

GIOVANNI BAGLIO, MARIA ELENA TOSTI

Introduzione	53
Elementi di analisi dei dati in uno studio ecologico	55
Punti di forza	57
Limiti	57

• *Fallacia ecologica* • *Confondimento* • *Qualità dei dati*

Introduzione

L'epidemiologia ecologica si colloca in una posizione intermedia rispetto a quella descrittiva e a quella analitica. L'affinità degli studi ecologici con l'epidemiologia descrittiva deriva dal fatto che anche gli studi ecologici analizzano la diffusione delle malattie nelle popolazioni; l'affinità con l'epidemiologia analitica discende invece dal fatto che gli studi ecologici tendono, in qualche modo, a mettere in relazione la diffusione delle malattie con la diffusione dei fattori di rischio. Il collegamento tra malattie e fattori di rischio è realizzato però in un'ottica che mira a generare ipotesi, più che a saggiarle.

La caratteristica principale degli studi ecologici è costituita dal fatto che l'esposizione e l'esito in studio sono determinati a livello di popolazione e non a livello di singoli individui. Questo significa che non si dispone di dati individuali, relativi all'esposizione e/o alla malattia all'interno della popolazione, ma di dati in forma aggregata riferiti all'intero gruppo in studio. Se, ad esempio, si è interessati a studiare l'associazione tra il fumo di sigaretta e il carcinoma dell'esofago, si identifica la proporzione di fumatori (o il volume delle vendite di sigarette) in una certa popolazione

e, in riferimento alla stessa popolazione, si ricerca il dato di incidenza del carcinoma esofageo. Questo procedimento viene ripetuto per diverse popolazioni a confronto: un possibile risultato di tale ricerca potrebbe essere che la malattia si presenti con maggiore frequenza nelle popolazioni in cui la proporzione di fumatori è più alta.

È proprio il confronto tra i gruppi (più che tra gli individui singolarmente considerati) lo scopo degli studi ecologici ed è anche il punto di contatto tra gli studi ecologici e quelli di epidemiologia analitica.

In base alle diverse modalità di confronto, è possibile distinguere due principali tipologie di disegni ecologici. La prima comprende gli studi cosiddetti *multi-group*, in cui vengono confrontati tra loro gruppi o popolazioni con diversi livelli di esposizione, allo scopo di individuare differenze nella distribuzione geografica della malattia (si pensi, ad esempio, a uno studio sull'associazione tra consumo pro capite di carne e cancro del colon in diversi Paesi europei). L'altra tipologia di disegni ecologici comprende gli studi definiti *time-series*, che confrontano la stessa popolazione nel tempo, alla ricerca di "trend" temporali (ad esempio, uno studio sull'andamento dei tassi di mortalità per cancro del colon

in una certa area geografica, in relazione all'aumento nel tempo del consumo pro capite di carne). Una tipologia mista di disegno è rappresentata dagli studi *multiple time-series*, in cui diversi gruppi vengono confrontati tra loro e rispetto a se stessi nel tempo, alla ricerca di modelli di distribuzione della malattia, sia geografici che temporali.

