

Metanalisi o analisi pooled? Un confronto basato su serie storiche per l'analisi degli effetti a breve termine dell'inquinamento atmosferico sulla salute umana

Meta-analysis or pooled analysis? A comparison based on time-series used for the analysis of short-term effects of air pollution on human health

Bianca Gherardi*,^{1,2} Giulia Tommaso*,^{1,2} Andrea Ranzi,¹ Stefano Zauli Sajani,¹ Aldo De Togni,³ Lorenzo Pizzi,⁴ Paolo Lauriola¹

¹ Centro tematico ambiente e salute, Direzione tecnica, Agenzia regionale prevenzione e ambiente Emilia-Romagna, Modena

² Dipartimenti di scienze statistiche, Università di Bologna

³ Dipartimento di sanità pubblica, AUSL Ferrara

⁴ UOC epidemiologia, promozione della salute e comunicazione del rischio, UOS rischio ambientale, Dipartimento di sanità pubblica, Azienda USL di Bologna

*Hanno contribuito in ugual misura alla ricerca

Corrispondenza
Bianca Gherardi
bianca.gherardi2@unibo.it

Giulia Tommaso
giulia.tommaso2@unibo.it

Epidemiol Prev 2015; 39 (2): 98-105

OBIETTIVI: confrontare l'approccio di metanalisi e analisi *pooled* nello studio degli effetti a breve termine dell'inquinamento atmosferico nelle città dell'Emilia-Romagna, caratterizzate da forte omogeneità delle caratteristiche ambientali e sociodemografiche.

METODI: applicazione di metanalisi a effetti fissi e analisi *pooled* a effetti fissi sui dati di serie storiche relativi a sette capoluoghi dell'Emilia-Romagna nel periodo 2006-2010. La relazione tra eventi sanitari avversi (decessi per cause naturali, per malattie cardiovascolari, cerebrovascolari e respiratorie) e concentrazioni di PM₁₀, PM_{2.5} e NO₂ è stata analizzata mediante l'uso di modelli GAM utilizzando il protocollo EpiAir.

RISULTATI: l'applicazione dell'analisi *pooled* ha portato a un aumento di precisione delle stime d'effetto rispetto all'approccio metanalitico. Le ampiezze degli intervalli di confidenza relative all'analisi *pooled* sono risultate inferiori rispetto a quelle metanalitiche, ottenendo riduzioni percentuali comprese tra il 7% e il 43%.

Questo aumento di potenza ha portato a una frequenza maggiore di stime statisticamente significative nell'analisi *pooled*. Si è riscontrata in generale una buona corrispondenza tra i due metodi per direzione e forza dell'associazione tra gli *outcome* sanitari e i vari inquinanti. Fa eccezione la stima dell'effetto del PM₁₀ sulla mortalità respiratoria, dove la stima metanalitica è risultata notevolmente più elevata e poco in linea con i dati di letteratura.

CONCLUSIONI: lo studio ha permesso di mettere in luce l'aumento di precisione e di stabilità delle stime d'effetto ottenuto dall'applicazione di un'analisi *pooled* rispetto a una metanalisi in una realtà regionale come quella dell'Emilia-Romagna, caratterizzata da assenza di eterogeneità nell'esposizione agli inquinanti e agli altri confondenti. In tali contesti, l'approccio *pooled* è, quindi, da ritenere preferibile rispetto a quello metanalitico.

Parole chiave: metanalisi, analisi *pooled*, effetti fissi, confronto, precisione

Cosa si sapeva già

- Metanalisi e analisi *pooled* sono due metodi di sintesi quantitativa per riassumere evidenze derivanti da studi volti a indagare la medesima relazione.
- Pochi studi hanno messo a confronto metanalisi e analisi *pooled*.

Cosa si aggiunge di nuovo

- E' il primo confronto applicato alle esposizioni ambientali partendo dagli stessi conteggi giornalieri.
- In condizioni di bassa numerosità di centri e omogeneità delle condizioni meteo-ambientali, quali quelle di alcune realtà regionali, l'approccio *pooled* permette di ottenere stime più stabili.

ABSTRACT

Meta-analysis or pooled analysis? A comparison based on time-series used for the analysis of short-term effects of air pollution on human health

Epidemiol Prev 2015; 39(2): 98-105

OBJECTIVES: to compare the meta-analysis and the pooled analysis approach to study short-term effects of air pollution on human health in Emilia-Romagna Region (Central Italy) cities, characterised by strong homogeneity of environmental and sociodemographic features.

METHODS: application of fixed-effects meta-analysis and fixed-effects pooled analysis on time-series data of seven cities in Emilia-Romagna in the period 2006-2010. The relationship among adverse health events (deaths due to natural causes, cardiovascular disease, cerebrovascular disease and respiratory disease) and concentrations of PM_{10} , $PM_{2.5}$

and NO_2 was investigated by means of GAM models, using the EpiAir protocol.

