

Dipartimento di Ancona – Servizio Epidemiologia Ambientale
Via C. Colombo, 106 - 60127 Ancona
Tel. 07128732760 - Fax 07128732761
e-mail: Epidemiologia.AmbientaleAN@ambiente.marche.it

STUDI SPAZIALI: COMPARABILITÀ DELLE RESIDENZE ANAGRAFICHE ED ABITATIVE REALI

Marco Baldini, Katiuscia Di Biagio, Silvia Bartolacci, Rita Simeoni, Mauro Mariottini

Osservatorio epidemiologico ambientale della regione Marche
Servizio di epidemiologia ambientale – Dipartimento provinciale ARPAM di Ancona



Parole chiave: studio spaziale, esposizione residenziale, residenza anagrafica, residenza abitativa, concordanza

Ancona, gennaio 2012

SOMMARIO

STUDI SPAZIALI: COMPARABILITÀ DELLE RESIDENZE ANAGRAFICHE ED ABITATIVE REALI ...	1
ABSTRACT	3
INTRODUZIONE	4
Campo d'azione dell'epidemiologia ambientale.....	4
Studi di epidemiologia spaziale	4
Obiettivi.....	6
MATERIALI E METODI	7
Dati utilizzati	7
Analisi statistica.....	7
RISULTATI	10
Analisi sull'intero campione degli anni-residenza	10
Definizione restrittiva di concordanza	11
Definizione ampia di concordanza	11
Analisi sull'ultima residenza.....	13
Definizione restrittiva di concordanza	13
Definizione ampia di concordanza	13
Analisi sulla residenza principale	13
Definizione restrittiva di concordanza	13
Definizione ampia di concordanza	13
Correlazione tra le distanze delle residenze storiche dalla sorgente di esposizione.....	14
DISCUSSIONE.....	15
CONCLUSIONI.....	17
PICCOLO GLOSSARIO	18
BIBLIOGRAFIA	19

ABSTRACT

Obiettivi

Il presente contributo intende fornire spunti di riflessione sugli studi epidemiologici che indagano i fenomeni sanitari in relazione alla variazione spaziale del rischio. In particolare si propone di effettuare alcune valutazioni sull'utilizzo dei dati residenziali correntemente disponibili nella conduzione degli studi spaziali e, nello specifico, sulla comparabilità della storia residenziale anagrafica ed abitativa reale di gruppi di popolazione.

Dati e Metodi

E' stata valutata la concordanza delle informazioni residenziali, ottenute dalle anagrafi comunali e da questionario, di residenti arruolati in uno studio caso-controllo realizzato in tre comuni della regione Marche limitrofi ad una raffineria, attraverso un modello marginale di regressione logistica basato su equazioni di stime generalizzate. A supporto della concordanza dei dati è stato effettuato un confronto tra le distanze delle residenze, anagrafiche ed abitative, dalla sorgente di rischio attraverso la determinazione dell'indice di correlazione lineare di Pearson.

Risultati

Sono state analizzate le informazioni sulle residenze anagrafiche ed abitative di 362 soggetti, per complessivi 5710 anni-residenza. La stima della percentuale di concordanza delle informazioni residenziali nell'intero gruppo di popolazione, tenendo conto della correlazione delle informazioni relative allo stesso soggetto, è stata dell'87,6% (IC95%: 84,2%-90,3%). L'indice di correlazione di Pearson tra le distanze delle residenze anagrafiche ed abitative dalla sorgente di rischio è risultato pari a 0,97 (p-value <0,0001).

Conclusioni

La valutazione delle informazioni relative alla storia residenziale dei soggetti in studio ha evidenziato il consistente grado di completezza dei dati anagrafici rilevati dai sistemi informativi comunali e la buona comparabilità di quest'ultimi con i dati inerenti le residenze abitative. I dati derivanti dai sistemi informativi delle anagrafi comunali, rispetto a quelli ottenuti mediante somministrazione di questionari, pur con i limiti della minore accuratezza nella stima delle esposizioni e dell'impossibilità di differenziare le residenze anagrafiche dai veri domicili, presentano vantaggi in termini di disponibilità e completezza d'informazione se rapportati ai relativi costi di reperibilità. Le residenze anagrafiche, in taluni disegni di studio, possono costituire, a nostro avviso, una valida alternativa all'utilizzo delle residenze abitative e fornire una risorsa aggiuntiva nelle indagini epidemiologiche ambientali.

INTRODUZIONE

Campo d'azione dell'epidemiologia ambientale

L'epidemiologia ambientale ha tre principali valenze: descrittiva, eziologica e valutativa.

Una funzione degli studi descrittivi è quella di favorire la generazione di ipotesi di interesse eziologico; obiettivo degli studi analitici è saggiare tali ipotesi verificando la presenza di un'associazione causale.

Nello specifico, gli studi descrittivi misurano la distribuzione delle variabili all'interno di una popolazione offrendo un'immagine della frequenza, della distribuzione spaziale e dell'andamento temporale degli eventi sanitari di interesse, senza considerazione di alcuna ipotesi causale; gli studi analitici rappresentano il metodo d'elezione per valutare in modo rigoroso l'associazione causale del verificarsi di eventi sanitari avversi in relazione all'esposizione a fattori ambientali e per misurare il rischio a livello aggregato (studi ecologici) o individuale¹.

Agli ambiti descrittivi ed eziologici si aggiunge la componente valutativa dell'epidemiologia ambientale che si propone di misurare gli effetti dell'adozione di interventi di prevenzione, quali possono essere rappresentati da campagne di educazione ambientale e sanitaria o da provvedimenti per la riduzione delle emissioni inquinanti.

Studi di epidemiologia spaziale

L'analisi delle associazioni tra i fattori predittivi e gli esiti sanitari può essere eseguita a diversi livelli di studio scelti in relazione all'ipotesi da verificare e, di conseguenza, al problema ambientale e/o sanitario da affrontare.² Nell'ultimo decennio hanno trovato largo impiego, nell'ambito dell'epidemiologia ambientale, gli studi spaziali (o geografici) che effettuano la descrizione e l'analisi delle variazioni geografiche di eventi sanitari in relazione a fattori di rischio demografici, ambientali, comportamentali, genetici e socio-economici.^{3,4} Spesso le analisi statistiche territoriali si prefiggono la finalità di studiare i fenomeni a livello di area geografica sulle popolazioni residenti, evidenziando ed eventualmente spiegando la variabilità spaziale che li caratterizza.^{5,6} Il ruolo degli studi geografici in epidemiologia ambientale, ed in particolare l'uso di nuovi approcci multilivello ed analisi temporali o/e spaziali che superano o attenuano alcune limitazioni derivanti dalla fallacia ecologica⁷, è stato più volte discusso in letteratura e gli aspetti riguardanti la metodologia di analisi, l'interpretazione dei risultati e la valutazione delle implicazioni sono stati affrontati da numerosi autori.^{8,9,10,11}

E' più recente l'implementazione di studi in cui la dimensione spaziale ha un ruolo fondamentale, ma l'obiettivo dell'analisi è di tipo individuale, ovvero l'intento è evidenziare determinate realtà a livello di singole unità statistiche quali individui o famiglie. Ciò è legato al fatto che la posizione relativa assunta sul territorio dagli eventi oggetto di analisi può essere uno strumento rilevante per spiegare la loro variabilità spaziale. L'interazione spaziale degli eventi si formalizza tramite concetti geometrici quali la distanza (non necessariamente intesa come una metrica in senso stretto), la direzione o l'adiacenza dei siti su cui gli eventi si verificano. La distanza è uno strumento utile a spiegare la variabilità spaziale di un fenomeno quando questa si dimostra sensibile allo scostamento degli eventi oggetto di studio da siti con particolari caratteristiche.¹²

In base alle finalità si possono identificare quattro tipologie di studi descrittivi di epidemiologia spaziale³:

- a. mappe di rischio (*disease mapping*),
- b. correlazione geografica (*geographical correlation studies*),
- c. valutazione del rischio in riferimento a fonti puntuali o lineari di emissione (*point and point-line source studies*),
- d. identificazione di cluster (*cluster detection*).