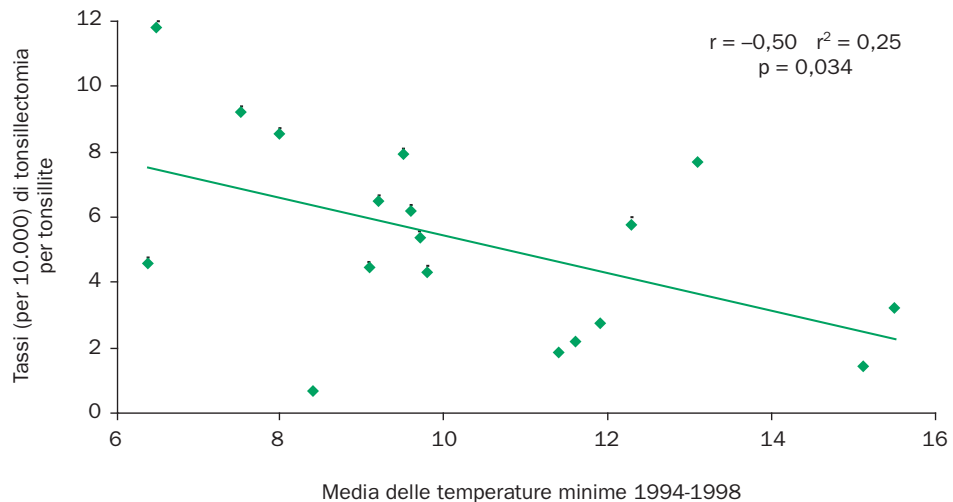
Le informazioni utilizzate possono essere misure medie all'interno del gruppo (incidenza o prevalenza di malattie, proporzione di fumatori, reddito medio, ecc.) e, in questo caso, si parla di *misure aggregate*: esse si caratterizzano per essere informazioni tipicamente individuali che poi vengono aggregate a livello di popolazione. Un altro tipo di dati spesso utilizzati negli studi ecologici è costituito dalle *misure ambientali*, come la temperatura, l'inquinamento atmosferico o delle acque, ecc. La caratteristica di questi dati è che sono propri del contesto in cui i singoli soggetti vivono, ma possono trovare un corrispettivo valore a livello individuale; in alcuni casi possono anche variare all'interno delle aree oggetto di studio.

Un'ultima categoria di informazioni riguarda le cosiddette *misure globali* (o *inte-*

grali) che invece non hanno variabilità a livello individuale, perché sono caratteristiche del gruppo in quanto tale: tra queste si possono citare il tipo di servizio sanitario, il sistema politico, la legislazione vigente, ecc.¹

Per meglio comprendere cosa sia uno studio ecologico, può essere utile ricorrere a un esempio: nella figura 5.1 vengono messi in relazione i tassi di tonsillectomia per tonsillite con le temperature minime ambientali nelle diverse regioni italiane.² Questo è un classico esempio di studio ecologico *multi-group*. Attraverso il Ministero della Salute è possibile ottenere i dati sui ricoveri per tonsillectomia effettuati in Italia in un certo anno; a partire da tali informazioni, è possibile calcolare i tassi di intervento di tonsillectomia per le diverse regioni italiane. Sempre da fonte istituzionale (ISTAT), è possibile conoscere la media delle temperature minime registrate nei diversi capoluoghi di regione in un determinato periodo. Ottenute le informazioni necessarie sull'esposizione (temperature minime) e sull'esito di interesse (tassi di tonsillectomia), si può procedere alla costruzione di un grafico che metta in relazione le due serie di

FIGURA 5.1
Correlazione tra tassi annuali di tonsillectomia per tonsillite e temperature minime (media periodo 1994-1998), per regione. Italia, 2000.²



Dall'esame del grafico risulta evidente che le unità di osservazione sono gruppi (in questo caso, popolazioni definite su base geografica) e non singoli individui. Emerge una correlazione negativa tra gli interventi di tonsillectomia e la temperatura atmosferica: all'aumentare delle temperature minime diminuisce il tasso di tonsillectomia per tonsillite.

dati nelle aree in studio (21 tra regioni e province autonome).

Elementi di analisi dei dati in uno studio ecologico

La figura 5.1 riporta la tipica rappresentazione dei dati di uno studio ecologico: sull'asse delle ascisse vengono posti i dati di esposizione (variabile indipendente) e sull'asse delle ordinate i dati sulla diffusione dell'esito in studio nelle aree considerate (variabile dipendente).

Per mostrare come si può ottenere una misura di associazione in uno studio ecologico, si consideri il caso di una relazione lineare tra esposizione ed esito, proprio come nell'esempio della figura 5.1. Il primo passo è rappresentato dalla stima dell'equazione che descrive, in termini matematici, la retta di regressione:

$$y = \beta_0 + \beta_1 \cdot x$$

dove:

- y è la misura di esito (tasso annuale di ospedalizzazione per tonsillectomia);
- x è la misura di esposizione (media delle temperature minime);
- β_0 è l'intercetta, ossia il punto in cui la retta interseca l'asse delle ordinate (corrisponde al valore di y quando x = 0);
- β_1 è il coefficiente di regressione (esprime la maggiore o minore pendenza della retta).