RESULTS: the pooled analysis application entailed a gain in terms of precision of effect estimates in respect to meta-analysis approach. The interval widths of pooled analysis are lower than those of meta-analytic estimates, with percentage reductions between 7% and 43%. This power increase led to a major number of statistically significant pooled analysis estimates. It has been a generally good correspondence between the two methods in terms of direction and strength of the association among health outcomes and the various pollutants. An exception is the PM_{10} effect estimate on res-

piratory mortality, where the meta-analytic estimate was significantly higher and not in line with literature data.

CONCLUSIONS: the study highlighted the increase in accuracy and stability of effect estimates obtained from a pooled analysis compared to a meta-analysis in a regional context such as the Emilia-Romagna Region, characterised by the absence of heterogeneity in exposure to pollutants and other confounders. In this context, the pooled approach is to be considered preferable to meta-analysis.

Keywords: meta-analysis, pooled analysis, fixed effects, comparison, precision

INTRODUZIONE

Metanalisi e analisi *pooled* sono strumenti per sintetizzare evidenze derivanti da diversi sottogruppi di popolazione e rappresentano una parte importante della ricerca epidemiologica, anche a causa della crescente necessità di valutare rischi che potrebbero risultare non statisticamente significativi in studi di piccole dimensioni, ma avere un impatto sull'intera popolazione con implicazioni importanti per la salute pubblica.¹ Esistono diversi metodi per riassumere le evidenze; in questo articolo ci si focalizzerà su due in particolare: la metanalisi e l'analisi *pooled*.

Nell'ambito dell'epidemiologia ambientale, in Italia sono state pubblicate numerose metanalisi sugli effetti a breve termine dell'inquinamento atmosferico sulla salute, denominate MISA² e MISA2,³ EpiAir⁴ e EpiAir2.⁵

Gli *outcome* analizzati sono mortalità giornaliera (per cause non accidentali, cardiovascolari e respiratorie) e ricoveri ospedalieri per problemi cardiorespiratori. I dati sono stati rilevati giornalmente nelle principali città italiane per periodi di studio di quattro anni. Per ottenere una misura d'effetto globale è stata effettuata una metanalisi a effetti casuali, in quanto le città non sono risultate omogenee.

In letteratura esistono pochi studi che hanno confrontato metanalisi e analisi *pooled*, la maggior parte dei quali soltanto dal punto di vista teorico, illustrandone vantaggi e svantaggi.^{1,6-8} L'unico lavoro che ha affrontato l'argomento basandosi su un esempio applicativo, che tratta dati occupazionali, è quello di Gordon et al,⁹ che ha posto a confronto un'analisi *pooled* a effetti fissi con una metanalisi a effetti fissi

di risultati pubblicati. Dal lavoro si evince che, sebbene l'analisi *pooled* possa essere considerata uno strumento migliore, in alcuni casi il suo utilizzo non è appropriato, ovvero quando si può supporre che non esistano o siano pochi gli studi non pubblicati e quando il controllo di potenziali confondenti è stato già trattato in modo esaustivo.

Nell'ambito dell'epidemiologia ambientale non si hanno termini di paragone con studi che trattino il confronto tra metanalisi e analisi *pooled*, in particolare per quanto riguarda lo studio degli effetti a breve termine dell'inquinamento sulla salute.

Il protocollo metanalitico EpiAir⁴ rappresenta il principale riferimento a livello nazionale per gli studi che trattano questa tematica. L'obiettivo del presente studio è mettere a confronto una metanalisi con un'analisi *pooled* al fine di dimostrare che, in un contesto caratterizzato da non eterogeneità (a livello di fattori di esposizione e fattori meteo-climatici) come quello emiliano-romagnolo, risulta preferibile un approccio *pooled*, che garantisce una maggiore precisione delle stime d'effetto. In particolare, dopo avere illustrato le ipotesi alla base di ciascun metodo, verrà presentata un'applicazione di entrambi i metodi sui dati di mortalità e di inquinamento relativi a sette città dell'Emilia-Romagna che fanno parte del gruppo di città indagate dallo studio EpiAir2.⁵

MATERIALI E METODI

In questo paragrafo le due tipologie di analisi vengono illustrate teoricamente e applicate entrambe allo stesso set di dati ambientali e sanitari.

Lo studio EpiAir⁴ ha rappresentato un riferimento metodologico sia per le scelte sulla raccolta dei dati sia per la definizione dei modelli; per i dettagli si rimanda alle relative pubblicazioni.

Le analisi sono state condotte utilizzando il software STATA12.

Dati ambientali e sanitari

I dati relativi ai 7 Comuni considerati (Bologna, Modena, Ferrara, Piacenza, Reggio Emilia, Rimini e Parma) sono stati raccolti secondo il protocollo EpiAir2.⁵ I dati ambientali sono stati forniti dall'Agenzia regionale prevenzione e ambiente (ARPA) dell'Emilia-Romagna, mentre quelli sanitari dal Servizio sistema informativo sanità e politiche sociali della Regione Emilia-Romagna.

I parametri ambientali analizzati in relazione agli effetti sulla salute sono le frazioni respirabili di particolato (PM₁₀ e PM_{2.5}) e il biossido di azoto (NO₂).