- a. Le mappe di rischio identificano l'insieme di analisi volte a sintetizzare la variazione spaziale o spazio-temporale del rischio di una patologia. Tale analisi risulta rilevante, tra l'altro, per fornire, attraverso la rappresentazione cartografica, una descrizione della distribuzione territoriale degli eventi sanitari, per indicare le necessità sanitarie della popolazione al fine della pianificazione di interventi o di ulteriori approfondimenti, per fornire indicazioni di carattere generale sul fenomeno osservato ed elaborare ipotesi investigative da verificare successivamente in studi specifici.^{12,13}
- b. Gli studi di correlazione geografica sono studi tipicamente di tipo ecologico e si focalizzano essenzialmente sugli aspetti eziologici delle patologie esaminando la variabilità spaziale di aspetti sanitari in relazione all'esposizione a variabili ambientali. Le associazioni misurate in studi ecologici, e quindi in aggregati di individui, non riflettono necessariamente le associazioni esistenti a livello individuale e possono esprimere distorsioni create dalla cosiddetta *fallacia ecologica*: il possibile errore che si commette estrapolando associazioni osservate da livello di popolazione a livello individuale. Essi possono produrre informazioni riferibili ad aggregazioni collettive (ecologiche) e comunque difficilmente inferenziabili all'individuo.^{1,12,14,15}
- c. Gli studi di valutazione del rischio in riferimento a fonti di emissione mirano ad identificare incrementi "focali" del rischio di determinati eventi sanitari come effetto di esposizione ad agenti emessi da specifiche fonti sia di tipo puntuale^{16,17,18,19,20} (per esempio un impianto produttivo o una stazione emittente onde elettromagnetiche) che lineare^{21,22,23} (per esempio una strada o una linea elettrica). Questi studi hanno una portata tipicamente locale dato che è realistico ipotizzare che la sorgente di rischio esaurisca il suo potenziale effetto con una risoluzione geografica relativamente ristretta.¹²
- d. Gli studi sull'identificazione dei cluster si pongono in genere come obiettivo quello di evidenziare la presenza di aggregati spazio-temporali di eventi senza una precisa premessa eziologica e mirano essenzialmente a supportare attività di monitoraggio e sorveglianza del territorio sebbene, spesso, possano costituire una fase preliminare per studi analitici volti ad approfondire cause specifiche di eccessi di rischio per una determinata patologia.^{12,24,25}

Le quattro tipologie di studi spaziali consentono di perseguire molteplici obiettivi di indagini in ambito epidemiologico. Tale distinzione, tuttavia, acquista un significato puramente didattico dato che, nell'implementazione degli studi, spesso sussistono sovrapposizioni tra le diverse metodologie. Ciò che accomuna tutte le categorie indicate è che alle unità di analisi sono associate specifiche informazioni spaziali (ed eventualmente temporali) che identificano le locazioni delle stesse.

In epidemiologia ambientale spesso la valutazione del rischio sanitario viene resa difficoltosa dalla complessità della stima delle esposizioni dovuta alla presenza di esposizioni multiple, composizione eterogenea, basse concentrazioni, diffusione e variabilità spazio-temporale dei contaminanti. Tali caratteristiche, tra loro collegate, incidono in vario modo sulla validità e/o accuratezza delle stime di esposizione e sulla valutazione dei rischi sanitari ad esse associati²⁶. Specie nell'ultimo decennio, si è cercato di ovviare all'incertezza nelle stime dell'esposizione umana attraverso l'effettuazione di misure dirette di esposizione mediante tecniche di biomonitoraggio umano. La misura di biomarcatori di esposizione permette, infatti, di superare i limiti della stima indiretta delle esposizioni quali ad esempio le diverse vie di interazione tra organismo umano e contaminanti o la suscettibilità individuale. Tuttavia, per quanto concerne le patologie cronico-degenerative, spesso caratterizzate da lunghi periodi di latenza, è molto difficoltoso ricostruire retrospettivamente i diversi scenari di esposizione ed in questi casi il biomonitoraggio è di scarsa utilità in quanto solo pochi indicatori biologici (come ad esempio la diossina) possono essere utili nella misura delle esposizioni su ampie finestre temporali. Laddove non possa essere intrapresa una strategia di monitoraggio dell'esposizione su base

individuale è appropriato valutare l'opportunità di impiegare strategie di valutazione dell'esposizione basate su un adeguato bilanciamento tra misure ambientali e modelli previsionali²⁷. Questi ultimi trovano largo impiego nella valutazione delle esposizioni ad inquinanti aerodispersi; studi di validazione basati sulla correlazione tra le misure dirette degli inquinanti e le concentrazioni stimate con modelli previsionali, hanno indicato una buona affidabilità dei modelli di dispersione, particolarmente per stime a lungo termine²⁸. Sebbene la maggior parte dei modelli disponibili siano in grado di offrire valide approssimazioni dei livelli di concentrazione, su brevi e lunghi periodi temporali, in assenza di stime espositive ritenute più attendibili, viene tutt'oggi utilizzata la sola distanza dalla sorgente per la determinazione indiretta di una stima delle esposizioni. In questi casi, l'assegnazione ai dati amministrativi territoriali (indirizzi, codici postali, aree amministrative come distretti scolastici, sezioni elettorali, ecc.) di coordinate geografiche consente di georeferenziare (posizionare sul territorio un oggetto secondo un sistema di coordinate) le informazioni disponibili di carattere ambientale e/o sanitario riferite ai soggetti in studio.

Le metodologie statistiche appropriate per le analisi spaziali richiedono, necessariamente per gli studi in cui l'unità di analisi è l'individuo, l'impiego di un'informazione georeferenziata a livello puntuale. Nelle situazioni, per altro frequenti, in cui il dettaglio spaziale individuale non è disponibile si ricorre ad analisi di dati aggregati per area geografica.

A tale scopo la georeferenziazione dei dati attraverso un sistema GIS (Geographic Information System), come osservato in precedenza, può costituire un'informazione adeguata ai fini della costruzione di sistemi di sorveglianza epidemiologica. E' bene sottolineare che negli studi caso-controllo, se è vero che il dettaglio spaziale dei casi coinvolti nell'analisi è spesso estrapolabile da database sanitari, quali registri di patologia, delle cause di morte, ecc., può essere difficoltoso e costoso il reperimento di un'altrettanto dettagliata informazione sui controlli. Tali informazioni, in questo disegno di studio, provengono in genere da altre fonti della statistica ufficiale (database di enti amministrativi) o dalla raccolta dei dati mediante interviste dirette.

In sintesi, nell'ambito degli studi in cui il modello di analisi statistica prevede l'introduzione del dato spaziale, gli studi descrittivi sono volti a fornire una preliminare istruttoria di analisi di una determinata popolazione; solo successivamente, e qualora valutazioni tecnico-economiche e di ponderazione del valore informativo ne verificano la fattibilità, le indagini epidemiologiche utilizzeranno studi osservazionali analitici al fine di individuare il rapporto causale tra esposizione ambientale ed occorrenza di eventi sanitari.

Obiettivi

Il presente contributo intende fornire spunti di riflessione sugli studi epidemiologici che indagano i fenomeni sanitari in relazione alla variazione spaziale del rischio. In particolare si propone di effettuare alcune valutazioni sull'utilizzo dei dati residenziali correntemente disponibili nella conduzione degli studi spaziali.

La storia residenziale individuale dei soggetti partecipanti ad un'indagine epidemiologica può essere raccolta attraverso la consultazione delle anagrafi comunali o attraverso interviste dirette.

