche (cognome, nome, data di nascita, comune di nascita, data di decesso e sesso), l'indirizzo di residenza e tutte le date relative ai movimenti da e per il comune (data di emigrazione, immigrazione). Ogni record relativo ad un individuo doveva essere provvisto di un codice identificativo (numero progressivo individuale) e del codice fiscale. Sono stati richiesti anche i cambiamenti di residenza all'interno dello stesso comune, ma per la maggior parte delle anagrafi oggetto dello studio queste informazioni non sono ancora disponibili in formato digitale.

I dati ricevuti sono stati sottoposti a diversi controlli di qualità:

- è stata verificata la presenza di tutti i campi richiesti: nei casi in cui l'anagrafe non ha fornito informazioni fondamentali, non ricavabili da altre variabili (cognome, nome, data di emigrazione, data di decesso, indirizzo di residenza ecc.), è stata inviata una nuova richiesta dati. In altri casi le informazioni non complete sono state ricavate da altre già disponibili (ad esempio l'attribuzione del sesso, sulla base del codice fiscale o del nome).

- è stata valutata la completezza dei dati (quante informazioni mancanti): negli archivi anagrafici spesso risultava non compilata l'informazione relativa alla data di iscrizione. Per attribuire questa informazione è stata utilizzata la data di nascita, se la persona era nata nel comune, la data di immigrazione, se la persona era nata altrove ma immigrata nel comune e la data di inizio *follow-up* per gli altri.

Dopo aver effettuato questi primi controlli, è stato verificato che l'anagrafe avesse effettuato una corretta estrazione dati. Sono state, quindi, esaminate le date di nascita, immigrazione, emigrazione e morte, per controllare che non vi fossero degli anni senza eventi e che la frequenza immigrazione degli eventi fosse approssimativamente costante nel periodo in studio (1996-2008). Per emigrazioni e decessi è stata necessaria un'ulteriore verifica: il confronto del numero di deceduti per anno con i dati del ReNCaM (Registro Nominativo delle Cause di Morte); il confronto del numero di emigrazioni con il dato ISTAT.

Una volta verificata la completezza dei dati anagrafici, è stata selezionata la coorte dei residenti:

1. è stata verificata la presenza/assenza di record doppi per codice individuale fornito dall'anagrafe e sono state eliminate le informazioni ripetute;
2. è stato verificato che il codice fiscale, dove presente, fosse attendibile e il più possibile completo;
3. a seguito di una lunga procedura di sistemazione delle variabili, è stato ricalcolato il codice fiscale mediante un algoritmo basato su cognome, nome, data di nascita, sesso e comune di nascita dell'individuo. Se incompleto, il codice fiscale ricalcolato è stato sistemato e integrato sulla base del codice fiscale fornito dall'anagrafe;
4. è stata verificata la presenza di record doppi per codice fiscale ricalcolato, sono state eliminate le informazioni ripetute e integrate quelle aggiuntive (es. diverse date di emigrazione ed immigrazione su più record relative allo stesso individuo);
5. è stata selezionata la coorte sulla base delle date di iscrizione e cancellazione dal comune. Sono stati, per esempio, eliminati i nati dopo la chiusura del *follow-up* (2009), i deceduti prima dell'inizio del *follow-up* (1996) e le emigrazioni o immigrazioni avvenute al di fuori del periodo di osservazione;

6. per ogni individuo è stato accertato lo stato in vita alla fine del *follow-up* (vivo/morto, emigrato o irreperibile).

Come fonte dati della mortalità (1996-2008) è stato utilizzato il Registro Nominativo delle Cause di Morte (ReNCaM) che contiene tutti i decessi avvenuti nel Lazio a prescindere dalla residenza del deceduto e tutti i decessi dei residenti nella regione a prescindere dal luogo di decesso. Come fonte dati dei ricoveri ospedalieri è stato utilizzato il Sistema Informativo Ospedaliero (SIO) del Lazio, che rileva e gestisce i dati analitici di tutti i ricoveri ospedalieri (in acuzie e post-acuzie) che ogni anno si verificano negli Istituti di Ricovero e Cura della Regione a partire dal 1994, con una copertura e qualità dei dati soddisfacente solo dal 1996.

#### *Georeferenziazione degli indirizzi di residenza delle coorti in studio*

Negli ultimi anni si è diffuso l'utilizzo di software basati su sistemi di informazione geografica (es. ArcGIS), utili non solo per elaborare mappe ma anche per integrare i dati sanitari o di popolazione in un sistema geografico, effettuare analisi spaziali e per la valutazione dell'esposizione. Gli indirizzi di residenza delle coorti arruolate nell'ambito del progetto ERAS Lazio sono stati georeferenziati per assegnare l'esposizione ed attribuire ad ogni individuo della coorte le variabili ambientali e di contesto sociale utili per definire al meglio l'associazione tra l'esposizione a discariche/inceneritori e mortalità/morbosità per causa.

