

SOSTENIBILITÀ PER L'AMBIENTE E LA SALUTE DEI CITTADINI NELLE CITTÀ PORTUALI IN ITALIA (progetto PNC PREV-B-2022-12376988 – regione capofila, Puglia)

Studio di coorte di popolazione nelle aree portuali

Roma, maggio 2024

Dipartimento di Epidemiologia del
Servizio Sanitario Regionale del Lazio

Via C. Colombo, 112 - 00147 Roma - **www.deplazio.net**

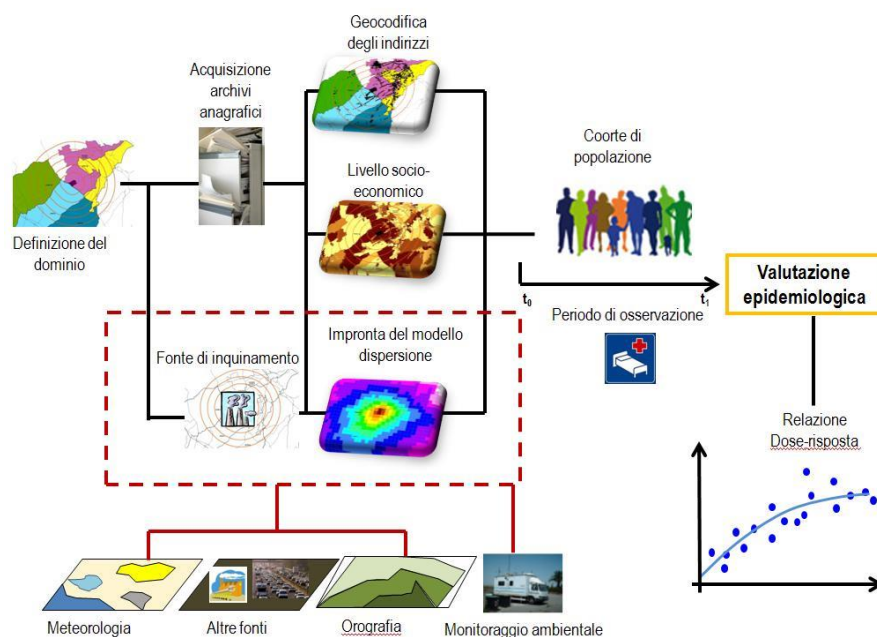
L'epidemiologia ambientale valuta la relazione tra i livelli di inquinamento nelle varie matrici ambientali e i possibili effetti sulla salute della popolazione (morbosità stimata attraverso i dati di ospedalizzazione, l'incidenza di tumori, o la mortalità per causa). Il disegno di studio epidemiologico più adatto allo studio delle relazioni ambiente e salute è certamente l'approccio di coorte residenziale che consente la valutazione degli effetti a lungo termine dell'esposizione ambientale e la stima del tempo esatto in cui ogni soggetto è stato a rischio di sviluppare l'evento in studio (tempo-persona).

Gli elementi necessari per poter condurre questo tipo di studio sono:

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

La metodologia proposta è riassunta nella Figura 1 (tratta da Ancona et al., 2014).

Figura 1. Metodologia integrata ambiente e salute



I Sistemi Informativi Sanitari regionali sono la fonte dei dati per l'arruolamento della coorte e per la definizione degli esiti sanitari. L'esposizione sarà definita sulla base dell'impronta delle concentrazioni al suolo di specifici inquinanti considerati come traccianti dell'inquinamento prodotto dalle navi e dalle attività portuali per i porti di Civitavecchia, Brindisi e Bari. Tale impronta verrà stimata utilizzando opportuni modelli di dispersione degli inquinanti scelti come traccianti. Per la corretta definizione dell'esposizione e per la corretta definizione delle variabili individuali (livello socio-economico) e di contesto (eventuali altri fattori di pressione ambientale presenti nel territorio, incluso il traffico stradale e/o la vicinanza di strade ad alta intensità di traffico) è necessario che le coorti siano georeferenziate (attribuzione delle coordinate geografiche agli indirizzi di residenza dei soggetti inclusi nello studio). Allegato al presente documento vengono riportate le esperienze di confronto anagrafe comunale con anagrafe assistibili per le coorti di Civitavecchia e Brindisi arruolate per precedenti studi epidemiologici (Ancona et al., 2016; Bauleo et al., 2017)

Arruolamento di coorti residenziali da fonte anagrafe assistibili e accertamento degli esiti sanitari

Dall'archivio regionale dell'anagrafe assistibili sarà arruolata la coorte dei residenti nelle aree interessate dal 1° gennaio 2011 al 31 dicembre 2023 includendo tutti gli ingressi (per nascita o immigrazione) e tracciando tutte le uscite (per morte o emigrazione) nel periodo in studio.