Una prima misura di associazione, molto utilizzata negli studi ecologici, è costituita dal coefficiente di correlazione, r. In questo contesto, il coefficiente di correlazione è indice del grado di dispersione dei punti intorno alla retta di regressione ed è diretta espressione della "strettezza" dell'associazione tra i livelli di esposizione e di malattia.³ Nell'esempio rappresentato nella figura 5.1 si ha un coefficiente di correlazione di -0,50, indicativo di una buona correlazione negativa tra temperatura e tasso di tonsillectomia per tonsillite.

Nei capitoli precedenti sono state illustrate le misure di associazione tipicamente utilizzate in epidemiologia quando l'esposizione in studio sia dicotomica, cioè quando essa sia espressa come presenza o assenza di una determinata condizione o caratteristica (vedi capitolo 2). Si è visto che le principali misure si ottengono attraverso il rapporto o la differenza tra i rischi (rischio relativo, RR, e rischio attribuibile, RA, rispettivamente). Anche negli studi ecologici è possibile ottenere queste misure di associazione. I dati fittizi della tabella 5.1 possono essere utilizzati per mostrare il procedimento di calcolo: si tratta di uno studio ecologico in cui vengono confrontati i tassi annuali di incidenza della malattia M con la proporzione di esposizione al fattore di rischio F, in cinque diverse popolazioni.

Nella figura 5.2 è riportata la rappresentazione grafica dei suddetti dati. Una volta stimata la formula della retta di regressione che lega i tassi annuali di incidenza (y) alla percentuale di esposti (x), oltre al calcolo della correlazione, è possibile ottenere anche stime dei RR e dei RA. Il procedimento è illustrato in dettaglio nella legenda della figura 5.2; in breve, si ottiene l'incidenza negli esposti ponendo x uguale a "1", analogamente si ottiene l'incidenza nei non esposti ponendo x uguale a "0": calcolando il rapporto, o la differenza, dei tassi di incidenza in esposti e non esposti si ottengono, rispettivamente, il RR e il RA.

L'esempio si basa su una relazione lineare che lega l'incidenza di malattia alla percentuale di esposizione. Procedimenti analoghi a quelli visti sono applicabili anche nel caso in cui la relazione sia non lineare.*

* In questi casi è possibile utilizzare trasformazioni dei dati per ottenere una relazione lineare tra esposizione e tassi di malattia (ad esempio, la trasformazione logaritmica, se la relazione originaria è esponenziale, ecc.), oppure è possibile interpolare direttamente i dati attraverso una relazione non lineare (ad esempio, polinomiale).

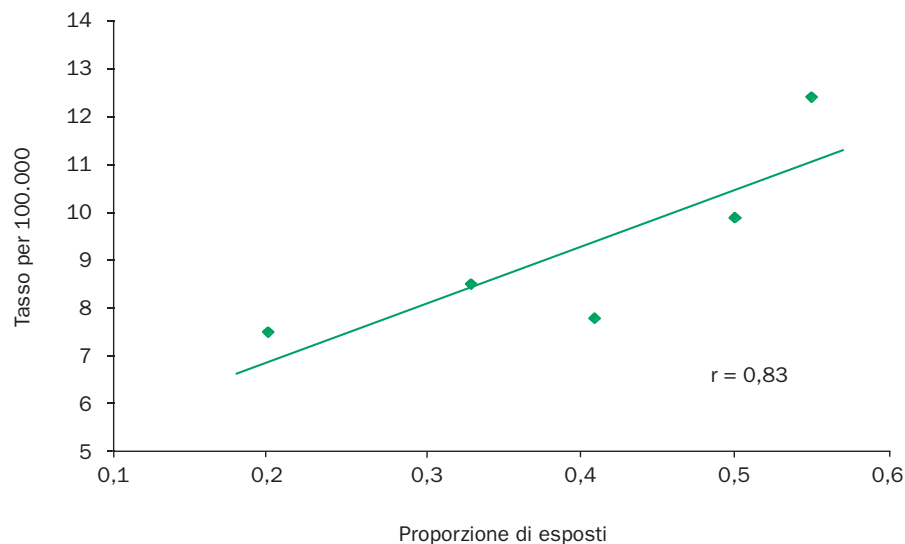
Popolazione	Tasso per 100.000 di M	Proporzione esposti a F
A	7,5	0,20
B	8,5	0,33
C	7,8	0,41
D	9,9	0,50
E	12,4	0,55

TABELLA 5.1

Esempio fittizio di studio ecologico: tassi annuali di incidenza (per 100.000) della malattia M e proporzione di esposti al fattore di rischio F.