Obiettivo specifico dello studio è quello di valutare il livello di comparabilità della storia residenziale e della storia abitativa reale di gruppi di popolazione ricostruite rispettivamente attraverso la consultazione delle liste anagrafiche comunali e la somministrazione di questionari mediante intervista. Nella presente indagine si vuole valutare sia la comparabilità della storia residenziale e della storia abitativa dell'intero periodo in studio che la comparabilità della sola ultima residenza o della residenza principale.

MATERIALI E METODI

Dati utilizzati

I dati analizzati nel presente lavoro provengono dai dati raccolti per la conduzione di uno studio di epidemiologia analitica, attuato con tecnica caso-controllo su base di popolazione, al fine di valutare il rischio di morte per tumori del sistema emolinfopoietico nei residenti nei territori comunali adiacenti alla raffineria API di Falconara Marittima. Allo stato attuale non sono stati ancora pubblicati i risultati dello studio giunto recentemente alla fase conclusiva, mentre si rimanda, per eventuale interesse, al testo integrale del report pubblicato al termine della fase precedente nel gennaio 2009²⁹.

I dati utilizzati riguardano la storia abitativa dei soggetti arruolati nell'indagine epidemiologica e rappresentano i casi (residenti nell'area di studio che sono deceduti per tumore del sistema emolinfopoietico nel periodo di osservazione) e i controlli (residenti nell'area di studio appaiati ai casi per età e sesso al verificarsi del decesso). La loro storia abitativa è relativa ai 40 anni precedenti all'occorrenza dell'evento. Le informazioni sul comune, via, numero civico, esponente e periodo di permanenza sono state raccolte tramite un questionario compilato da un parente del soggetto arruolato. Dell'intervistato è stata acquisita l'informazione sul grado di parentela con il caso/controllo e sul livello di istruzione. Parallelamente, le stesse informazioni sulla storia residenziale dei soggetti in studio sono state acquisite, relativamente ai 20 anni antecedenti l'evento, mediante la consultazione delle liste anagrafiche comunali.

Le residenze anagrafiche ed abitative sono state georeferenziate con rilevazione manuale GPS (*Global Positioning System*) e, laddove non possibile rintracciare fisicamente l'indirizzo dell'arruolato, con l'ausilio di software per la ricerca automatica delle coordinate geografiche (CSH Geolocate) o l'utilizzo della toponomastica informatizzata fornita dalla locale amministrazione comunale. Successivamente, sono state misurate le distanze delle residenze individuate dalla sorgente di rischio per la stima delle esposizioni individuali (Mapinfo v. 10.0.1).

Analisi statistica

Il periodo vissuto nella residenza è stato specificato da una data iniziale ed una data finale; esse sono state espresse in *anni* (AAAA) nell'archivio proveniente dai questionari (DBQ) ed in *giorni/mesi/anni* (GG/MM/AAAA) in quello dell'anagrafe comunale (DBA). Al fine di rendere confrontabili i periodi di permanenza dei due archivi, il formato data di DBA è stato convertito nel formato data di DBQ. In entrambi i database, inoltre, per ridurre i possibili bias di informazione, sono stati eliminati gli anni-residenza per i quali si è verificata la sovrapposizione di due indirizzi di residenza nello stesso anno solare (esempi in tabella 1).

Tabella 1. Esempi di esclusione di anni-residenza per cambio di indirizzo del soggetto nell'anno solare e di conversione del formato delle date provenienti dall'anagrafe comunale, GG/MM/AAAA, nel formato delle date provenienti dal questionario, AAAA.

Dati dal Registro Anagrafe Comunale (DBA)								
Formato data di residenza: GG/MM/AAAA								
Id. Soggetto	Res. 3		Res. 2		Res. 1		Anni esclusi	Anni analizzati
	da	a	da	a	da	a		
1			01/01/1981	22/11/1995	23/11/1995	31/12/1995	1995	1981-1994
2	01/01/1982	24/05/1984	25/05/1984	10/12/1992	11/12/1992	31/12/1996	1984, 1992	1982-1983, 1985-1991, 1993-1996
Dati dal Questionario (DBQ)								
Formato data di residenza: AAAA								
Id. Soggetto	Res. 3		Res. 2		Res. 1		Anni esclusi	Anni analizzati
	da	a	da	a	da	a		
1			1964	1970	1970	1998	1970	1964-1969, 1971-1998
2	1958	1980	1980	1991	1991	1998	1980, 1991	1958-1979, 1981-1990, 1992-1998

La completezza è stata valutata sul totale degli anni-residenza dei soggetti per i quali, i dati provenienti dai due archivi, DBA e DBQ, presentavano almeno una delle seguenti informazioni sulle residenze: comune, via, civico, esponente. Il dato atteso è stato definito "completo" se erano noti comune, via e numero civico di residenza. La completezza è stata determinata come rapporto tra gli anni-residenza per i quali il dato era completo e il totale degli anni di residenza confrontabili.

La concordanza delle informazioni derivanti dalle due sorgenti di dati è stata valutata sulle residenze storiche comparabili di 362 soggetti relative a complessivi 5710 anni-residenza per i quali, per ogni anno, le informazioni residenziali nei due archivi sono state ritenute concordi. Al fine di rendere confrontabili le residenze anagrafiche ed abitative sono state individuate due definizioni di concordanza differenti a seconda dei criteri di selezione più o meno stringenti. Secondo una definizione più restrittiva di concordanza, sono stati ritenuti concordi i dati residenziali che, non dando un particolare peso all'esponente, coincidevano per:

- comune, via, numero civico ed esponente;
- comune, via, numero civico ed un esponente diverso o mancante;
- comune, via e mancante numero civico in entrambi i database.

Si è ritenuto considerare concordi anche le informazioni residenziali che, in entrambi i database, sono risultate mancanti del numero civico, in quanto in tal caso il software per la georeferenziazione automatica degli indirizzi avrebbe attribuito loro la stessa posizione geografica; posizione che avrebbe coinciso con il punto centrale dell'arco stradale considerato.

Non sono state ritenute concordi le residenze che presentavano:

- uguale comune e via (numeri civici diversi ed esponenti diversi o mancanti);
- uguale comune e via (numero civico presente in un solo database);
- uguale comune ma diversa via (con le informazioni presenti su entrambe le vie);
- uguale comune ma diversa via (con l'informazione presente solo su una via);
- diverso comune (con le informazioni su entrambi i comuni presenti).

Successivamente, dato il possibile recall bias sulle informazioni residenziali ottenute dai questionari circa il ricordo da parte dell'intervistato riguardo il numero civico, è stata valutata

anche la concordanza dei due database (seconda definizione di concordanza) includendo tra i criteri, oltre a quelli sopra indicati:

- comune, via, numero civico ed esponente;
- comune, via, numero civico ed un esponente diverso o mancante;
- comune, via e mancante numero civico in entrambi i database.

anche:

- uguale comune e via (numeri civici diversi ed esponenti diversi o mancanti);
- uguale comune e via (numero civico presente in un solo database).

Conseguentemente, considerando quest'ultima definizione di concordanza, non sono state ritenute concordi le residenze che presentavano:

- uguale comune ma diversa via (con le informazioni presenti su entrambe le vie);
- uguale comune ma diversa via (con l'informazione presente solo su una via);
- diverso comune (con le informazioni su entrambi i comuni presenti).

Tenendo conto che le osservazioni misurate sullo stesso soggetto tendono ad essere correlate, la percentuale di concordanza e il relativo intervallo di confidenza al 95% di probabilità (IC 95%) sono stati stimati attraverso un modello marginale di regressione logistica basato su equazioni di stime generalizzate (GEE, *Generalized Estimating Equation*)³⁰ assumendo una matrice di correlazione di tipo scambiabile.

Nel modello marginale è stato valutato l'effetto dell'anno di riferimento della residenza, ridefinito come una variabile dicotomica con il cut-off nell'anno 1980, come fattore in grado di esercitare una possibile influenza nel ricordo dell'informazione comunicata nei questionari (recall bias).

