

I metodi di analisi statistica per la valutazione degli effetti a breve termine dell'inquinamento atmosferico nel progetto EpiAir

Massimo Stafoggia, per il gruppo collaborativo EpiAir

Obiettivi del progetto EpiAir

Il progetto EpiAir si propone di:

1. avviare un sistema di sorveglianza delle caratteristiche ambientali con rilevanza sanitaria nelle città italiane;
2. raccogliere in modo sistematico i dati relativi ai fenomeni sanitari rilevanti, quali mortalità e ricoveri ospedalieri, per cui è noto l'impatto dell'inquinamento atmosferico;
3. produrre rapporti con le stime degli effetti sanitari attribuibili all'inquinamento atmosferico;
4. individuare i gruppi di popolazione particolarmente vulnerabili ai danni dell'inquinamento atmosferico.

Popolazioni in studio e dati disponibili

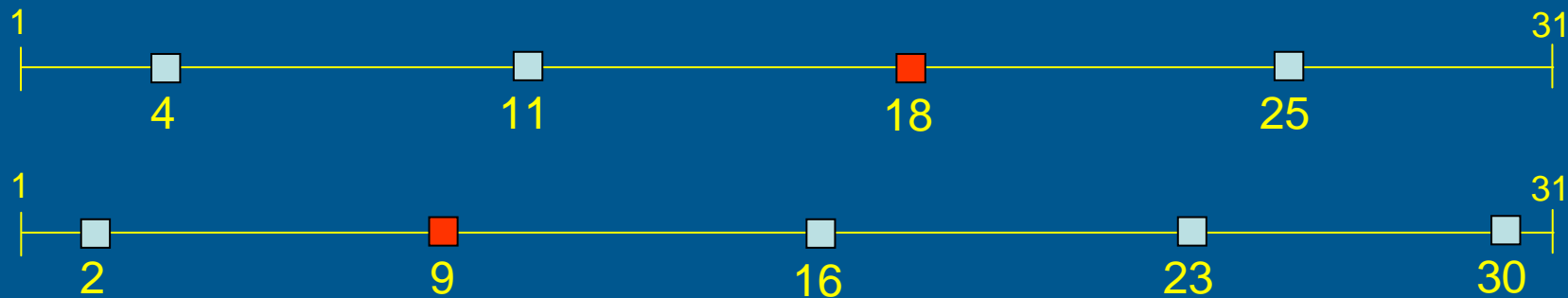
- 10 città coinvolte;
- Periodo in studio 2001-2005;
- Principali variabili ambientali: PM10, NO₂, O₃, temperatura, umidità, pressione;
- Principali esiti sanitari: mortalità naturale, mortalità per gruppi di cause, ricoveri per specifiche patologie;
- Caratteristiche individuali: età, genere, luogo del decesso, condizioni croniche, condizioni acute

Obiettivi della presentazione

- Disegno di studio
- Controllo città-specifico del **confondimento**
- Analisi città-specifica dell'**associazione** tra inquinanti ed esiti sanitari
- Analisi città-specifica della **modificazione d'effetto** per caratteristiche individuali
- Produzione di stime congiunte: **meta-analisi**
- **Interpretazione** dei risultati

Disegno di studio (1)

- Disegno case-crossover;
- Disegno caso-controllo in cui ogni soggetto deceduto/ricoverato è “matchato” con se stesso;
- Il “caso” è rappresentato dal giorno di decesso/ricovero, i “controlli” da altri giorni opportunamente scelti
- Approccio “time-stratified”: stessi giorni della settimana all’interno dello stesso mese ed anno del caso. Esempi:



Disegno di studio (2)

- Si confronta l'esposizione nel giorno del caso con l'esposizione media nei giorni dei controlli
- Si aggiusta per i fattori tempo-dipendenti che variano tra casi e controlli (es. meteorologia, festività, ecc.)
- Non si aggiusta per i fattori individuali che non variano nel tempo, o variano lentamente (es. genere, fumo, ecc.) in quanto casi e controlli sono identici rispetto ad essi
- Tali fattori possono invece **modificare la relazione** tra esposizione ed esito