FIGURA 5.2

Correlazione tra tassi annuali di incidenza (per 100.000) della malattia M e proporzione di esposti al fattore di rischio F.



Nel grafico sono riportati i dati della tabella 5.1 con la retta di regressione che stima l'andamento dei tassi di incidenza al variare della proporzione di esposizione. Il coefficiente di correlazione, r , fornisce una prima misura di associazione. In questo caso si registra un'alta correlazione positiva tra esposizione e malattia ($r = 0,83$): al crescere della percentuale di esposti a F, cresce l'incidenza di M.

La retta di regressione stimata in questo esempio è la seguente:

$$y = 4,47 + 11,94 \cdot x$$

dove y = incidenza della malattia e x = proporzione di esposti.

Ponendo la proporzione di esposti $x = 0$ (nessun esposto), si ottiene la stima del tasso annuale di incidenza nei non esposti:

$$y_0 = 4,47 + 11,94 \cdot 0 = 4,47 \text{ per } 100.000$$

Ponendo $x = 1$ (tutti esposti), si ottiene la stima del tasso annuale di incidenza negli esposti:

$$y_1 = 4,47 + 11,94 \cdot 1 = 4,47 + 11,94 = 16,41 \text{ per } 100.000$$

Una volta stimati i tassi di incidenza negli esposti e nei non esposti, è facile ottenere il rischio relativo (RR) e il rischio attribuibile (RA):

$$RR = \frac{16,41}{4,47} = 3,67$$

$$RA = 16,41 - 4,47 = 11,94 \text{ per } 100.000$$

L'interpretazione di questi risultati è la seguente: gli esposti hanno un rischio 3,7 volte più elevato rispetto ai non esposti di contrarre la malattia M; inoltre, circa 12 casi di malattia all'anno ogni 100.000 persone esposte al fattore F sono attribuibili all'esposizione.

Punti di forza

Dal momento che in epidemiologia l'interesse è generalmente incentrato sulle relazioni esistenti tra esposizioni ed esiti a livello individuale, ci si potrebbe chiedere come mai vengano condotti degli studi a livello di popolazione e non si ricorra direttamente a osservazioni centrate sugli individui. La risposta è che gli studi ecologici offrono una serie di vantaggi tali da renderli, in alcune situazioni, addirittura preferibili agli studi individuali.

Un primo punto a favore è che in genere gli studi condotti a livello di popolazione utilizzano dati routinari, desumibili da fonti ufficiali e prontamente disponibili nel momento stesso in cui si decida di iniziare lo studio. Questo fa sì che questi studi risultino estremamente economici, rapidi da condurre e particolarmente utili, nelle fasi preliminari di una ricerca, per generare ipotesi di associazione che potranno essere successivamente confermate attraverso disegni di studio individuali.

Un secondo punto di forza è che l'approccio ecologico è il solo che permetta di indagare gli effetti sulla salute esercitati da fattori ambientali e da caratteristiche proprie dei gruppi. Un esempio viene dal campo delle vaccinazioni. È noto che il rischio di contrarre un'infezione per i membri di una comunità non dipende solo dal loro stato individuale di suscettibilità/immunità, ma anche dal livello di immunità collettiva, la cosiddetta *herd immunity* o "immunità di gregge". Al di sopra di un certo livello di *herd immunity*, la malattia non ha infatti possibilità di diffondersi in quanto la popolazione immune agisce da barriera tra i soggetti contagiosi e quelli recettivi. Quindi, se ad esempio si volesse condurre uno studio sull'efficacia di un nuovo vaccino, potrebbe essere preferibile utilizzare l'approccio ecologico, per meglio valutare l'effetto delle dinamiche di popolazione sui rischi individuali di infezione.⁴

Un altro esempio di felice applicazione degli studi ecologici alla comprensione dei fenomeni epidemiologici riguarda le associazioni tra malattie e fattori di rischio che ope-

rano prettamente a livello di popolazione, come la relazione inversa riscontrata tra stato di salute e distribuzione del reddito.^{5,6}