In mancanza della disponibilità di migliori stime di esposizione individuale, spesso negli studi di epidemiologia ambientale, così come nell'indagine caso-controllo citata in precedenza, è utilizzata, quale proxy di esposizione, l'informazione relativa alla distanza della residenza dalla sorgente di rischio. Nel presente studio a supporto della concordanza dei dati residenziali è stato effettuato un confronto tra le distanze delle residenze, anagrafiche ed abitative, tratte dai rispettivi archivi (DBA e DBQ), dalla sorgente di rischio attraverso la determinazione dell'indice di correlazione lineare di Pearson.

I risultati sono stati ritenuti statisticamente significativi al livello di probabilità del 95% (p-value <0,05). Tutte le analisi sono state effettuate utilizzando il software Sas System v.9.2.

RISULTATI

Lo studio ha analizzato le informazioni sulle residenze di 362 soggetti per i soli anni di residenza nei comuni dell'area in studio (Falconara Marittima, Chiaravalle e Montemarignano) confrontando, per ogni anno-residenza, i dati provenienti dai questionari dello studio epidemiologico caso-controllo e i dati registrati nelle liste comunali anagrafiche, per complessivi 5710 anni (anni-residenza).

Di questi, 4953 anni-residenza (86,7%) sono risultati pari o successivi al 1980 (≥ 1980), mentre i restanti 757 (13,3%) sono da attribuire al periodo precedente (< 1980).

Dei 362 intervistati si è osservato il grado di parentela con i soggetti in studio ed il livello di scolarità. Di essi, l'82,3% (n=298) presentava con gli arruolati un livello di parentela "stretto" (coniuge/figlio/fratello/genitore), il 10,2% (n=37) un livello di parentela "largo" (nipote/zio/cognato) e il restante 7,5% (n=27) aveva un diverso grado di parentela. Per quanto concerne il livello di istruzione, gli intervistati sono stati suddivisi in due gruppi: istruzione medio-alta e bassa; al primo gruppo appartengono il 45,1% di essi (n=163) e sono coloro che avevano conseguito una licenza media superiore, un diploma universitario o una laurea, mentre nel secondo gruppo risulta il restante 54,9% (n=198), cioè coloro che non avevano alcuna istruzione o un grado di istruzione inferiore (licenza elementare, licenza media inferiore). Solo per un intervistato l'informazione sul grado di istruzione è risultata mancante (tabella 2).

Tabella 2. Descrittiva del periodo di residenza dei soggetti in studio e del livello di istruzione e del grado di parentela degli intervistati

Periodo	Frequenza assoluta	Perc. assoluta (%)	Frequenza cumulata	Perc. cumulata (%)
≥ 1980	4953	86,7	4953	86,7
< 1980	757	13,3	5710	100,0
Istruzione	Frequenza assoluta	Perc. assoluta (%)	Frequenza cumulata	Perc. cumulata (%)
Bassa	198	54,9	198	54,9
Medio/Alta	163	45,1	361*	100,0
Parentela	Frequenza assoluta	Perc. assoluta (%)	Frequenza cumulata	Perc. cumulata (%)
Stretta	298	82,3	298	82,3
Larga	37	10,2	335	92,5
Altra	27	7,5	362	100,0

*un valore mancante sul grado di istruzione dell'intervistato

Analisi sull'intero campione degli anni-residenza

Sono stati 362 i soggetti per i quali si sono ricercate nei due database le residenze anagrafiche ed abitative.

Atteso che è stata ritenuta completa l'informazione sulla residenza quando recante comune, via e numero civico, è possibile ritenere che, come risulta dalla tabella sottostante (tabella 3), gli archivi utilizzati abbiano mostrato un buon livello di completezza di informazione. Dei 5710 anni di residenza è stata rilevata una completezza del 99,8% dei dati estrapolati dall'archivio delle anagrafi comunali e del 96,6% dei dati acquisiti dai questionari. I due archivi, DBA e DBQ, hanno evidenziato il 96,6% di informazioni complete riferibili agli stessi anni-residenza.

Tabella 3. Anni di residenza con informazioni provenienti dai questionari e dalle anagrafi comunali complete (comune, via e numero civico di residenza) e non complete

		Questionario		
		completo	non completo	
Anagrafica	completa	5515	181	5696
	non completa	0	14	14
		5515	195	5710

Percentuale di completezza		
Completezza	Anagrafica:	99,8%
Completezza	Questionari:	96,6%
Completezza	Anagrafica e Questionari:	96,6%

Definizione restrittiva di concordanza

La percentuale di concordanza delle informazioni sulle residenze anagrafiche ed abitative per i soli anni-residenza nei comuni all'interno dell'area (tabella 4), con uguali comune, via e numero civico o con i numeri civici assenti in entrambi i database, è risultata pari all'84,5% (IC95%: 83,6%-85,5%).

La percentuale di anni-residenza concordi dopo il 1980 è stata pari all'87,7% (4342 su 4953) mentre quella relativa al periodo precedente il 1980 è stata del 64,1% (485 su 757). Il periodo di residenza ≥ 1980 vs < 1980 è risultato essere un fattore in grado di esercitare una distorsione da ricordo (p-value = 0,0001).

Tenendo conto della correlazione delle informazioni misurate sullo stesso soggetto, la percentuale di concordanza, aggiustata per il periodo di residenza (≥ 1980 vs < 1980), è stata dell'87,6% (IC95%: 84,2%-90,3%).

In sintesi, per l'intero campione, la differenza tra l'effettivo/reale domicilio e la residenza registrata nelle anagrafi comunali è stata stimata al 12,4% (IC95%: 9,7%-15,9%).

Definizione ampia di concordanza

Il confronto delle informazioni sulle residenze anagrafiche ed abitative per i soli anni-residenza nei comuni all'interno dell'area (tabella 4), con almeno uguali comune e via in entrambi i database, ha prodotto una percentuale di concordanza pari all'89,7% (IC95%: 88,8%-90,4%).

La percentuale di anni-residenza concordi dopo il 1980 è risultata pari al 92,2% (4566 su 4953) mentre quella relativa al periodo precedente il 1980 è stata del 73,1% (553 su 757). Il periodo di residenza (≥ 1980 vs < 1980) è risultato essere un fattore in grado di esercitare una distorsione da ricordo (p-value = 0,0530).

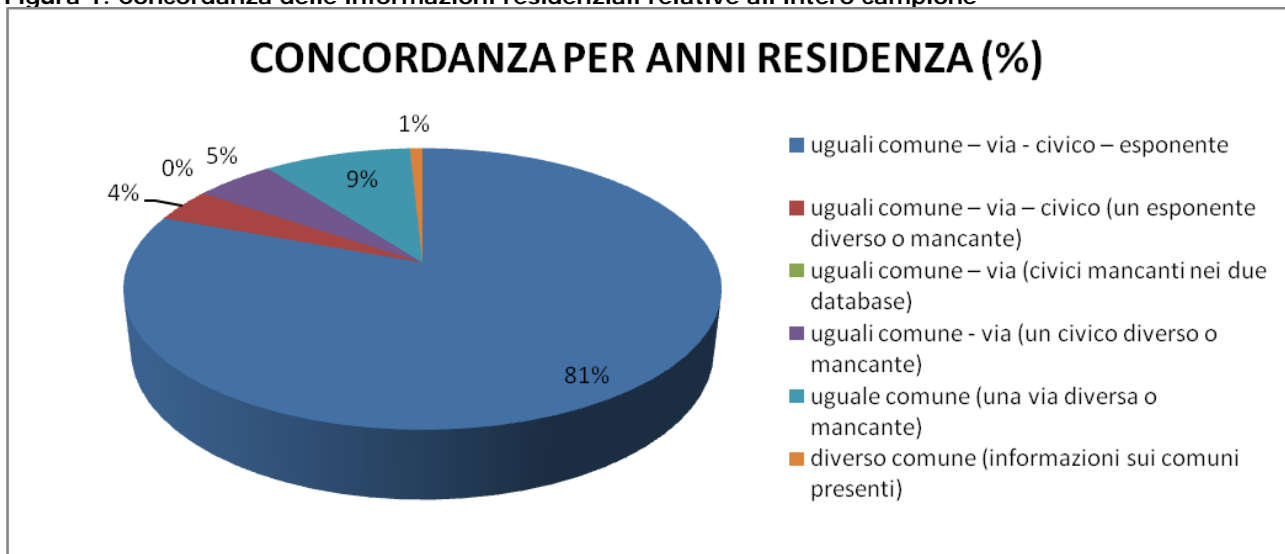
Tenendo conto della correlazione delle informazioni misurate sullo stesso soggetto, la percentuale di concordanza, aggiustata per il periodo di residenza (≥ 1980 vs < 1980), è stata del 92,1% (IC95%: 89,3%-94,3%).