Controllo città-specifico del confondimento (1)

- **Confondimento:** distorsione della stima di associazione tra esposizione ed esito per la mancata correzione di fattori (confondenti) associati ad entrambi
- **Nelle serie temporali:** i principali confondenti sono variabili tempo-dipendenti che variano con l'esposizione (inquinamento) e l'esito (mortalità/ricovero)
- **Esempi principali:** trend di medio-lungo periodo, meteorologia, festività, giorno della settimana, ecc.
- **Il case-crossover (con approccio "time-stratified")** aggiusta per trend e giorno della settimana *by design*. Gli altri confondenti devono essere aggiustati in fase di analisi

Controllo città-specifico del confondimento (2)

- Decremento estivo di popolazione:

- 2 Due settimane intorno a Ferragosto
- 1 Da metà luglio a fine agosto (eccetto il precedente periodo)
- 0 Tutti gli altri giorni

- Giorni di festa e periodi di festività:

- 1 Festività nazionali e patronali
 - 0 Tutti gli altri giorni
- Mortalità

- 3 Giorni di Natale e Pasqua
 - 2 Giorni intorno a Natale e Pasqua
 - 1 Altre festività nazionali e patronali
 - 0 Tutti gli altri giorni
- Ricoveri ospedalieri

Controllo città-specifico del confondimento (3)

- Epidemie influenzali: basate sul sistema di sorveglianza della sindrome influenzale INFLUNET, promosso dal Ministero della Salute

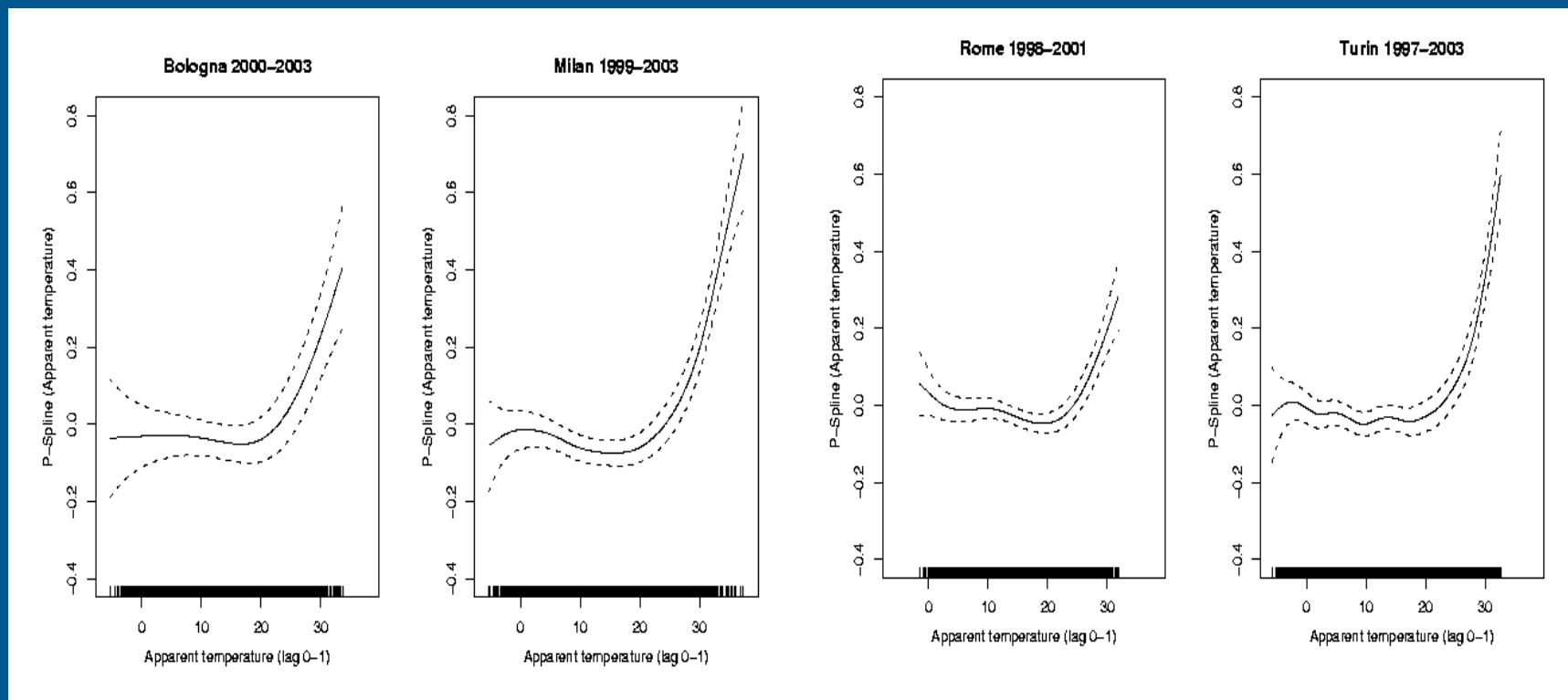
- 1 Giorni di epidemie influenzali
- 0 Tutti gli altri giorni