Per ogni individuo della coorte, laddove disponibili, verranno selezionate le variabili relative a informazioni anagrafiche, l'indirizzo di residenza (*al baseline*) e tutte le date relative ad eventuali movimenti da e per il comune (data di emigrazione, immigrazione). Saranno inoltre raccolte informazioni sui cambiamenti di residenza all'interno dello stesso comune se disponibili.

La coorte selezionata sarà confrontata con eventuali coorti di popolazioni esistenti e selezionate dagli archivi comunali al fine di caratterizzare e descrivere l'eventuale errore che si commette passando da una coorte di tipo comunale ad una di tipo sanitaria.

Successivamente si procederà alla selezione della coorte dei residenti:

- verrà selezionata la coorte dei residenti nell'area in studio per 0-99 anni nel periodo di studio;
- si verificherà l'eventuale presenza di record doppi per codice individuale e in quel caso verranno conseguentemente eliminate le informazioni ripetute, salvo che si tratti di cambi di residenza, in quel caso da mantenere;
- verranno eventualmente eliminati i record con errori nelle date (per esempio data di nascita > data ingresso anagrafe assistibili).
- per ogni individuo verrà accertato lo stato in vita alla fine del follow-up (vivo/morto, emigrato o irreperibile).

L'attribuzione della causa di decesso verrà effettuata attraverso una procedura di record linkage con il Registro Regionale delle Cause di Morte se disponibile o sfruttando

l'Indagine su Decessi e Cause di Morte dell'ISTAT, mentre come fonte dati dei ricoveri ospedalieri si utilizzerà il Sistema Informativo Ospedaliero (SIO) regionale.

Georeferenziazione degli indirizzi di residenza delle coorti in studio

Gli indirizzi di residenza di tutti i soggetti arruolati, saranno georeferenziati per elaborare mappe del territorio e attribuire ad ogni individuo della coorte le variabili ambientali e di contesto sociale utili per definire al meglio l'associazione tra l'esposizione e mortalità/morbosità per causa.

Per la geocodifica degli indirizzi di residenza si suggerisce di usare la metodologia proposta da Mataloni et al. (2022) che permette di georeferenziare indirizzi di qualità non ottimali senza ricorrere a dispendiose procedure di normalizzazione². I passaggi tecnici per la procedura sono riportati in figura 2.

Figura 2 – le fasi del processo di geocodifica



In breve:

- Fase 1: Attraverso l'utilizzo delle *Application Programming Interface* (API) di Google Maps, e di una procedura scritta in Python, è possibile interrogare i servizi di Google per ottenere coordinate geografiche di ogni indirizzo. L'output che viene restituito da questa procedura non contiene solo le coordinate geografiche (latitudine e longitudine), ma anche l'indirizzo per esteso per il quale è stata verificata la corrispondenza rispetto alla stringa originale e il livello di accuratezza con cui è stata geocodificata ogni via, secondo le impostazioni di Google. L'accuratezza può essere definita come la precisione delle coordinate attribuite rispetto alla reale localizzazione della residenza dei soggetti e può essere di diversi livelli (in ordine decrescente di precisione):
 1. **rooftop** per gli indirizzi geocodificati al civico puntuale;
 2. **range interpolated** nel caso in cui non sia stato trovato il numero civico di input e siano state riportate le coordinate del punto mediano tra due civici vicini a quello indicato;
 3. **geometric center** col quale l'indirizzo viene attribuito al centroide di una piccola area nei dintorni all'indirizzo richiesto;
 4. **approximate** se l'indirizzo è stato geocodificato in un punto centrale o di interesse del Comune nel quale si ricercava l'indirizzo specifico.
- Fase 2: Le coordinate degli indirizzi geocodificati nella prima fase vengono mappati per verificare che fossero stati proiettati esattamente all'interno del Comune afferente. Tutti gli indirizzi con geocodifica nulla o geocodificati in un altro comune o con un livello di accuratezza non adeguato sono stati sottoposti a una nuova geocodifica automatica descritta nella fase 3.
- Fase 3: Attraverso il pacchetto *taRifx.geo* 1.0.7 sviluppato attraverso il software R vengono nuovamente geocodificati gli indirizzi mediante API di Bing Maps.
- Fase 4: Come per la fase 2, anche in questo caso si procede alla verifica della geocodifica ottenuta.
- Fase 5: I restanti indirizzi che non sono stati geocodificati con le procedure automatiche descritte, vengono poi ricercati manualmente utilizzando diverse fonti online, al fine di raggiungere la migliore copertura possibile.