Le informazioni necessarie per la georeferenziazione degli indirizzi di residenza sono:

- nome della via (es. "Viale Manzoni");
- numero civico (es. "103");
- indirizzo (es. "Viale Manzoni 103");
- CAP (es. "00182");
- Comune (es. "Roma");
- Stato (es. "Ita").

I nomi delle vie di residenza sono stati modificati perché vi fosse una corrispondenza con le vie dello stradario di TELEATLAS, contenute nel software ArcGIS (es. da "Viale Manzoni" a "Viale Alessandro Manzoni"). La via modificata e il numero civico sono stati utilizzati per definire l'indirizzo di residenza corretto. Questo indirizzo è stato poi geocodificato attraverso una procedura automatica del software ArcGIS che restituisce

le coordinate geografiche di tutti gli indirizzi individuati sul territorio dal software. È stato necessario procedere con una geocodifica manuale nei seguenti casi:

1. nello stradario era presente la via ma non quel particolare numero civico: sono state quindi attribuite le coordinate geografiche dei numeri civici vicini;
2. la via non era presente nello stradario: è stata effettuata una ricerca sul sito <http://it.bing.com/> che restituisce le coordinate geografiche corrispondenti (in termini di Latitudine e Longitudine).

Chiaramente, in caso di quote elevate di indirizzi non georeferenziati o di informazioni mancanti/irreperibili, il comune è stato escluso dallo studio per evitare un *bias* di selezione.

#### *Definizione dell'esposizione e modelli di dispersione*

La valutazione dell'esposizione agli impianti di trattamento dei rifiuti si basa solitamente sulla distanza tra due punti: l'impianto (un punto, nel caso di un camino, un perimetro nel caso di una discarica) e la residenza del soggetto, nell'ipotesi che il rischio aumenti all'avvicinarsi agli impianti. Tale approccio non considera la complessità dei meccanismi che intervengono nella diffusione degli inquinanti in aria, dovuta principalmente alla meteorologia e all'orografia del territorio. I modelli di dispersione consentono di tenere conto di questi aspetti, restituendo mappe di concentrazione degli inquinanti che rappresentano l'impronta al suolo dell'impianto considerato. Nel progetto ERASLazio, sono state dunque considerati due approcci per la misura dell'esposizione della popolazione residente nei pressi degli impianti per il trattamento dei rifiuti: la distanza e le impronte degli inquinanti scelti come traccianti degli impianti in studio (Idrogeno Solforato - H<sub>2</sub>S - per le discariche, polveri PM<sub>10</sub> per i termovalorizzatori). Per le 9 discariche e i 2 termovalorizzatori presenti nel Lazio è stato definito un dominio di simulazione modellistica (30x30 Km) centrato su ciascun impianto. Le emissioni di questi inquinanti sono state stimate considerando quantità autorizzate, tipologia di impianto (puntiforme o areale) e modalità di emissione (altezza di rilascio, velocità dei fumi e relativa temperatura). Per stimare le tonnellate/anno emesse dalle sorgenti areali (discariche) è stato usato il modello Landfill Gas Emissions, sviluppato dalla US Environmental Protection Agency<sup>11</sup>, il quale considera gli anni di attività della discarica, la sua capacità e il quantitativo di rifiuto/anno conferito. I lotti delle discariche sono stati suddivisi in celle di 125x125m al fine di distribuire sull'area della discarica le emissioni ottenute; le emissioni stimate

in ogni cella hanno costituito l'input per il modello SPRAY. Nel caso di termovalorizzatori e raffinerie, invece, non è necessario stimare il quantitativo di inquinante emesso, perché si utilizzano i valori autorizzati.

Le emissioni degli impianti (discarica, inceneritore), così ottenute, sono state fornite al modello lagrangiano a particelle 3D SPRAY insieme ai campi meteorologici 3D generati da RAMS. Il risultato dell'integrazione di questi modelli è la definizione dell'impronta al suolo di ogni singolo impianto. La scelta di un modello lagrangiano a particelle è dovuta alla complessità dell'orografia del territorio laziale. In zone che presentano un'orografia meno complessa e una meteorologia meno variabile è possibile eseguire le stesse stime utilizzando modelli di dispersione più semplici (modello gaussiano)<sup>12</sup>.