Città	Giorni di epidemia influenzale				
	2001	2002	2003	2004	2005
Milano	5/02-25/02	21/01-10/02	17/02-9/03	2/02-22/02	31/01-20/02
Mestre-Venezia	5/02-25/02	28/01-17/02	17/02-9/03	19/01-8/02	31/01-20/02
Torino	29/01-18/02	21/01-10/02	17/02-9/03	26/01-8/02 16/02-22/02	31/01-20/02
Bologna	5/02-25/02	21/01-10/02	17/02-9/03	26/01-8/02 16/02-22/02	31/01-20/02
Firenze	5/02-25/02	21/01-10/02	17/02-9/03	26/01-15/02	31/01-20/02
Pisa	5/02-25/02	21/01-10/02	17/02-9/03	26/01-15/02	31/01-20/02
Roma	5/02-25/02	21/01-10/02	10/02-2/03	9/02-29/02	24/01-30/01 7/02-20/02
Taranto	5/02-25/02	21/01-10/02	17/02-9/03	8/03-28/03	24/01-13/02
Cagliari		24/02-16/03	24/02-16/03	19/01-25/01 2/02-8/02 16/02-22/02	7/02-27/02
Palermo		21/01-10/02	17/02-9/03	2/02-8/02 16/02-29/02	7/02-27/02

- Pressione barometrica: misurata in HectoPascal

Controllo città-specifico del confondimento (4)