Gli studi ecologici, inoltre, si dimostrano particolarmente utili per valutare esposizioni che variano poco all'interno delle popolazioni, ma variano molto tra popolazioni. Ad esempio, se vi fosse l'interesse a indagare gli effetti della carenza di iodio sul rischio di ipotiroidismo cronico in un'area pedemontana (dove la quasi totalità dei soggetti è esposta a un basso apporto di iodio), uno studio condotto a livello individuale potrebbe non essere in grado di evidenziare alcuna associazione, laddove uno studio comparativo tra aree geografiche potrebbe far emergere importanti differenze tra le popolazioni e, conseguentemente, il ruolo svolto dalla carenza di iodio nell'insorgenza di ipotiroidismo.

Un ultimo punto a favore degli studi ecologici è che sono particolarmente adatti per evidenziare andamenti geografici e/o temporali nella distribuzione delle malattie.

Limiti

Nonostante i loro pregi, gli studi ecologici pongono alcuni problemi di ordine metodologico, che limitano la possibilità di fare, sulla base di essi, inferenze causali. I principali problemi che si incontrano durante la conduzione e l'interpretazione dei risultati di uno studio ecologico sono di seguito riportati.

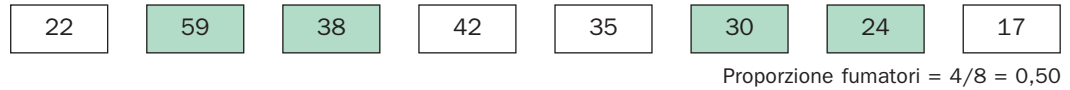
FALLACIA ECOLOGICA

La cosiddetta "fallacia ecologica" è il principale limite degli studi ecologici: fa riferimento al fatto che le associazioni, saggiate a livello aggregato, possono non essere vere anche a livello individuale. Non essendo, peraltro, disponibili informazioni dettagliate sulle popolazioni in studio, non è possibile sapere se il fattore di rischio e la malattia siano compresenti nello stesso individuo. Nella figura 5.3 è riportato un esempio di fallacia ecologica, applicato a uno studio fittizio sull'associazione tra età e abitudine al fumo.

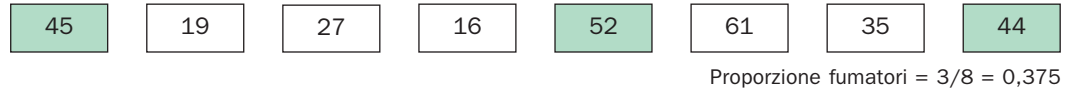
FIGURA 5.3

Studio ecologico sull'associazione tra età (esposizione) e fumo di sigaretta (esito).

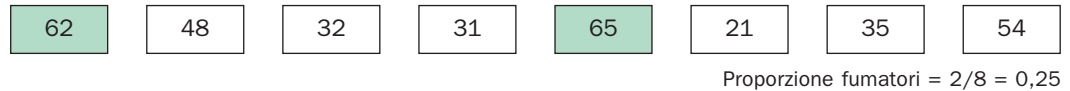
Popolazione A - Età media = 33,4 anni



Popolazione B - Età media = 37,4 anni



Popolazione C - Età media = 43,5 anni



In questo esempio fittizio sono riportate tre popolazioni (A, B e C), ciascuna composta da 8 individui, dei quali è indicata l'età nei riquadri che li rappresentano. I riquadri verdi si riferiscono ai soggetti fumatori. Le tre popolazioni hanno un'età media crescente, mentre la percentuale dei fumatori nelle tre popolazioni è decrescente (50% nella popolazione A e 25% nella popolazione C che è la più anziana). Analizzando i dati in forma aggregata si concluderebbe a favore di una relazione inversa tra fumo di sigarette ed età: la popolazione più giovane tende a fumare di più. Per contro, l'analisi dei dati a livello individuale (in questo caso disponibili) porta a conclusioni opposte: l'età media dei fumatori è 46,6 anni mentre quella dei non fumatori è 33,0; dunque, la relazione tra fumo ed età non è inversa, ma diretta.