Per quanto riguarda i dati di mortalità, sono stati analizzati i deceduti di 35 anni o più nel comune di residenza per quattro cause di morte: • naturali (ICD-9-CM 1-799), • cardiovascolari (ICD-9-CM 390-459), • cerebrovascolari (ICD-9-CM 430-438), • respiratorie (ICD-9-CM 460-519). La distribuzione di frequenza dei decessi è riportata in tabella 1.

Modelli specifici per città

I modelli costruiti hanno la finalità di studiare la relazione tra variazioni delle concentrazioni giornaliere degli inquinanti e le variazioni nel numero giornaliero di decessi per le quattro cause sopracitate.

Nei modelli si è tenuto conto del confondimento dovuto alle variabili: decremento estivo di popolazione, giorni di festa, epidemie influenzali, basse temperature, temperature elevate e trend temporale di medio-lungo periodo; quest'ultimo confondente è stato trattato seguendo l'approccio *time-stratified* del disegno *case-crossover*.⁴

Si è scelto di non considerare come confondente la pressione barometrica (variabile considerata in EpiAir),^{4,5} poiché non si disponeva dei dati per l'anno 2006 e perché analisi preliminari hanno mostrato l'irrelevanza dell'inserimento di tale variabile sulle stime d'effetto. In questo modo è stato

possibile far coincidere il nostro periodo di studio con quello dello studio di riferimento (EpiAir).^{4,5}

Il periodo di osservazione è di 5 anni (2006-2010) e le rilevazioni di ogni dato sono state effettuate giornalmente. In ogni studio specifico per città effettuato, l'associazione tra inquinamento atmosferico ed esiti sanitari avversi viene indagata mediante la costruzione di tanti modelli quanti sono gli eventi sanitari di interesse a tre diversi lag temporali che permettono di descrivere un effetto immediato (0-1 giorni), ritardato (2-5 giorni) e prolungato (0-5 giorni) degli inquinanti sugli esiti sanitari.

La variabile dipendente è il numero giornaliero di decessi per patologie legate all'inquinamento in una regressione di Poisson, in cui la funzione di link è logaritmica.

Negli studi di serie storiche, i *Generalised additive model* (GAM) sono il metodo più applicato, poiché costituiscono un criterio flessibile rispetto all'approccio completamente parametrico; infatti, permettono di considerare covariate i cui effetti non sono lineari rispetto alla variabile dipendente tramite l'inserimento di funzioni non parametriche (*spline*). Il numero osservato di eventi nel giorno t ($t=1, \dots, 1.826$) è espresso dalla variabile Y_t e il numero di eventi attesi può essere espresso dalla relazione:

$$E(Y_t) = \exp(\beta X_t + \delta C_t)$$

dove:

X_t è la variabile di esposizione e si assume abbia un effetto comune su ogni soggetto;

C_t rappresenta l'insieme dei fattori di confondimento specifici per giorno;

β e δ sono i rispettivi vettori di coefficienti.

Le variabili «basse temperature» e «alte temperature» sono state inserite nei modelli sotto forma di *spline*.

Per ogni modello specifico per città sono state utilizzate *spline* naturali con nodi corrispondenti a quattro specifici percentili delle distribuzioni delle due variabili sopracitate. Per la variabile «basse temperature» si sono scelti il 5°, 10°, 25° e 50° percentile, mentre per la variabile «alte temperature» il 50°, 75°, 90° e 95° percentile.

CITTÀ	POPOLAZIONE RESIDENTE	CAUSE NATURALI	MALATTIE CARDIO VASCOLARI	MALATTIE CEREBRO VASCOLARI	MALATTIE RESPIRATORIE
Bologna	259.418	19.223	7.466	1.742	1.808
Ferrara	95.816	7.875	3.013	778	496
Modena	118.712	8.599	3.218	702	665
Parma	118.333	8.545	3.454	812	579
Piacenza	67.558	5.019	1.919	568	361
Reggio Emilia	100.090	6.519	2.833	707	619
Rimini	90.385	5.724	2.250	491	418
Totale (modello <i>pooled</i>)	850.312	61.504	24.153	5.800	4.946

Tabella 1. Distribuzione di frequenza dei decessi di pazienti residenti e deceduti nelle 7 città in studio e in totale, per patologia di decesso, nell'intero periodo di analisi.

Table 1. Frequency distribution of the deaths of resident died patients in the 7 cities included in the study and overall, for pathology of death, throughout the period of analysis.

Le misure d'effetto sono state espresse come incrementi percentuali del rischio di morte per tutte le cause naturali, per patologie cardiovascolari, cerebrovascolari e respiratorie, per incrementi di $10 \mu\text{g}/\text{m}^3$ delle concentrazioni di inquinanti; inoltre, per ciascun incremento è stato calcolato l'intervallo di confidenza al 95%.

Metanalisi a effetti fissi

Analisi effettuate all'interno dello studio EpiAir2⁵ e ripetute dagli autori per i dati del seguente studio hanno evidenziato una non eterogeneità nelle stime d'effetto tra le città emiliano-romagnole. Pertanto, a differenza di quanto fatto a livello nazionale in EpiAir,⁴ si è scelto di svolgere una meta-analisi a effetti fissi.

La meta-analisi è una tecnica che produce una stima d'effetto sintetica a partire dalle stime specifiche per città che si ottengono effettuando analisi distinte per le sette aree urbane esaminate.