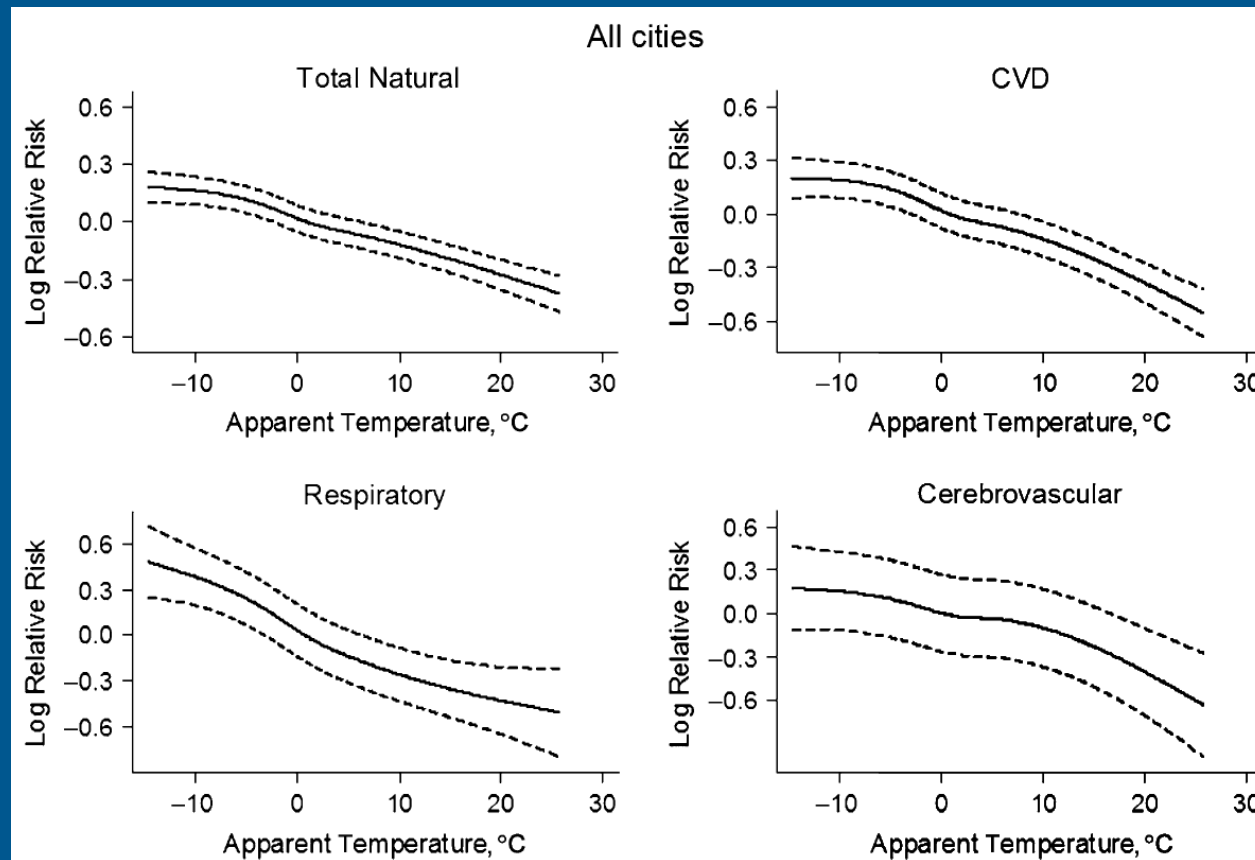
- **Temperature elevate:** evidenze di una relazione forte e non lineare tra temperatura apparente elevata ed esiti sanitari, a lag immediati (0-1 giorno) (es. mortalità SISTI)



Stafoggia et al. *Epidemiology* 2006; 17:315-323

Controllo città-specifico del confondimento (5)

- **Temperature basse:** evidenze di una relazione consistente tra temperature basse ed esiti sanitari, a lag cumulati (fino a 15 giorni) (es. mortalità PHEWE)



Analitis et al. *Am J Epidemiol* 2008;168:1397–1408

Controllo città-specifico del confondimento (6)

Modalità di aggiustamento dei confondenti:

- **Decremento estivo di popolazione, festività, epidemie influenzali:** variabili categoriche
- **Pressione barometrica:** spline penalizzata
- **Elevate temperature:** spline penalizzata della temperatura apparente (lag 0-1) nei giorni in cui essa è superiore al suo valore mediano città-specifico
- **Basse temperature:** spline penalizzata della temperatura (lag 1-6) nei giorni in cui essa è inferiore al suo valore mediano città-specifico

Analisi città-specifica dell'associazione tra inquinanti ed esiti sanitari (1)

- Una volta identificato il set di confondenti e la modalità di aggiustamento, si inserisce il **termine di esposizione** (concentrazione di inquinante)
- A seconda degli obiettivi dell'analisi, **si definiscono diversi modelli** che differiscono per come l'inquinante viene inserito nel modello:
 - **Modelli a lag singolo**
 - **Modelli a lag distribuiti vincolati**
 - **Modelli a lag distribuiti non vincolati**

Analisi città-specifica dell'associazione tra inquinanti ed esiti sanitari (2)

Modelli a lag singolo:

- **Obiettivo:** stimare l'associazione tra la concentrazione di inquinante di uno specifico giorno (definito dal lag) ed il rischio di esito
- Si realizzano aggiungendo al modello di aggiustamento la concentrazione di inquinante relativa al giorno di interesse
- Il coefficiente stimato rappresenta il $\log(\text{RR})$ di esito associato ad una variazione di $1 \mu\text{g}/\text{m}^3$ di incremento dell'inquinante, verificatasi nel giorno di interesse