Ovviamente, si potrebbero fare molti esempi di studi ecologici nei quali l'associazione evidenziata a livello aggregato è riscontrabile anche a livello individuale. Il problema principale è che, non disponendo di dati per ogni singolo soggetto, non è possibile verificare se il risultato sia valido tanto per le popolazioni quanto per gli individui.

Peraltro, si può obiettare che esiste un analogo ma opposto rischio di "fallacia atomistica". Ciò si realizza quando si applicano alle popolazioni i risultati di studi condotti sugli individui (ad esempio, si muore di più a causa del cancro nei Paesi ricchi, ma gli individui ricchi muoiono meno di cancro rispetto ai poveri). Entrambi questi errori sono in realtà evitabili, tenendo ben presente il livello di organizzazione degli studi.

CONFONDIMENTO

Operando a livello ecologico è certamente più difficile che non a livello individuale tenere sotto controllo il confondimento, ossia

evitare di evidenziare relazioni tra esposizione ed esito dovute in realtà alla presenza di altre caratteristiche associate, a livello ecologico, tanto all'esposizione quanto all'esito (vedi capitolo 11).

Vi sono diversi motivi per i quali non sempre si riesce efficacemente a eliminare il confondimento:

- i gruppi a confronto, appartenendo spesso a popolazioni diverse, possono differire tra loro in molti aspetti, difficilmente controllabili;
- non sempre sono disponibili dati routinari sui potenziali confondenti;
- molti fattori sono più strettamente correlati tra loro a livello di popolazione che a livello individuale; è anche possibile che un fattore non sia un confondente a livello individuale, ma lo sia a livello ecologico.

QUALITÀ DEI DATI

Gli studi ecologici sono soggetti a bias di informazione (errori sistematici nel processo

di acquisizione delle informazioni, vedi capitolo 11) in misura maggiore rispetto agli studi individuali; in questi ultimi, infatti, le informazioni vengono in massima parte rilevate proprio ai fini dello studio e quindi presentano una maggiore accuratezza rispetto ai dati di fonte routinaria, generalmente raccolti per scopi diversi dalle indagini epidemiologiche.

Inoltre, poiché i dati utilizzati negli studi ecologici fanno riferimento a popolazioni diverse o anche a tempi diversi, non sempre i metodi di rilevazione e i criteri di classificazione sono gli stessi nei gruppi a confronto; ciò può causare errori che danno poi luogo a risultati distorti in termini di associazione tra esposizione e malattia. Un esempio è costituito dalle classificazioni di alcune malattie che cambiano rapidamente nel tempo e possono anche essere diverse da luogo a luogo: facendo confronti basati su popolazioni analizzate in tempi differenti, si potrebbero riscontrare variazioni nei tassi di incidenza dovute più che altro a modifiche nei sistemi di classificazione.

Un altro problema nei confronti temporali può essere legato all'utilizzo dei dati di mortalità al posto dei dati di morbosità: nonostante la vera misura di rischio di una malattia sia costituita dal suo tasso di incidenza, molto

spesso questa misura non è disponibile e, per la conduzione di uno studio ecologico, si tende a utilizzare la mortalità. Uno studio del tipo *time-series* che utilizzi i tassi di mortalità come misura dell'esito potrebbe far emergere una tendenza alla diminuzione che non corrisponde a una reale riduzione della morbosità, ma al miglioramento dei sistemi di cura.

Bibliografia

1. Rothman KJ, Greenland S. Modern epidemiology. Second Edition. Philadelphia: Lippincott-Raven, 1998.
2. Matera E, Di Domenicantonio R, Baglio G, et al. Epidemiologia degli interventi di tonsillectomia e/o adenoidectomia in Italia. *Pediatr Med Chir* 2004; 26 (3): 179-86.
3. Blalock HM jr. Statistica per la ricerca sociale. Bologna: Il Mulino, 1969.
4. Smith S, Sinclair D, Raine R, Reeves B. Health care evaluation. Maidenhead: Open University Press, 2005.
5. Wilkinson RG. Unhealthy societies. The afflictions of inequalities. London: Routledge, 1996.
6. Mackenbach JP. Income inequalities and population health. *BMJ* 2002; 324: 1-2.