Sotto l'ipotesi che le stime dei coefficienti provengano da studi tra loro indipendenti e siano realizzazioni di variabili casuali normali con stessa media e varianza nota, la stima sintetica a effetti fissi massimizza la verosimiglianza per il parametro di interesse, che rappresenta l'effetto unico sottostante il fenomeno osservato.¹⁰

La stima di massima verosimiglianza dell'effetto globale è:

$$\hat{\theta}_{FE}^{ML} = \frac{\sum_{i=1}^k \omega_i \hat{\theta}_i}{\sum_{i=1}^k \omega_i}$$

dove:

$\hat{\theta}$ è la stima d'effetto specifica per città;

$$\omega_i = \frac{1}{s_i^2}$$

ovvero l'inverso della varianza città-specifica.

La stima di massima verosimiglianza si distribuisce come una normale avente media θ_{FE} e varianza

$$\left[\sum_{i=1}^k \omega_i \right]^{-1}$$

Ovviamente, nella stima dell'effetto globale avranno più peso le stime specifiche per città con varianza minore.

Analisi *pooled* a effetti fissi

Anche nell'analisi *pooled* si è indagata l'associazione tra inquinamento atmosferico e gli stessi esiti sanitari avversi considerati nella meta-analisi a tre diversi lag temporali.

Il modo standard per impostare il *dataset* per *pooled time-series* (cioè serie temporali aggregate per unità costanti nel tempo) è quella di raggruppare i dati prima per città e poi per giorni, cioè si dispongono uno sotto l'altro i blocchi contenenti i dati giornalieri relativi alle 7 città.

Per ogni città e per ogni giorno si ha il valore delle variabili già considerate nelle analisi specifiche per città svolte prima della meta-analisi. Oltre a queste, l'impostazione del modello ha richiesto l'inserimento di altre due variabili: una indicatrice della città e l'altra corrispondente al termine di *offset*. Il termine di *offset* serve a considerare che i conteggi giornalieri del numero di decessi si riferiscono a città diverse con differente ammontare di popolazione residente; equivale a inserire un ulteriore regressore con coefficiente pari a 1. Nell'applicazione proposta in questo studio corrisponde al valore medio della popolazione ≥ 35 anni residente nei 7 Comuni studiati nei 5 anni del periodo considerato; nelle analisi specifiche per città non è stato inserito, poiché i dati erano tutti riferiti allo stesso Comune.

Sul *dataset* generato è stato applicato un modello GAM del tipo:

$$\log(E(Y_{it})) = \beta X_{it} + \delta C_{it} + \alpha_i + \ln(z_i)$$

dove:

Y_{it} è il valore di Y per l'i-esima città nella t-esima unità temporale;

X è il vettore delle variabili di esposizione;

C quello dei fattori di confondimento;

β e δ sono, rispettivamente, i vettori dei loro coefficienti; α_i rappresenta la variabile indicatrice della città, è fissa e cattura l'effetto sulla variabile dipendente dei fattori osservabili e non osservabili che variano tra le città, ma sono costanti nel tempo;

$\ln(z_i)$ rappresenta il termine di *offset*.

Nel modello sono presenti *spline*, analogamente ai modelli specifici per città, per le variabili «basse temperature» e «alte temperature» (all'interno della matrice C_{it}). Per ognuna di queste variabili, ciascuno dei quattro nodi delle *spline* è stato calcolato come media dei relativi nodi che erano stati ottenuti nelle analisi specifiche per città.

RISULTATI

La tabella 2 mostra alcune statistiche descrittive delle variabili meteo-climatiche e di esposizione. Emerge un'evidente omogeneità tra città relativamente a tutte le variabili, a eccezione della concentrazione media giornaliera di NO_2 . Per questo inquinante risulta, infatti, una minore uniformità a