In sintesi, per l'intero campione, secondo i criteri più estesi di concordanza, la differenza tra l'effettivo/reale domicilio e la residenza registrata nelle anagrafi comunali è stata del 7,9% (IC95%: 5,7%-10,7%).

Tabella 4. Concordanza delle informazioni residenziali relative all'intero campione

Criteri di Concordanza/discordanza	Concordanza in anni-residenza			
	Frequenza assoluta	perc. ass. (%)	Frequenza cumulata	perc. cum. (%)
uguali comune – via - civico – esponente	4606	80,7	4606	80,7
uguali comune – via – civico (un esponente diverso o mancante)	221	3,9	4827	84,5
uguali comune – via (civici mancanti nei due database)	0	0	4827	84,5
uguali comune - via (un civico diverso o mancante)	292	5,1	5119	89,7
uguale comune (una via diversa o mancante)	544	9,5	5663	99,2
diverso comune (informazioni sui comuni presenti)	47	0,8	5710	100,0

Figura 1. Concordanza delle informazioni residenziali relative all'intero campione



Analisi sull'ultima residenza

Definizione restrittiva di concordanza

Dal confronto dei database relativi all'ultimo anno di residenza è stata rilevata una concordanza tra le residenze abitative ed anagrafiche (tabella 5) dell'87,9% (IC95%: 84,1%-90,8%).

Definizione ampia di concordanza

La percentuale di concordanza delle informazioni sulle ultime residenze anagrafiche ed abitative (tabella 5), con almeno uguali comune e via in entrambi i database, è risultata pari al 91,4% (IC95%: 88,1%-93,9%).

Tabella 5. Concordanza delle informazioni sulle residenze anagrafiche ed abitative relative all'ultimo anno di residenza nell'intero campione

Criteri di Concordanza/discordanza	Concordanza residenze			
	Frequenza assoluta	perc. ass. (%)	Frequenza cumulata	perc. cum. (%)
uguali comune – via - civico – esponente	16	4,4	16	4,4
uguali comune – via – civico (un esponente diverso o mancante)	302	83,4	318	87,8
uguali comune – via (civici mancanti nei due database)	0	0	318	87,8
uguali comune - via (un civico diverso o mancante)	13	3,6	331	91,4
uguale comune (una via diversa o mancante)	26	7,2	357	98,6
diverso comune (informazioni sui comuni presenti)	5	1,4	362	100,0

Analisi sulla residenza principale

Definizione restrittiva di concordanza

Il confronto delle informazioni sulle residenze principali anagrafiche ed abitative (tabella 6) ha prodotto una percentuale di concordanza pari all'86,5 % (IC95%: 82,5% – 89,6%).

Definizione ampia di concordanza

Dai risultati del confronto dei due archivi risulta che il 90,6 % (IC95%: 87,1% – 93,2%) delle informazioni residenziali sono concordi (tabella 6).

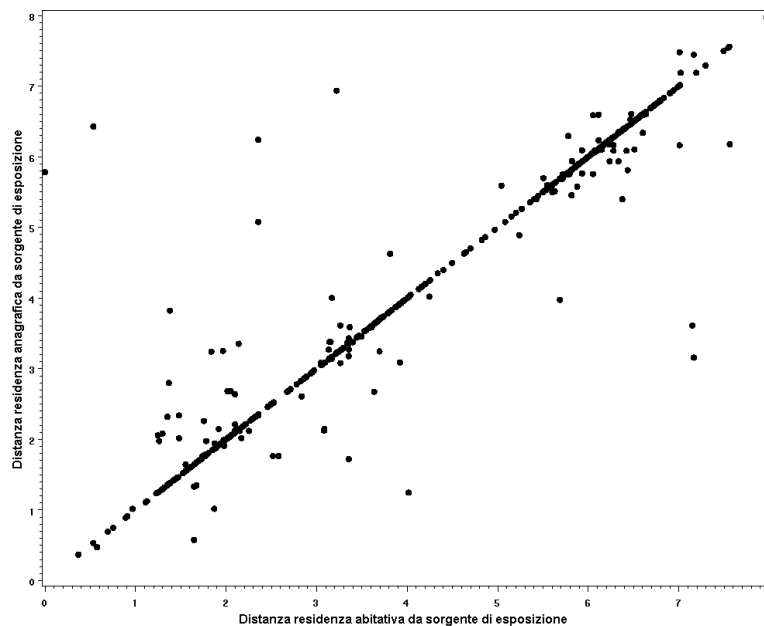
Tabella 6. Concordanza delle informazioni sulle residenze anagrafiche ed abitative relative alla residenza principale nell'intero campione

Criteri di Concordanza/discordanza	Concordanza in anni-residenza			
	Frequenza assoluta	perc. ass. (%)	Frequenza cumulata	perc. cum. (%)
uguali comune – via - civico – esponente	13	3,6	13	3,6
uguali comune – via – civico (un esponente diverso o mancante)	300	82,9	313	86,5
uguali comune – via (civici mancanti nei due database)	0	0	313	86,5
uguali comune - via (un civico diverso o mancante)	15	4,1	328	90,6
uguale comune (una via diversa o mancante)	29	8,0	357	98,6
diverso comune (informazioni sui comuni presenti)	5	1,4	362	100,0

Correlazione tra le distanze delle residenze storiche dalla sorgente di esposizione

Sono stati utilizzati i dati residenziali dei soggetti arruolati nell'ambito dello studio caso-controllo citato in precedenza estratti dall'anagrafe comunale e dal questionario per la stima dell'esposizione individuale ambientale agli inquinanti emessi dalla raffineria. Nello specifico l'indagine ha fatto riferimento, per ogni anno-residenza, alla misura della distanza delle residenze anagrafiche ed abitative dal centroide della sorgente di rischio (figura 2). Tenendo conto che per 2 anni-residenza non è stato possibile calcolare la distanza recettore-fonte di esposizione, l'indice di correlazione di Pearson tra le due distanze, calcolato su 5708 anni-residenza, è risultato pari a 0,97 (p-value <0,0001).

Figura 2. Distanze delle residenze anagrafiche ed abitative dalla sorgente di esposizione di 5708 anni-residenza



DISCUSSIONE

Nel presente studio è stata valutata la comparabilità della storia residenziale e della storia abitativa dei soggetti per i quali sono stati raccolti gli indirizzi delle residenze e delle abitazioni reali attraverso, rispettivamente, la consultazione delle liste anagrafiche comunali e la somministrazione diretta di questionari.

Nella conduzione di un'indagine epidemiologica, idealmente, la conoscenza dei dati sulla popolazione in studio dovrebbe comprendere informazioni dettagliate sulle caratteristiche individuali, sui movimenti demografici, sulle esposizioni ad eventuali sorgenti di rischio e sulle registrazioni di carattere sanitario. Solitamente queste informazioni, siano esse individuali o aggregate, sono variamente disponibili; ciò comporta che le diverse metodologie statistiche di volta in volta impiegate si adeguino al livello di dettaglio delle informazioni raccolte.