Esito = confondenti + $I_{\text{lag}0}$ → Concentrazione di inquinante di oggi

Esito = confondenti + $I_{\text{lag}1}$ → Concentrazione di inquinante di ieri

Esito = confondenti + $I_{\text{lag}2}$ → Concentrazione di inquinante di due giorni fa

Analisi città-specifica dell'associazione tra inquinanti ed esiti sanitari (3)

Modelli a lag distribuiti vincolati:

- **Obiettivo:** fornire una rappresentazione grafica della associazione tra inquinante ed esito, con riferimento a più giorni consecutivi di esposizione
- Si realizzano aggiungendo al modello di aggiustamento le concentrazioni di inquinante relative ai giorni di interesse
- I coefficienti stimati vengono vincolati a seguire un andamento prestabilito, ad esempio polinomiale
- L'assunzione è che l'effetto dell'inquinamento si manifesti su più giorni consecutivi, e che l'andamento di tale effetto sia approssimabile con una relazione matematica

Analisi città-specifica dell'associazione tra inquinanti ed esiti sanitari (4)

Modelli a lag distribuiti vincolati:

- 1° passo : $I_{lag0} + I_{lag1} + I_{lag2} + I_{lag3} + I_{lag4} + I_{lag5} + I_{lag6}$
- 2° passo : I coefficienti $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ vengono vincolati a seguire l'andamento polinomiale descritto dalla formula:

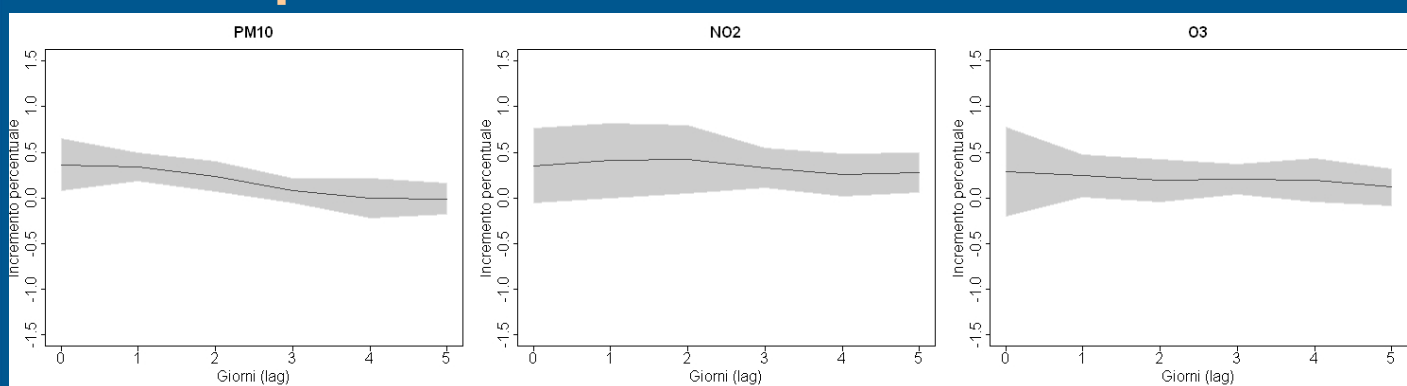
$$\beta_k = \eta_0 k^0 + \eta_1 k^1 + \eta_2 k^2 + \eta_3 k^3$$

- 3° passo: i nuovi coefficienti $\eta_0, \eta_1, \eta_2, \eta_3$ vengono stimati dal modello
- 4° passo : i coefficienti finali $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ vengono ottenuti dalle stime di $\eta_0, \eta_1, \eta_2, \eta_3$ per trasformazione inversa
- 5° passo : tali coefficienti vengono rappresentati nel grafico finale