		CITTÀ						
		BOLOGNA	FERRARA	MODENA	PARMA	PIACENZA	REGGIO EMILIA	RIMINI
Temperatura (°C) media giornaliera	Media	14,7	14,6	14,6	15,2	14,4	14,8	15,1
	SD	8,6	8,3	8,5	8,4	8,5	8,5	7,6
	Range IQ	14,4	14,1	14,1	14,0	14,2	14,1	12,6
	50° percentile	14,7	14,8	14,7	15,3	14,6	14,8	15,2
	90° percentile	26,6	25,8	26,1	26,6	25,8	26,3	25,3
Temperatura apparente (°C) media giornaliera	Media	13,7	13,5	13,2	13,9	13,1	13,5	14,2
	SD	9,9	9,6	9,4	9,4	9,5	9,5	8,9
	Range IQ	16,8	16,3	15,9	15,8	16,0	16,1	15,3
	50° percentile	13,3	13,4	13,0	13,7	12,9	13,3	13,8
	90° percentile	27,7	27,0	26,3	26,9	26,3	26,7	26,6
Umidità relativa (%) media giornaliera	Media	69,0	69,0	65,0	64,0	66,0	66,0	68,0
	SD	17,0	17,0	18,0	19,0	17,0	18,0	15,0
	Range IQ	27,0	27,0	31,0	32,0	29,0	31,0	29,0
	50° percentile	69,0	68,0	63,0	62,0	64,0	65,0	67,0
	90° percentile	93,0	92,0	90,0	90,0	89,0	91,0	88,0
Concentrazione media giornaliera di PM ₁₀	Media	38,4	37,5	41,0	35,4	37,8	37,1	34,9
	SD	21,6	22,7	24,0	20,2	22,2	21,1	18,5
	Range IQ	24,0	25,0	28,3	23,5	26,0	23,8	20,6
	50° percentile	32,0	31,0	34,3	30,5	32,0	31,7	30,2
	90° percentile	66,0	70,0	73,7	61,0	69,0	64,3	59,9
Concentrazione media giornaliera di PM _{2,5}	Media	25,8	25,3	21,0	21,0	–	21,5	20,1
	SD	18,5	18,9	14,3	15,8	–	14,5	15,5
	Range IQ	18,0	21,0	15,0	15,0	–	15,0	16,0
	50° percentile	20,0	19,0	16,0	15,0	–	17,0	15,0
	90° percentile	49,0	52,2	41,0	43,0	–	42,0	43,0
Concentrazione media giornaliera di NO ₂	Media	50,3	42,9	52,6	38,9	33,6	43,1	39,2
	SD	18,3	15,2	18,4	15,8	14,5	16,3	15,9
	Range IQ	24,7	20,3	25,0	24,5	20,2	22,2	20,7
	50° percentile	49,3	41,7	51,8	37,5	32,4	41,5	37,2
	90° percentile	73,5	62,6	76,8	59,5	52,6	64,8	60,6

Tabella 2 Statistiche descrittive dei dati meteorologici e dei dati di esposizione elaborati nel periodo in studio.

Table 2 Descriptive statistics of meteorological data and exposition data during the study period.

causa della diversa tipologia di centraline di monitoraggio distribuite nelle aree urbane considerate.

La tabella 3 mostra che le ampiezze degli intervalli di confidenza delle stime d'effetto dell'analisi *pooled* risultano sempre minori rispetto a quelle degli intervalli di confidenza delle stime metanalitiche, sottolineando una precisione maggiore delle stime d'effetto derivanti dall'analisi *pooled*. In certi casi il **restringimento degli intervalli di confidenza** porta ad avere risultati statisticamente significativi nell'analisi *pooled* e non nella metanalisi. In particolare:

- per ogni incremento di 10 µg/m³ di PM₁₀ si rileva un incremento immediato del rischio di mortalità per cause naturali pari a 0,55 sia per l'analisi *pooled* sia per la metanalisi con intervalli di confidenza al 95% rispettivamente di (0,00;1,10) e (-0,07;1,17);

- per ogni incremento di 10 µg/m³ di PM_{2,5} si rileva un incremento ritardato del rischio di mortalità naturale pari a 1,08 (IC95% 0,12;2,05) nell'analisi *pooled* e 0,91 (IC95% -0,16;1,99) nella metanalisi e un incremento prolungato del

rischio di mortalità naturale pari a 1,32 (IC95% 0,21;2,44) nell'analisi *pooled* e 1,17 (IC95% -0,08;2,45) nella metanalisi;

- per ogni incremento di 10 µg/m³ di NO₂ si riscontra un incremento immediato del rischio di mortalità per cause naturali di 0,84 (IC95% 0,00;1,68) e 0,86 (IC95% -0,21;1,96) rispettivamente per analisi *pooled* e metanalisi.

In altri casi la **significatività statistica** di certe stime risultata solamente dall'analisi *pooled* deriva, oltre che da un restringimento degli intervalli di confidenza, anche da stime puntuali più elevate. In particolare:

- per ogni incremento di 10 µg/m³ di PM₁₀ si rileva un incremento ritardato del rischio di mortalità naturale di 0,78 (IC95% 0,15;1,42) nell'analisi *pooled* e di 0,44 (IC95% -0,27;1,16) nella metanalisi e un incremento prolungato del rischio di mortalità naturale di 0,89 (IC95% 0,16;1,63) nell'analisi *pooled* e di 0,57 (IC95% -0,27;1,42) nella metanalisi;

- per ogni incremento di 10 µg/m³ di PM₁₀ si rileva un in-

		MORTALITÀ NATURALE			MORTALITÀ CARDIOVASCOLARE			MORTALITÀ CEREbroVASCOLARE			MORTALITÀ RESPIRATORIA		
PM ₁₀	LAG	PA	MA	Δ (%)*	PA	MA	Δ (%)*	PA	MA	Δ (%)*	PA	MA	Δ (%)*
	0-1	1,10	1,24	11,07	1,74	1,95	10,62	3,62	3,88	6,77	3,76	4,33	13,12
	2-5	1,27	1,43	11,11	2,01	2,25	10,66	4,18	4,69	10,90	4,33	4,93	12,23
	0-5	1,47	1,69	13,06	2,33	2,67	12,57	4,84	5,57	13,17	5,02	5,80	13,45
PM _{2,5}	0-1	1,64	1,83	10,46	2,61	2,92	10,73	5,46	6,10	10,53	5,50	6,13	10,32
	2-5	1,93	2,15	10,31	3,07	3,40	9,80	6,43	7,37	12,83	6,46	7,24	10,72
	0-5	2,24	2,53	11,61	3,56	4,04	11,89	7,44	8,64	13,90	7,49	8,44	11,19
NO ₂	0-1	1,68	2,17	22,44	2,67	4,70	43,24	5,63	7,21	21,95	5,97	7,60	21,45
	2-5	1,90	2,55	25,29	3,05	4,08	25,29	6,30	8,64	27,13	6,65	8,71	23,69
	0-5	2,10	2,95	28,82	3,36	4,72	28,76	7,00	10,03	30,17	7,42	10,08	26,41