L'approccio descrittivo in epidemiologia ambientale ha il merito di evidenziare potenziali difformità nella distribuzione territoriale dello stato di salute delle popolazioni ed ha il ruolo di far emergere possibili spunti eziologici per l'avvio di indagini osservazionali analitiche. E' ritenuto pertanto utile, in questo campo, un utilizzo dello studio geografico come elemento di integrazione delle conoscenze biologiche, patogenetiche e cliniche e per consentire di individuare situazioni in cui sia opportuno condurre studi epidemiologici analitici più approfonditi.³¹

Agli studi descrittivi di epidemiologia spaziale, però, per motivazioni legate prevalentemente alla complicata acquisizione dei dati necessari allo svolgimento dell'indagine e all'ingente impegno richiesto in termini di risorse economiche ed umane, solo raramente fanno seguito studi epidemiologici di tipo analitico in grado di misurare l'associazione tra salute ed esposizioni ambientali e valutare ipotesi di rilevanza eziologica utilizzando gli individui come unità di analisi.

Entrambi i disegni di studio, comunque, presentano limitazioni intrinseche legate all'assenza di dati di esposizione certi e verificabili individualmente. Negli studi di epidemiologia ambientale, spesso, l'interesse del ricercatore si polarizza sulla valutazione retrospettiva dell'esposizione e, in genere, vengono formulate ipotesi eziologiche di patologie con relativamente lungo periodo di latenza. Per questi motivi, in assenza di misure dirette o di modellizzazione dell'esposizione, quest'ultima viene stimata mediante l'utilizzo di un indicatore indiretto quale la misura della distanza delle abitazioni dalla fonte espositiva.^{32,33,34,35}

In tale contesto, la possibilità di disporre di informazioni georeferenziate risulta di fondamentale importanza per affrontare analisi quantitative di fenomeni connessi con la qualità dell'ambiente ed individuare le interazioni presenti fra i fenomeni ambientali e gli eventi sanitari. Nei casi di utilizzo della distanza tra residenza e sorgente di rischio come proxy di esposizione individuale, le indagini epidemiologiche spaziali possono far riferimento, tra le altre, all'*esposizione residenziale cumulata*; questa rappresenta un valido indicatore che include informazioni sulla pregressa storia individuale abitativa delle persone coinvolte nello studio considerando la distanza delle abitazioni dalla presunta sorgente emissiva e il tempo di permanenza in ciascuna di tali residenze. Altri indicatori frequentemente utilizzati per studi relativi al rischio abitativo/residenziale sono rappresentati dall'*ultima residenza*^{36,37,38} e dalla *residenza principale*^{33,39}. Nel primo caso si tratta della residenza abitata nel periodo di tempo più recente rispetto al verificarsi dell'evento sanitario oggetto d'indagine, mentre nel secondo caso si tratta della residenza abitata per il periodo di tempo più lungo nel corso del periodo di vita studiato. L'*esposizione residenziale cumulata*, tuttavia, sembra essere più accurata degli altri indicatori per i quali è nota solo l'informazione distanza della abitazione/residenza dalla sorgente di rischio e non la durata del tempo di permanenza in tali residenze e per i quali non si hanno informazioni sull'intera storia residenziale.

Occorre tenere presente, tuttavia, che tutti gli indicatori considerati sono gravati da un certo grado di inesattezza rispetto al livello di esposizione ambientale dei soggetti coinvolti imputabile alla loro incostante permanenza in abitazione per tutto il periodo in studio.

Quale che sia l'indicatore individuale di esposizione ambientale utilizzato, negli studi epidemiologici spaziali che si avvalgono della distanza del domicilio dalla sorgente di rischio per valutare l'associazione tra l'occorrenza di eventi sanitari e l'esposizione a fattori di rischio ambientale, è necessario conoscere l'esatta localizzazione georeferenziata delle residenze/abitazioni dei soggetti arruolati. Le informazioni sulle residenze, come detto in precedenza, possono essere ottenute dalla rilevazione dei dati anagrafici raccolti in database predisposti per fini amministrativi o dalla raccolta della storia abitativa di gruppi di popolazione mediante la somministrazione di questionari. Successivamente, mediante tecniche di georeferenziazione che si avvalgono dell'utilizzo manuale di strumenti GPS (Global Positioning System), servizi WEB, toponomastica informatizzata gestita in genere dalle amministrazioni locali e software per la ricerca automatica delle coordinate geografiche, gli indirizzi vengono geocodificati e riferiti ad un sistema di coordinate geografiche.

E' chiaro che la valutazione di completezza e di accuratezza dei dati è una condizione indispensabile affinché essi possano essere usati per studi epidemiologici. In letteratura è presente una discreta quantità di lavori scientifici circa la valutazione della qualità dei dati sanitari estrapolabili dai certificati di morte e dalle schede di dimissione ospedaliera^{40, 41, 42, 43, 44} (noti come *re-abstract study*⁴⁵) al fine di un loro utilizzo per la conduzione di studi epidemiologici. Sono tutt'altro che numerosi, invece, i richiami bibliografici a studi che, per lo stesso fine, hanno indagato la validità dei dati desumibili dagli archivi informatizzati comunali relativi alle residenze anagrafiche.

L'obiettivo dello studio, a tal proposito, è stato quello di valutare il livello di completezza e di concordanza dei due archivi utilizzati quali fonti informative per gli indirizzi abitativi ed anagrafici. Come evidenziato nella sezione dei risultati (tabella 3), entrambi gli archivi hanno mostrato un buon livello di completezza di informazione; comune, via e numero civico sono noti nella quasi totalità dei dati ottenuti dall'archivio delle anagrafi comunali (99,8%) e, sebbene in presenza di possibile *recall bias*, nel 96,6% dai dati acquisiti dai questionari.

La concordanza delle informazioni relative alle residenze anagrafiche ed abitative, tenendo conto della correlazione delle informazioni misurate sullo stesso soggetto, è risultata elevata mostrando una percentuale dell'87,6%. A conferma del buon livello di concordanza evidenziato dai dati ottenuti dalle due fonti informative è il risultato del calcolo dell'indice di correlazione di Pearson (0,97) sulle distanze delle residenze anagrafiche/abitative dalla sorgente di esposizione. Tale coefficiente può assumere valori compresi tra [-1; +1]. Quando l'indice di correlazione vale 0 le due variabili sono indipendenti, mentre se è negativo o è positivo le due serie di dati sono inversamente o direttamente correlate tra loro. E' chiaro che più il valore del coefficiente si avvicina a 1 e più la correlazione lineare positiva sarà massima.

Questo tipo di valutazione, pertanto, ha evidenziato la considerevole comparabilità dei dati raccolti dai sistemi informativi comunali con i dati inerenti le residenze abitative e quindi l'utilizzabilità del supporto informativo comunale per fornire dati residenziali utilizzabili a fini epidemiologici.

Negli studi epidemiologici, la raccolta delle sole residenze anagrafiche, rispetto all'insieme delle informazioni sulle residenze abitative e sulla eventuale presenza di covariate ottenibili con la somministrazione del questionario, permette di raggiungere un più alto numero di soggetti, ma risulta meno capace di stimare le esposizioni e, conseguentemente, il rischio dell'occorrenza di malattia associabile alle stesse. La residenza, infatti, non sempre coincide con l'abitazione (misclassificazione dell'esposizione), né i possibili effetti di confondimento, quali ad esempio le informazioni sull'abitudine al fumo o il domicilio nelle vicinanze di potenziali altre fonti di esposizione, possono essere acquisiti attraverso la raccolta di dati presenti nelle anagrafi comunali.

In generale si può presumere che l'approccio della valutazione delle esposizioni individuali mediante l'utilizzo delle residenze anagrafiche, a fronte di quello che utilizza le residenze abitative, potendo contare su un maggior numero di osservazioni disponga di una potenza statistica maggiore e quindi meno soggetto a fluttuazione di tipo statistico. Lo stesso, tuttavia,

essendo meno accurato nella stima dell'esposizione potrebbe portare a possibili distorsioni nella valutazione del rischio.