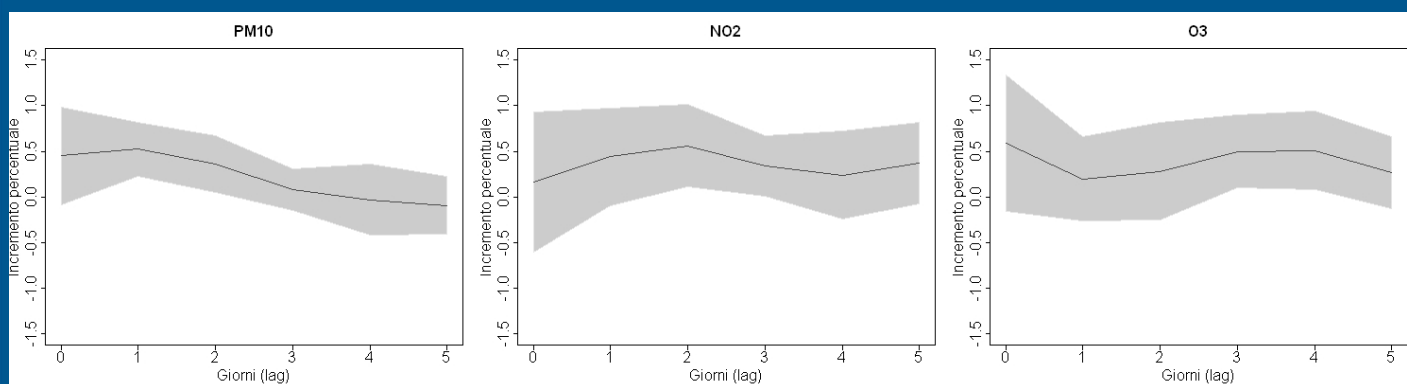
Analisi città-specifica dell'associazione tra inquinanti ed esiti sanitari (5)

Modelli a lag distribuiti vincolati: ESEMPI

Mortalità per cause naturali



Ricoveri per cause cardiache



Analisi città-specifica dell'associazione tra inquinanti ed esiti sanitari (6)

Modelli a lag distribuiti non vincolati:

- **Obiettivo:** fornire una stima cumulativa della associazione tra inquinante ed esito, con riferimento a più giorni consecutivi di esposizione
- Si realizzano aggiungendo al modello di aggiustamento le concentrazioni di inquinante relative ai giorni di interesse
- I coefficienti stimati non vengono vincolati a seguire alcun andamento prestabilito
- Sebbene i singoli coefficienti siano molto instabili per via della elevata correlazione tra le concentrazioni di inquinanti di giorni contigui, la loro somma fornisce una stima non distorta dell'effetto cumulato dell'inquinante sull'esito

Analisi città-specifica dell'associazione tra inquinanti ed esiti sanitari (7)

Modelli a lag distribuiti non vincolati:

• 1° passo : $\text{Esito} = \text{confondenti} + I_{\text{lag}0} + I_{\text{lag}1} + I_{\text{lag}2} + I_{\text{lag}3} + I_{\text{lag}4} + I_{\text{lag}5}$

• 2° passo : I coefficienti $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ vengono sommati:

$$\beta_{05} = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5$$

• Il coefficiente β_{05} rappresenta il $\log(\text{RR})$ di esito associato ad un incremento di $1 \mu\text{g}/\text{m}^3$ dell'inquinante verificatosi in ognuno dei giorni considerati (dal giorno presente a 5 giorni prima, nell'esempio)

• Tale coefficiente viene trasformato per ottenere l'incremento percentuale di rischio associato ad incrementi di $10\mu\text{g}/\text{m}^3$ dell'inquinante (per ognuno dei giorni considerati):

$$\%_{05} = (\exp\{\beta_{05} * 10\} - 1) * 100$$

Analisi città-specifica dell'associazione tra inquinanti ed esiti sanitari (8)

Modelli a lag distribuiti non vincolati: ESEMPIO

Incrementi percentuali di rischio (%) e intervalli di confidenza al 95% (L, U) per variazioni di 10 unità dell'inquinante