#### *Analisi statistica e misure di associazione*

L'associazione tra gli indicatori di esposizione (distanza o concentrazioni di inquinanti) e la mortalità/morbidità delle coorti è stata valutata mediante il modello di sopravvivenza di Cox. Questo modello semiparametrico è basato sull'assunto che il rapporto tra i rischi sia costante nel tempo, senza specificare il rischio al baseline. Il Rischio Relativo (RR), stimato dal modello di Cox che nello specifico prende il nome di Hazard Ratio (HR), è calcolato come rapporto tra due rischi di morire (o contrarre la malattia) ed esprime l'eccesso (o il difetto) di rischio del gruppo posto al numeratore rispetto al gruppo posto al denominatore. Per esempio, se il RR di mortalità del gruppo A rispetto al gruppo B (riferimento) risulta pari a 1.58, si può concludere che i soggetti del gruppo A hanno un rischio di morte del 58% superiore a quello del gruppo B. Viceversa, se per il gruppo C il RR di mortalità è pari a 0.25, i soggetti di questo gruppo hanno un rischio di morte quattro volte più piccolo di quello osservato tra i soggetti di riferimento.

Negli studi osservazionali la valutazione epidemiologica deve tenere conto delle possibili disomogeneità tra esposti e non esposti per diverse caratteristiche quali età, genere e stato socio-economico, che a loro volta sono associate al rischio di malattia. Questi fattori possono agire come confondenti dell'associazione tra esposizione ed esito, è pertanto necessario effettuare un aggiustamento attraverso l'utilizzo di metodi multivariati. Questi metodi permettono di studiare le differenze tra esposti e non esposti (espresse in termini di RR) "al netto" del possibile effetto confondente della disomogenea distribuzione delle caratteristiche date.



Negli studi del progetto ERAS Lazio sono state considerate come possibili confondenti della relazione tra inquinanti emessi da discariche e inceneritori e mortalità/morbosità alcune variabili ambientali e di condizione sociale: età, periodo di calendario, livello socio-economico, inquinamento atmosferico di fondo (PM<sub>10</sub>), residenza in prossimità di strade principali, autostrade e industrie.

## **RISULTATI**

Nell'ambito del progetto ERAS, sono state ricostruite le coorti di popolazione residente in più di 20 comuni della Regione Lazio. La grande mole di dati ha consentito di mettere in evidenza differenze nella qualità degli archivi anagrafici comunali, pertanto in questo rapporto vengono riportati alcuni esempi sui controlli e le procedure che sono state effettuate prima di accettare i dati nella analisi. I comuni "virtuosi" sono stati confrontati con altri che hanno fornito dati di scarsa qualità e per i quali è stato necessario attuare diverse procedure di ottimizzazione o addirittura l'eliminazione dallo studio.

Si consideri per esempio il comune di Bracciano. Tra il 1996 ed il 2007 risultano 1,432 deceduti da fonte ReNCaM e 1,455 da fonte anagrafica (Tabella 1). Anche il confronto del numero di deceduti per singolo anno conferma che i dati forniti dall'anagrafe risultano completi, corretti e coerenti. La stessa cosa non si può affermare per il comune di Albano Laziale. Secondo i dati originariamente forniti dall'anagrafe, infatti, il numero annuale di deceduti aumentava col passare del tempo, solo dal 2005 diventava confrontabile con il dato ReNCaM. L'incompletezza dell'estrazione dati è dimostrata dal fatto che non risultava alcun decesso nel 1996, uno solo nel 1997 e 3 nel 1998.

Anche le frequenze annuali delle nascite hanno fornito indicazioni in merito alla qualità dei dati inviati dalle anagrafi. È anomalo che ci siano, nell'arco di 12 anni, dei trend troppo marcati o che ci siano delle forti variazioni da un anno all'altro. Come si può notare dalla tabella 2, la distribuzione dell'anno di nascita del comune di Roccasecca risulta decisamente più omogenea, quindi più attendibile, di quella del comune di Albano Laziale, in cui il numero di nati passava da 166 nel 1996 a 23 nel 2008.

Nei casi in cui gli uffici anagrafici non siano stati in grado di soddisfare le richieste è stato necessario ridurre il periodo in studio o eliminare il comune dalle analisi. L'anagrafe del comune di Civitavecchia ha incontrato diverse difficoltà nell'estrazione dei dati riferiti agli anni 1996-2003, gli unici dati sistematizzati ed utilizzabili sono risultati quelli dal 2004 in poi. Diverso è il caso del comune di Pontecorvo per il quale non siamo mai riusciti ad ottenere i dati anagrafici richiesti.