*Variazione percentuale delle ampiezze degli intervalli di confidenza secondo la relazione: [(ampiezzaMA – ampiezzaPA) / ampiezzaMA * 100]

Tabella 3 Ampiezze degli intervalli di confidenza delle stime d'effetto di analisi *pooled* (PA) e metanalisi (MA) ai vari lag.

Table 3 Confidence intervals widths of pooled analysis (PA) and meta-analysis (MA) effect estimates at the various lag.

cremento del rischio di mortalità per cause cardiovascolari di 1,22 (IC95% 0,22;2,23) per l'analisi *pooled* e di 0,84 (IC95% -0,28;1,97) per la metanalisi;

■ gli unici due casi in cui risultano statisticamente significative solo le stime prodotte con metanalisi sono relativi a incrementi di 10 µg/m³ di PM₁₀ sulla mortalità per cause cerebrovascolari al lag 2-5 (2,01; IC95% -0,05;4,12 per l'analisi *pooled* – 2,42; IC95% 0,10;4,79 per la metanalisi) e respiratorie al lag 0-1 (1,00; IC95% -0,86;2,90 per l'analisi *pooled* – 4,05; IC95% 1,91;6,24 per la metanalisi); questo avviene poiché le stime metanalitiche risultano più elevate delle stime *pooled*, ma non migliori in termini di precisione.

DISCUSSIONE

Metanalisi e analisi *pooled* sono due metodi per sintetizzare dati relativi a sottogruppi di unità al fine di ottenere una misura globale dell'effetto studiato. I due metodi non sono interscambiabili, poiché non sempre si possono applicare entrambi: infatti, mentre l'analisi *pooled* richiede la disponibilità dei dati relativi ai singoli studi che si vogliono sintetizzare, la metanalisi può essere condotta anche avendo a disposizione le sole stime specifiche per città.

L'applicazione del modello *pooled* ha fornito stime più precise. Questo guadagno di potenza statistica è da attribuire al numero inferiore di gradi di libertà rispetto al modello metanalitico. Nel modello *pooled*, infatti, i confondenti vengono modellati come effetto medio delle 7 città e non in maniera specifica per città (come avviene nella metanalisi). In altre parole, l'approccio *pooled* è più parsimonioso di quello metanalitico, poiché effettua un aggiustamento medio del confondimento, evitando così un iper-aggiustamento specifico per città, sotto l'ipotesi di omogeneità delle esposizioni e dei confondenti.

Le analisi hanno mostrato corrispondenza tra le stime d'effetto ai vari lag, evidenziando una forte analogia tra metanalisi e analisi *pooled* relativamente a direzione e forza del-

l'associazione tra gli *outcome* sanitari e i vari inquinanti. Osservando i *forest plot* (figura 1) si evince che, in generale, il modello *pooled* mette in luce un numero maggiore di effetti statisticamente significativi al 95%: come si è detto precedentemente, talvolta ciò è dovuto a un valore più elevato delle stime *pooled* rispetto a quelle metanalitiche e/o a una maggiore precisione delle stime, che si riflette in intervalli di confidenza più stretti.

Una maggiore attendibilità delle stime *pooled* sembra emergere dall'analisi dell'effetto del PM₁₀ sulla mortalità respiratoria al lag 0-1. La stima metanalitica è molto diversa dalla stima *pooled* e mostra un incremento di rischio del 4,05%, molto anomalo rispetto ai dati di letteratura, ai dati delle altre aggregazioni di patologia e al corrispondente dato ottenuto per il PM_{2,5}. Tale valore metanalitico risente principalmente della stima ottenuta per la città di Bologna (IR% 8,67; IC95% 4,19-12,55) molto più elevata rispetto alle altre città. La stima *pooled*, al contrario, mostra una maggiore coerenza interna e con i dati di letteratura.