CONCLUSIONI

Nello studio caso-controllo a cui si è fatto riferimento per l'estrapolazione dei dati residenziali, il questionario è stato ritenuto dai ricercatori come lo strumento più valido per reperire le informazioni sulle residenze abitative degli arruolati. La rilevazione di quest'ultime mediante intervista, presenta indubbiamente il vantaggio di verificare la possibile azione di fattori confondenti e di individuare il vero domicilio dei soggetti in studio. Tuttavia il buon livello di completezza e l'elevato grado di concordanza tra le residenze anagrafiche e le residenze abitative evidenziata dall'analisi dei dati porta alla conclusione che, visti i vantaggi in termini di disponibilità e costi di reperibilità, l'uso delle residenze anagrafiche possa ritenersi un valido surrogato a quello delle residenze abitative.

In sintesi i dati derivanti dai sistemi informativi delle anagrafi comunali o di altri archivi informatizzati, seppur non completamente sostitutivi dei dati relativi alle residenze abitative raccolti con la somministrazione di un questionario, possono fornire una risorsa concreta per la valutazione epidemiologica delle esposizioni ambientali.

Il citato studio epidemiologico analitico sui residenti in un'area adiacente ad una raffineria, per la stima indiretta delle esposizioni individuali, ha utilizzato in parallelo i due approcci: dati ricavati da questionario con controllo dei confondenti ed il solo utilizzo dei dati residenziali ricavati dalle anagrafi comunali. Successivamente alla completa disponibilità dei risultati dello studio suddetto sarà possibile valutare anche la concordanza dei rischi rilevati con i due metodi sulle medesime popolazioni al netto del peso dei fattori confondenti.

PICCOLO GLOSSARIO

Cluster spaziale	Una aggregazione anomala di eventi (sanitari, economici, etc.) limitata geograficamente.
Esposizione residenziale cumulata	Indicatore che include informazioni sulla pregressa storia individuale abitativa delle persone coinvolte nello studio, considerando la distanza delle abitazioni dalla presunta sorgente emissiva e il tempo di permanenza in ciascuna di tali residenze.
GIS – Geographic(al) Information System	Sistema informativo computerizzato che permette l'acquisizione, la registrazione, l'analisi, la visualizzazione e la restituzione di informazioni derivanti da dati geografici (geo-riferiti).
GPS – Global Positioning System	<p>Il Sistema di Posizionamento Globale, (in inglese: Global Positioning System, abbreviato GPS), è un sistema di posizionamento e navigazione satellitare che fornisce posizione ed orario in ogni condizione meteorologica, ovunque sulla Terra, o nelle sue immediate vicinanze, ove vi sia un contatto privo di ostacoli con almeno quattro satelliti del sistema.</p> <p>Il suo grado attuale di accuratezza è dell'ordine dei metri, in dipendenza dalle condizioni meteorologiche, dalla disponibilità e dalla posizione dei satelliti rispetto al ricevitore, dalla qualità e dal tipo di ricevitore, dalla riflessione del segnale, ecc.</p>
Indice di correlazione di Pearson	Indice che esprime l'intensità del legame associativo tra due variabili.
Intervallo di confidenza	L'intervallo di confidenza fornisce informazioni riguardo alla precisione dei valori ottenuti attraverso lo studio di un campione, in modo tale che si può confidare - con un margine di certezza ragionevole - che un dato intervallo contenga il <i>valore vero</i> ricercato. L'intervallo di confidenza rappresenta un parametro di fondamentale importanza soprattutto negli studi epidemiologici in cui la variabilità del campione (molto spesso dovuta al fatto che il campione è piccolo) potrebbe distorcere l'interpretazione dei risultati.
Residenza abitativa	Informazione residenziale ottenuta tramite questionario.
Residenza anagrafica	Informazione residenziale ottenuta tramite Servizi demografici comunali.
Residenza principale	Residenza abitata per il periodo di tempo più lungo nel corso del periodo di vita studiato.
Significatività statistica	Si ottiene quando determinate misurazioni nell'ambito di uno studio di comparazione portano a ipotizzare che un dato risultato sia non casuale (cioè non una conseguenza di un caso).
Studi descrittivi di epidemiologia spaziale (o geografici)	Studi che effettuano la descrizione e l'analisi delle variazioni geografiche di eventi sanitari in relazione a fattori di rischio demografici, ambientali, comportamentali, genetici e socio-economici. In base alle finalità perseguite, si possono suddividere in quattro tipologie: mappe di rischio, correlazione geografica, valutazione del rischio in riferimento a fonti puntuali o lineari di emissione, identificazione di cluster.
Ultima residenza	Residenza abitata nel periodo di tempo più recente rispetto al verificarsi dell'evento sanitario oggetto d'indagine.

BIBLIOGRAFIA

- ¹ Terracini B. Aree oggetto di bonifica: inquadramento teorico e metodologico. Rapporti ISTISAN 05/1. ([Full text in Italian](#))
- ² Morgenstern H, Thomas D. Principles of study design in environmental epidemiology. *Environ Health Perspect.* 1993 December; 101(Suppl 4): 23–38. ([Abstract](#)) ([Full text](#))
- ³ Elliot P, Wakefield JC, Best NG, Brings SE (2000). Spatial epidemiology: methods and applications. In Elliott, P., Wakefield, J. C., Best, N.G., Brings D. J. editors. *Spatial Epidemiology: Methods and applications*, Oxford: University Press: 3-14.
- ⁴ Elliott P, Wartenberg D. Spatial Epidemiology: Current Approaches and Future Challenges. *Environ Health Perspect.* 2004 June; 112(9): 998–1006.
- ⁵ Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. *Annu Rev Public Health.* 1995;16:61-81. ([Abstract](#))
- ⁶ Wakefield J. Ecologic studies revisited. *Annu Rev Public Health.* 2008;29:75-90. ([Abstract](#))
- ⁷ Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Public Health,* 1988; 127: 893–904. ([Full text](#))
- ⁸ Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998; 88: 216-22. ([Abstract](#)) ([Full text](#))
- ⁹ Terracini B. Piccole aree, grandi problemi. *Epidemiol Prev* 2006; 30: 3-4. ([Article in Italian](#))
- ¹⁰ Elliott P, Martuzzi M, Shaddick, G. Spatial statistical methods in environmental epidemiology: a critique. *Stat Methods Med Res* 1995;4(2):137-59. ([Abstract](#))
- ¹¹ Olsen SF, Martuzzi M, Elliott P. Cluster analysis and disease mapping why, when, and how? A step by step guide. *Br Med J* 1996;313:863-6. ([Abstract](#)) ([Full text](#))
- ¹² Petrucci A, Bocci C, Borgoni R, Civardi M, Salvati N, Salvini S, Vignoli D. Indagine sulla georeferenziazione dei dati nella statistica ufficiale. CoGIS, Commissione per la Garanzia dell'Informazione Statistica. Roma, 2009 ([Full text in Italian](#))
- ¹³ Lawson AB, Williams FL. An introductory guide to disease mapping. New York, NY: John Wiley and Sons, Inc, 2001. ([Summary](#))
- ¹⁴ Martuzzi M, Mitis F. Metodi e strumenti per studi epidemiologici di mortalità su base geografica in Italia. Rapporti ISTISAN 05/1. ([Full text in Italian](#))
- ¹⁵ Bianchi F, Terracini B. Potenzialità, criticità e prospettive dell'integrazione ambiente e salute. Rapporti ISTISAN 05/1. ([Full text in Italian](#))
- ¹⁶ Gouveia N, Prado RR. Spatial analysis of the health risks associated with solid waste incineration: a preliminary analysis. *Rev Bras Epidemiol.* 2010 Mar;13(1):3-10. ([Abstract](#))
- ¹⁷ Biggeri A, Barbone F, Lagazio C, Bovenzi M, Stanta G. Air pollution and lung cancer in Trieste, Italy: spatial analysis of risk as a function of distance from sources. *Environ Health Perspect.* 1996 July; 104(7): 750–754. ([Abstract](#)) ([Full text](#))
- ¹⁸ Chu-Ling Yu, Su-Fen Wang, Pi-Chen Pan, Ming-Tsang Wu, Chi-Kung Ho, Thomas J. Smith, Yi Li, Lucille Pothier, David C. Christiani. Residential Exposure to Petrochemicals and the Risk of Leukemia: Using Geographic Information System Tools to Estimate Individual-Level Residential Exposure. *Am. J. Epidemiol.* (1 August 2006) 164 (3): 200-207. ([Abstract](#)) ([Full text](#))
- ¹⁹ Selvin S, Schulman J, Merrill DW. Distance and risk measures for the analysis of spatial data: a study of childhood cancers. *Soc Sci Med.* 1992 Apr;34(7):769-77. ([Abstract](#))
- ²⁰ Michelozzi P, Kirchmayer U, Capon A, Forastiere F, Biggeri A, Barca A, Ancona C, Fusco D, Sperati A, Papini P, Pierangelini A, Rondelli R, Perucci CA. Leukemia mortality and incidence of infantile leukemia near the Vatican Radio Station of Rome. *Epidemiol Prev.* 2001 Nov-Dec;25(6):249-55. ([Abstract](#)) ([Article in Italian](#))
- ²¹ Salam MT, Islam T, Gilliland FD. Recent evidence for adverse effects of residential proximity to traffic sources on asthma. *Curr Opin Pulm Med.* 2008 Jan;14(1):3-8. ([Abstract](#))
- ²² Draper G, Vincent T, Kroll ME, Swanson J. Childhood cancer in relation to distance from high voltage power lines in England and Wales: a case-control study. *BMJ* 330 : 1290 doi: 10.1136/bmj.330.7503.1290 ([Abstract](#)) ([Full text](#))
- ²³ Kleinerman RA, Kaune WT, Hatch EE, Wacholder S, Linet MS, Robison LL, Niwa S, Tarone RE. Are Children Living Near High-Voltage Power Lines at Increased Risk of Acute Lymphoblastic Leukemia? *Am. J. Epidemiol.* (2000) 151 (5): 512-515. ([Abstract](#)) ([Full text](#))