Alcuni comuni presentavano un'alta quota di record senza indirizzo di residenza: in alcuni casi si è cercato di recuperare il più possibile questa informazione trascrivendo gli indirizzi manualmente dai cartellini anagrafici; in altri casi, anche a causa dell'impossibilità di una collaborazione con gli uffici anagrafici, i file si sono dimostrati inutilizzabili. Nei dati del comune di Fonte Nuova, per esempio, su un totale di 43,089 residenti, per 4,615 (10.7%) non veniva riportato l'indirizzo di residenza.

Per alcuni comuni, invece, la procedura di georeferenziazione è stata determinante nella valutazione della qualità dei dati: se la proporzione di persone con indirizzo geocodificato era bassa (<85%) è stato necessario escludere il comune dall'analisi, se invece questa proporzione era più elevata ( $\geq 85\%$ ) sono state confrontate le caratteristiche (in termini di stato in vita) delle persone geocodificate e non, per non incorrere in *bias* di selezione. Per il comune di Latina (Tabella 3) la quota di persone non geocodificate era molto bassa (1.3%) e non si evidenziavano grandi differenze nella distribuzione dello stato in vita nei due gruppi (geocodificati/non geocodificati). Al contrario, il comune di Paliano presentava un'alta quota di persone con indirizzi non geocodificati (38.4%), pertanto Paliano non è stato inserito nello studio.

La modellistica di dispersione degli inquinanti si sta affermando come strumento complementare a disposizione dell'epidemiologo nella valutazione dell'esposizione della popolazione residente nei pressi di impianti industriali. I risultati delle simulazioni condotte nell'ambito del progetto ERAS Lazio, per la stima dei modelli di dispersione degli inquinanti, mostrano aree molto eterogenee dal punto di vista dell'inquinamento ambientale.

Come si può notare dalla tabella 4, sebbene le dimensioni delle discariche di Civitavecchia e Colleferro siano molto simili, la discarica di Colleferro raggiunge valori

di concentrazione di H<sub>2</sub>S quasi sette volte più elevati rispetto a quella di Civitavecchia (0.765 µg/m<sup>3</sup> VS 0.113 µg/m<sup>3</sup>). I livelli di concentrazione massimi di H<sub>2</sub>S stimati per la discarica di Latina raggiungono il valore di 1.196 µg/m<sup>3</sup> mentre per la discarica di Roccasecca, la cui area è circa 24 volte più piccola, il valore più alto è di 0.314 µg/m<sup>3</sup>. Nell'area di Malagrotta (Roma) si raggiungono le più alte concentrazioni di H<sub>2</sub>S (2.999 µg/m<sup>3</sup>).

## DISCUSSIONE E CONCLUSIONI

Uno studio di coorte prevede la definizione della popolazione in studio attraverso il suo arruolamento e la determinazione a priori di un periodo di osservazione durante il quale i componenti della coorte vengono seguiti nel tempo attraverso contatti ripetuti. La tecnica di cui si è trattato in questo rapporto, invece, ha previsto l'arruolamento di coorti di popolazione senza nessun contatto diretto con chi partecipa allo studio. Si utilizzano infatti i dati anagrafici di tutti i residenti in un certo comune nell'arco di diversi anni e, solo successivamente, sulla base dei dati ottenuti, viene selezionata la popolazione di interesse.

La metodologia utilizzata nel progetto ERASLazio presenta diversi punti di forza:

- rappresenta una proposta innovativa rispetto alle metodologie utilizzate fino ad oggi;
- è una tecnica poco dispendiosa in termini economici e di tempo. Non richiede un vasto impiego di forza lavoro: non è necessario aggiornare lo stato in vita della coorte effettuando visite a domicilio o interviste telefoniche perché la definizione dello stato in vita si basa sui dati forniti dalle anagrafi comunali;
- rispetto alla costruzione di una coorte con metodologia classica, si guadagna in termini di potenza statistica, in quanto, non si estrae un campione della popolazione, ma si studiano tutti i residenti nel comune. Non vi è quindi il problema della rappresentatività tipico degli studi campionari;
- non si rischia di incorrere nel *response bias* in quanto la partecipazione delle persone allo studio non dipende dal loro assenso;
- questa tecnica, associata alla georeferenziazione degli indirizzi di residenza, permette di ottenere una distribuzione spaziale dei soggetti in studio a cui si può associare la distribuzione spaziale dell'esposizione;

- l'utilizzo dei modelli di dispersione, associato a dati di popolazione georeferiti, permette di superare un grande limite degli studi di epidemiologia ambientale che per molto tempo hanno considerato esclusivamente la distanza dagli impianti come esposizione a fonti di inquinamento, ipotizzando che l'inquinante si disperdesse nell'aria in fasce concentriche dal centroide dell'impianto.
- le coorti di popolazione così costruite possono essere facilmente riutilizzate per altri studi attraverso l'attribuzione di nuove informazioni quali esposizione e variabili di confondimento.