Le differenze nei valori delle stime derivanti dai due approcci potrebbero essere dovute principalmente a tre ragioni:

1. nell'analisi *pooled* l'effetto delle covariate è modellato allo stesso modo su tutte le città in un unico passaggio, che restituisce direttamente stime d'effetto globali per ogni inquinante; nella metanalisi l'effetto delle covariate è, invece, modellato separatamente per ognuna delle 7 città, nel primo di due passaggi, ovvero quello delle analisi specifiche per città; la stima d'effetto globale per ogni inquinante si ottiene come sintesi delle stime d'effetto specifiche per città nel secondo e ultimo passaggio;

2. in entrambe le analisi ci si è trovati di fronte alla scelta di quattro nodi per le *spline* usate (in ogni modello) per il confondimento indotto dalle variabili «basse temperature» ed «alte temperature». Mentre nei modelli specifici per città i nodi delle *spline* coincidono con determinati percentili delle distribuzioni delle due covariate, nell'analisi *pooled* ognuno dei

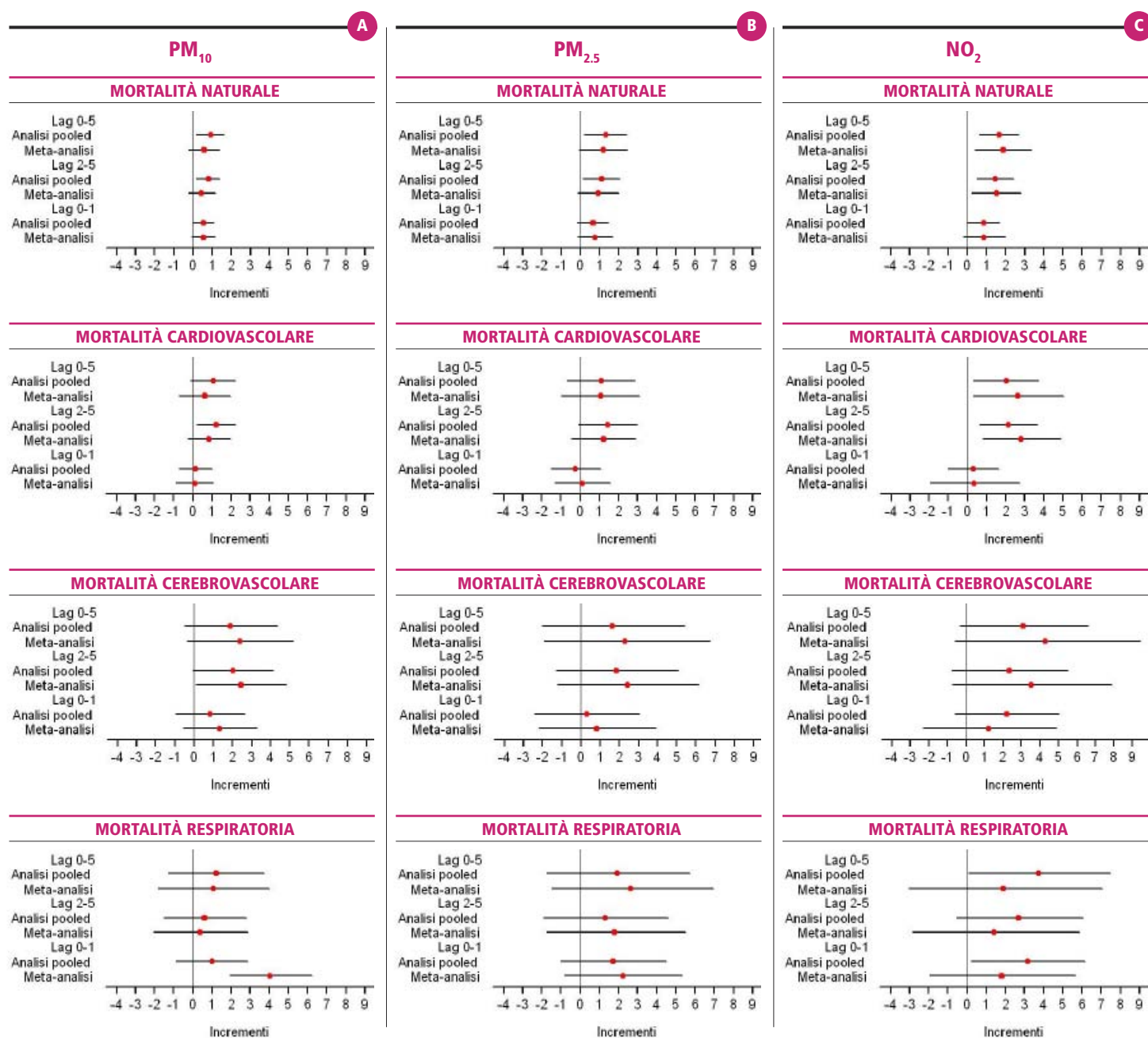


Figura 1. Forest plot dei risultati di metanalisi a effetti fissi e analisi pooled a effetti fissi. Incrementi percentuali del rischio di morte di soggetti di 35 anni o più per cause naturali, patologie cardiache, cardiovascolari e respiratorie, per incrementi di $10 \mu\text{g}/\text{m}^3$ di PM_{10} , $\text{PM}_{2.5}$ e NO_2 con relativi intervalli di confidenza al 95%.

Figure 1. Forest plot of the results of the fixed effects meta-analysis and fixed effects pooled analysis. Percent increases of death risk for individuals aged 35+ years for natural causes, heart disease, cardiovascular and respiratory diseases, relative to $10 \mu\text{g}/\text{m}^3$ variation in PM_{10} , $\text{PM}_{2.5}$ and NO_2 with confidence intervals at 95%.

quattro nodi delle *spline* è stato ricavato come media dei rispettivi sette nodi delle analisi specifiche per città;

3. si potrebbe essere in presenza di confondimento residuo nel caso in cui l'ipotesi di omogeneità non sia completamente soddisfatta e, quindi, non è più sufficiente un aggiustamento medio.