- ²⁴ Ball W, LeFevre S, Jarup L, Beale L. Comparison of Different Methods for Spatial Analysis of Cancer Data in Utah. *Environ Health Perspect*. 2008 August; 116(8): 1120–1124. ([Abstract](#)) ([Full text](#))
- ²⁵ Beale L, Abellan JJ, Hodgson S, Jarup L. Methodologic Issues and Approaches to Spatial Epidemiology. *Environ Health Perspect*. 2008 August; 116(8): 1105–1110. ([Full text](#))
- ²⁶ Ivano Iavarone. Valutazione dell'esposizione ad inquinanti ambientali. *Rapporti ISTISAN 06/19 Rev.* ([Full text](#)).
- ²⁷ Hodgson S, Nieuwenhuijsen MJ, Colvile R, Jarup L. Assessment of exposure to mercury from industrial emissions: comparing "distance as a proxy" and dispersion modelling approaches. *Occup Environ Med* 2007; 64: 380-388. ([Abstract](#))
- ²⁸ Williams ML. Atmospheric dispersal of pollutants and the modelling of air pollution. In: Harrison RM (Ed.). *Pollution: Causes, Effects and Control*, 3rd edn. Cambridge. *R Soc Chem* 1996:221-40.
- ²⁹ Meneghini E, Baili P, Mariottini M, Micheli A. Indagine epidemiologica presso la popolazione residente a Falconara Marittima e comuni limitrofi. Rapporto finale, gennaio 2009. ([Full text](#))
- ³⁰ Zeger S.L., Liang K., ; Albert P.S., 1988, Models for Longitudinal Data: A Generalized Estimating Equation Approach, *Biometrics*, Vol. 44, 1049-1060.
- ³¹ Comba P. Integration of geographic and analytical approach in environmental epidemiology. *Epidemiol Prev* 1994; 18(58):4-7. ([Abstract](#))
- ³² Biggeri A, Barbone F, Lagazio C, Bovenzi M, Stanta G. Air pollution and lung cancer in Trieste. *Environ Health Perspect* 104:750-754 (1996). ([Abstract](#)) ([Full text](#))
- ³³ Comba P, Ascoli V, Belli S et al. Risk of soft tissue sarcomas and residence in the neighbourhood of an incinerator of industrial wastes. *Occup Environ Med* 2003; 60: 1-4. ([Abstract](#)) ([Full text](#))
- ³⁴ Chellini E, Pizzo AM, Barbieri A et al. Studio geografico sulla mortalità per tumore del polmone nei residenti a Piombino attorno alla locale cokeria. *Epidemiol Prev* 2006; 29(5-6) Suppl 13-16: 50-2 . ([Abstract](#))
- ³⁵ Marinaccio A, Belli S, Binazzi A, Scarselli A, Massari S, Bruni A, Conversano M, Crosignani P, Minerba A, Zona A, Comba P. Residential proximity to industrial sites in the area of Taranto (Southern Italy). A case-control cancer incidence study. *Ann Ist Super Sanità* 2011 | Vol. 47, No. 2: 192-199. [Full text](#)
- ³⁶ López-Cima MF, García-Pérez J, Pérez-Gómez B, Aragonés N, López-Abente G, Tardón A, Pollán M. Lung cancer risk and pollution in an industrial region of Northern Spain: a hospital-based case-control study. *Int J Health Geogr*. 2011; 10: 10. ([Full text](#))
- ³⁷ McConnell R, Berhane K, Yao L, Jerrett M, Lurmann F, Gilliland F, Künzli N, Gauderman J, Avol E, Thomas D, Peters J. Traffic, Susceptibility, and Childhood Asthma. *Environ Health Perspect*. 2006 May; 114(5): 766–772. ([Full text](#))
- ³⁸ Salam MT, Islam T, Gilliland FD. Recent evidence for adverse effects of residential proximity to traffic sources on asthma. *Curr Opin Pulm Med*. 2008 Jan; 14(1):3-8. ([Abstract](#))
- ³⁹ Belli S, Benedetti M, Comba P, Lagravinese D, Martucci V, Martuzzi M, Morleo D, Trinca S, Viviano G (2004). Case-control study on cancer risk associated to residence in the neighbourhood of a petrochemical plant. *Eur J Epidemiol*. 19(1):49-54. ([Abstract](#))
- ⁴⁰ Agabiti N, Picconi O, Papini P, Schifano P, De Luca A, Cardo S, Gentile D, Scarinci M, Forastiere F, Arcà M, Perucci CA. La valutazione della qualità della compilazione e codifica della scheda di dimissione. Roma: Agency for Public Health; 2002. ([Full text in Italian](#))
- ⁴¹ Alberti V, Flor L. The quality of the data on hospital discharge. *Epidemiol Prev*. 1992 Sep; 14(52):32-8. ([Abstract](#))
- ⁴² Lorenzoni L, Da Cas R, Aparo UL. The quality of abstracting medical information from the medical record: the impact of training programmes. *Int J Qual Health Care* 1999; 11(3):209-13. ([Abstract](#)) ([Full text](#))
- ⁴³ Chio A, Ciccone G, Calvo A, Vercellino M, Di Vito N, Ghiglione P, Mutani R. Piemonte and Valle d'Aosta Register for ALS. Validity of hospital morbidity records for amyotrophic lateral sclerosis. A population-based study. *J Clin Epidemiol* 2002; 55(7):723-7. ([Abstract](#))
- ⁴⁴ Fano V, Forastiere F, Perucci CA. Utilizzo dei dati delle Schede di Dimissione Ospedaliera per le analisi geografiche in epidemiologia ambientale. *Rapporti ISTISAN 05/1*. ([Full text in Italian](#))
- ⁴⁵ Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP. Accuracy of diagnostic coding for medicare patients under the prospective-payment system. *N Engl J Med* 1998; 318: 352-5. ([Abstract](#))