Oltre ai tanti vantaggi, questa nuova procedura ha anche alcuni punti di debolezza:

- i rapporti con gli uffici anagrafici del Lazio non sono stati sempre semplici; spesso, a causa della mancata risposta al primo tentativo di contatto, è stato necessario ricontattarli (attraverso telefonate, e-mail e fax) o inviare un nostro tecnico per estrarre i dati dal software anagrafico. Nonostante tutti gli sforzi, i dati di alcuni comuni non sono mai arrivati;
- nonostante questa tecnica permetta un risparmio in termini economici e di tempo, è certamente complessa e laboriosa dal punto di vista della programmazione;
- alcune delle scelte prese durante le fasi di lavorazione dei dati anagrafici sono difficili, sia perché i file hanno una struttura diversa da comune a comune (pertanto non è possibile costruire un programma unico per tutti i file) sia perché, a volte, ci si è trovati a decidere sull'esclusione o l'inclusione di alcuni soggetti con informazioni dubbie;
- questa tecnica può introdurre *bias* di selezione. Si ha un *bias* di selezione nei casi in cui si ha un difetto selettivo di partecipazione che è correlato con l'esposizione e con l'esito in studio. Ciò potrebbe avvenire, per esempio, se vi fosse una proporzione molto diversa di deceduti con indirizzo non geocodificato (quindi non inclusi nello studio) rispetto al gruppo dei geocodificati.

In conclusione, la metodologia proposta rappresenta una strategia ottimale per lo sviluppo di studi di coorte utili alla valutazione del danno da esposizioni ambientali. Ovviamente, debbono essere considerate con cura le possibili distorsioni che si

possono generare in uno studio di coorte e sono necessari controlli rigorosi dei dati anagrafici acquisiti.

**BIBLIOGRAFIA**

1. Baker D, Nieuwenhuijsen MJ: Environmental epidemiology: Study methods and application. *Oxford University Press* 2008.
2. Nyberg F, Gustavsson P, Järup L, Bellander T, Berglind N, Jakobsson R, Pershagen G.: Urban air pollution and lung cancer in Stockholm. *Epidemiology* 2000 Sep; 11(5):487-95.
3. Bellander T, Berglind N, Gustavsson P, Jonson T, Nyberg F, Pershagen G, Järup L.: Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. *Environ Health Perspect* 2001 Jun; 109(6):633-9.
4. Dolk H, Thakrar B, Walls P, Landon M, Grundy C, Sáez Lloret I, Wilkinson P, Elliott P.: Mortality among residents near cokeworks in Great Britain. *Occup Environ Med* 1999 Jan; 56(1):34-40.
5. Dolk H, Elliott P, Shaddick G, Walls P, Thakrar B.: Cancer incidence near radio and television transmitters in Great Britain. II. All high power transmitters. *Am J Epidemiol* 1997 Jan 1;145(1):10-7.
6. Elliott P, Shaddick G, Kleinschmidt I, Jolley D, Walls P, Beresford J, Grundy C.: Cancer incidence near municipal solid waste incinerators in Great Britain. *Br J Cancer* 1996 Mar; 73(5):702-10.
7. Elliott P, Briggs D, Morris S, de Hoogh C, Hurt C, Jensen TK, Maitland I, Richardson S, Wakefield J, Jarup L.: Risk of adverse birth outcomes in populations living near landfill sites. *BMJ* 2001 Aug 18; 323(7309):363-8.
8. Livingstone AE, Shaddick G, Grundy C, Elliott P.: Do people living near inner city main roads have more asthma needing treatment? Case control study. *BMJ* 1996 Mar 16; 312(7032):676-7.
9. English P, Neutra R, Scalf R, Sullivan M, Waller L, Zhu L.: Examining associations between childhood asthma and traffic flow using a geographic information system. *Environ Health Perspect* 1999 Sep; 107(9):761-7.
10. Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA.: Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *Lancet* 2002 Oct 19; 360(9341):1203-9.
11. EPA <http://www.epa.gov/ttn/catc1/dir1/landgem-v302-guide.pdf>.

12. Regione Emilia-Romagna. I Risultati del Progetto Moniter. Gli Effetti degli Inceneritori sull'Ambiente e la Salute in Emilia-Romagna. Quaderni di Moniter 04>11. Bologna, Novembre 2011.