La perfetta equivalenza dei due approcci in termini di risultati potrebbe essere dimostrata teoricamente inserendo nel modello *pooled*, oltre alle variabili dei modelli specifici per città, anche le interazioni tra queste e la variabile indicatrice della città. Tuttavia, un modello di questo tipo coinvolge un numero elevato di parametri e, in questo studio, si è preferito utilizzare un modello più parsimonioso che garantisce una maggiore precisione delle stime d'effetto sotto l'ipotesi di assenza di eterogeneità nell'esposizione agli inquinanti e agli altri confondenti.

Infine, va sottolineato che, in entrambe le analisi, gli intervalli di confidenza corrispondenti alle stime d'effetto sulla mortalità cerebrovascolare e respiratoria sono molto ampi (tabella 2); ciò potrebbe essere imputabile a un numero giornaliero di decessi inferiore per queste patologie rispetto a quello delle altre cause di morte considerate (tabella 1). Analisi di tipo aggregato riducono le variazioni dovute all'errore casuale e consentono guadagni in termini di precisione e accuratezza delle stime.¹¹

CONCLUSIONI

Questo studio ha mostrato che in un contesto non eterogeneo come quello emiliano-romagnolo, in cui i centri analizzati sono uniformi relativamente alle variabili di confondimento

e di esposizione, quando si dispone dei dati dei conteggi giornalieri è preferibile fare ricorso a un modello *pooled* rispetto a una metanalisi. Infatti, sebbene entrambi i metodi abbiano la finalità di produrre stime sintetiche d'effetto, il modello *pooled* è uno strumento più potente e consente un controllo più robusto delle variabili di confondimento, poiché utilizza un numero meno elevato di parametri e, di conseguenza, di gradi di libertà. Inoltre, può accadere che per città medio-piccole, come quelle dell'Emilia-Romagna, i modelli specifici per città producano stime d'effetto poco accurate a causa della scarsa numerosità di eventi in esse rilevati, problematica che non si presenta utilizzando un modello *pooled*.

Tuttavia, se si tratta un numero elevato di centri si genera una maggiore eterogeneità a livello di variabili di confondimento e di esposizione. Pertanto, l'applicazione di un approccio *pooled* si rivela poco efficace e, quindi, in questo caso è più indicato usare una metanalisi. Inoltre, se le dimensioni dei campioni di ciascun centro sono elevate, i guadagni in termini di stabilità del modello *pooled* si riducono rispetto a quelli specifici per città.

La ripetizione del confronto dei due metodi su un arco temporale maggiore e/o su dati che seguono altre distribuzioni potrà confermare o meno le analogie e le differenze riscontrate nel presente lavoro.

Conflitti di interesse dichiarati: nessuno.

Ringraziamenti: un sentito ringraziamento alla professoressa Rossella Miglio per averci suggerito l'idea che ci ha portato all'elaborazione dello studio proposto nell'articolo e alla dottoressa Anna Freni Sterrantino per averci fornito consigli validi e preziosi.

BIBLIOGRAFIA

1. Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* 1999;28(1):1-9.
2. Biggeri A, Bellini P, Terracini B; Italian MISA Group. Meta-analysis of the Italian studies on short-term effects of air pollution. *Epidemiol Prev* 2001;25(2) Suppl:1-72.
3. Biggeri A, Bellini P, Terracini B. Meta-analysis of the Italian studies on short-term effects of air pollution – MISA 1996-2002. *Epidemiol Prev* 2004;28(4-5) Suppl:4-100.
4. Berti G, Galassi C, Faustini A, Forastiere F (eds). Inquinamento atmosferico e salute: sorveglianza epidemiologica e interventi di prevenzione. *Epidemiol Prev* 2009;33(6) Suppl:1-144.
5. Alessandrini E, Faustini A, Chiusolo M; Gruppo collaborativo EpiAir et al. Inquinamento atmosferico e mortalità in venticinque città italiane: risultati del progetto EpiAir2. *Epidemiol Prev* 2013;37(4-5):220-9.
6. Checkoway H. Data pooling in occupational studies. *J Occup Med* 1991;33(12):1257-60.
7. Friedenreich CM. Methods for pooled analysis of epidemiologic studies. *Epidemiology* 1993;4(4):295-302.
8. Smith SJ, Steinberg KK, Thacker SB. Method for pooled analysis of epidemiologic studies. *Epidemiology* 1994;5(3):381-3.
9. Gordon I, Boffetta P, Demers PA. A case study comparing a meta-analysis and a pooled analysis of studies of sinonasal cancer among wood workers. *Epidemiology* 1998;9(5):518-24.
10. Böhning D. *Computer assisted analysis of mixtures and applications. Meta-analysis, disease mapping and others*. Boca Raton (FL), Chapman and Hall CRC, 1999.
11. Clayton D. The EURODEM collaborative re-analysis of case-control studies of Alzheimer's disease: some methodological considerations. *Int J Epidemiol* 1991;20 Suppl 2:S62-4.