	Mortalità naturale			
	lag	%	L	U
PM10				
	0-1	0.69	0.40	0.98
	2-5	0.49	0.06	0.93
	0-5	0.93	0.40	1.46
NO2				
	0-1	0.99	0.24	1.74
	2-5	1.87	0.96	2.79
	0-5	2.09	0.96	3.24
O3 (aprile-settembre)				
	0-1	0.70	0.20	1.21
	2-5	0.97	0.50	1.45
	0-5	1.54	0.92	2.15

Analisi città-specifica della modificazione d'effetto per caratteristiche individuali (1)

Potenziali fattori di suscettibilità considerati:

- Età, genere
- Luogo del decesso:
 - fuori dall'ospedale, senza recenti ricoveri
 - fuori dall'ospedale, con recente ricovero
 - in ospedale
 - in residenza sanitaria assistita
- Specifiche condizioni croniche (sulla base dei ricoveri degli ultimi 2 anni, escluse le ultime 4 settimane)
- Specifiche condizioni acute (sulla base dei ricoveri delle ultime 4 settimane)

Analisi città-specifica della modificazione d'effetto per caratteristiche individuali (2)

Patologie	ICD9-CM
<i>croniche</i>	
diabete mellito	250
disturbi della coagulazione	286-287
ipertensione arteriosa	401-405
infarto del miocardio	410, 412
malattie ischemiche del miocardio	410-414
malattie del circolo polmonare	415-417
turbe della conduzione	426
aritmie	427
scompenso cardiaco	428
disturbi circolatori dell'encefalo	430-438
malattie polmonari croniche	490-505
<i>acute</i>	
malattie del circolo polmonare	415-417
turbe della conduzione	426
aritmie	427
scompenso cardiaco	428
infezioni respiratorie acute	460-466, 480-486
insufficienza renale	584-588

Analisi città-specifica della modificazione d'effetto per caratteristiche individuali (3)

- L'associazione tra inquinante e mortalità naturale è stata stimata in ogni categoria dei presunti modificatori (es. maschi e femmine per il modificatore “genere”, assenza o presenza di una patologia, ecc.)
- Le relative stime di associazione sono state confrontate in termini statistici, ed è stato prodotto un p-REM (p-value di modificazione relativa d'effetto), allo scopo di identificare differenze significative
- Standardizzazione per età: poiché l'età è fortemente associata a tutti i fattori considerati, le analisi di suscettibilità sono state standardizzate per età, secondo la seguente formula

$$\beta_{STD}^M = \frac{\omega_{35-64} \beta_{35-64}^M + \omega_{65-74} \beta_{65-74}^M + \omega_{75-84} \beta_{75-84}^M + \omega_{85+} \beta_{85+}^M}{\omega_{35-64} + \omega_{65-74} + \omega_{75-84} + \omega_{85+}}$$

Analisi città-specifica della modificazione d'effetto per caratteristiche individuali (4)

ESEMPIO:

Incrementi percentuali di rischio (%) e intervalli di confidenza al 95% (L, U) per variazioni di 10 unità del PM10

Variabili	%	L	U	p-rem
Popolazione totale (età 35+ anni) deceduta per cause naturali	0.80	0.41	1.19	-
Età (anni)				
35-64	0.02	-0.87	0.91	-
65-74	0.44	-0.32	1.21	0.481
75-84	0.69	0.04	1.35	0.234
85+	1.47	0.92	2.02	0.007
Genere*				
Maschi	0.84	0.36	1.32	-
Femmine	0.74	0.07	1.42	0.813
Luogo del decesso*				
Fuori dall'ospedale	0.56	0.03	1.10	-
Fuori dall'ospedale, con dimissione da 2 a 28 giorni prima del decesso	0.59	-0.51	1.70	0.970
In ospedale ^a	0.95	0.47	1.42	0.297
Residenza sanitaria assistenziale (RSA) ^b	-0.11	-1.83	1.63	0.465

Produzione di stime congiunte: meta-analisi (1)

Meta-analisi ad effetti fissi:

- La stima congiunta è una **media pesata** delle stime città-specifiche, con pesi **proporzionali alla precisione delle stime**
- Si assume che le **stime città-specifiche siano omogenee** e differiscano solo per effetto del caso
- **Le città più grandi**, che generalmente forniscono stime più precise, **pesano maggiormente nella media pesata** rispetto alle città piccole
- La formula che definisce la stima meta-analitica ad effetti fissi è:

$$\hat{\beta}_{FE}^{ML} = \frac{\sum_{j=1}^{10} \omega_j \hat{\beta}_j}{\sum_{j=1}^{10} \omega_j}$$

con pesi pari a

$$\omega_j = \frac{1}{\hat{\sigma}_j^2}$$

Produzione di stime congiunte: meta-analisi (2)

Meta-analisi ad effetti casuali:

- La stima congiunta è una **media pesata** delle stime città-specifiche, con pesi proporzionali alla precisione delle stime ed alla variabilità delle stime tra le città
- Si assume che le **stime città-specifiche siano eterogenee** e differiscano non solo per effetto del caso, ma **anche per altri fattori inerenti le città stesse** (es. latitudine, geografia, composizione della popolazione, ecc.)
- Quanto più elevata è la variabilità tra le città, tanto più esse pesano in modo simile nella media pesata
- La formula che definisce la stima meta-analitica ad effetti casuali è analoga alla precedente, ma con diversa formulazione dei pesi:

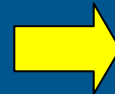
$$\omega_j = \frac{1}{\hat{\sigma}_j^2 + \hat{\tau}^2}$$

Produzione di stime congiunte: meta-analisi (3)

Meta-analisi ad effetti casuali:

- Il termine τ^2 è una stima dell'eterogeneità tra città e si basa sulla statistica test Q di omogeneità:

$$Q = \sum_{j=1}^{10} \omega_j (\hat{\beta}_j - \hat{\beta}_{FE}^{ML})^2 \sim \chi^2_{10-1}$$



$$\hat{\tau}^2 = \max \left\{ 0, \frac{Q - (10 - 1)}{\sum_{j=1}^{10} \omega_j - \frac{\sum_j \omega_j^2}{\sum_j \omega_j}} \right\}$$

- Più elevato il valore, più le stime città-specifiche sono diverse e la stima congiunta a effetti casuali diverge dalla stima a effetti fissi
- Più basso il valore, più le stime di associazione tra inquinante ed esito delle diverse città sono omogenee e la stima a effetti casuali si avvicina a quella a effetti fissi

Produzione di stime congiunte: meta-analisi (4)

Interpretazione dell'eterogeneità:

- Sulla base della distribuzione X^2 di Q, è stato prodotto il relativo p-value del test di omogeneità (p-HET)
 - Valori di p-HET inferiori a 0.05 supportano l'ipotesi di eterogeneità delle stime città-specifiche: esse differiscono non solo per il caso, ma anche per altri fattori
 - Valori di p-HET superiori a 0.05 supportano l'ipotesi di omogeneità delle stime città-specifiche: esse differiscono solo per effetto del caso
- In caso di eterogeneità: la stima congiunta non ha altro significato se non quello di effetto medio di realtà diverse tra di loro.
- In caso di omogeneità: la stima meta-analitica rappresenta davvero una sintesi degli effetti città-specifici, in quanto essi provengono da un'unica popolazione

Interpretazione dei risultati

- Tutti i risultati di associazione tra inquinanti ed esiti sono espressi come incrementi percentuali di rischio, corrispondenti ad incrementi di $10 \mu\text{g}/\text{m}^3$ dell'inquinante
- Tuttavia, i diversi inquinanti presentano valori medi e distribuzioni diverse nelle varie città, cosicché la scala comune di $10 \mu\text{g}/\text{m}^3$ può corrispondere a livelli di esposizione diversi tra gli inquinanti
- In particolare, mentre PM10 e NO2 mostrano livelli medi di concentrazioni simili, l'ozono ha livelli più elevati e distribuzioni più disperse
- Ne consegue che la popolazione delle 10 città italiane è mediamente esposta alla stessa variabilità nelle concentrazioni di PM10 e NO2, mentre è esposta a una variabilità più elevata per quanto riguarda l'ozono
- Nell'interpretare i risultati dello studio occorre tenere conto di questo aspetto