Al termine di tutta la procedura, a ogni indirizzo è stato attribuito un punteggio che indica la qualità

delle coordinate ottenute:

- 0:** l'indirizzo non è stato geocodificato;
- 1:** l'indirizzo è stato geocodificato con un buon livello di accuratezza (in questo caso viene indicata anche l'accuratezza della geocodifica tra *rooftop*, *range interpolated*, *geometric center*, *ricerca con Bing Maps*, *ricerca a mano*);
- 2:** l'indirizzo è stato geocodificato al confine del comune indicato sulla stringa dell'indirizzo;
- 3:** l'indirizzo è stato geocodificato con un basso livello di accuratezza (*approximate*).

Definizione dell'esposizione e modelli di dispersione

Saranno utilizzati una serie di indicatori ambientali geo-referenziati con elevata risoluzione spaziale (1 km²) e temporale (giornaliera) estesi alle città portuali in studio.

Per questi scopi è necessario disporre di uno strumento in grado di stimare la variabilità spaziale e temporale delle concentrazioni ad alta risoluzione spaziale applicabile omogeneamente ai domini di interesse.

Saranno sviluppati modelli bayesiani a struttura gerarchica con approccio INLA-SPDE ad alta risoluzione spazio-temporale, per la stima del particolato, PM₁₀ e PM_{2,5}, e del biossido di azoto (NO₂), con l'uso di predittori spazio-temporali in cui la variabile dipendente sarà rappresentata dai dati di monitoraggio della qualità dell'aria. Si prevede di utilizzare, tra le variabili indipendenti, i dati delle concentrazioni di inquinanti stimati attraverso modelli deterministici (ad esempio modelli di trasporto chimico o di dispersione). L'introduzione di stime di concentrazioni da modellistica ha lo scopo di migliorare l'accuratezza e la copertura sia spaziale che temporale delle stime di concentrazione degli inquinanti.

Il framework metodologico e di validazione è quello descritto in Fioravanti et al., 2021 e nel protocollo dello studio: "Attività 2.1.0: analisi retrospettiva dei dati raccolti - Attività 2.2.0: studio della relazione con le attività portuali e il trend delle emissioni locali legate alle attività portuali" dell'U.O.2.

I modelli ad alta risoluzione spazio-temporale presentano una stima della variabilità spaziale alla risoluzione di 1kmx1km omogenea su tutto il territorio nazionale e offrono una base di partenza per migliorare ulteriormente la distribuzione spaziale, introducendo variabili predittive locali misurate o calcolate all'interno della cella di 1kmx1km.

Al fine dello studio di coorte di popolazione nelle aree portuali verrà stimata questa variabilità spaziale in alcune città portuali, partendo proprio dalle stime del modello omogeneo su scala nazionale, per giungere alla stima dei livelli di concentrazione alla più alta risoluzione spaziale utilizzabile al fine della stima di esposizione all'indirizzo della popolazione residente.

Il dominio di riferimento sarà rappresentato dalle 3 città portuali target, Civitavecchia, Bari e Brindisi.

I modelli di predizione saranno costruiti per il 2022 e per il 2013. I modelli relativi al 2013 si considerano validi retroattivamente anche per gli anni 2012 e 2011, in modo da poter assegnare un'esposizione a ciascun individuo della coorte, inclusi i soggetti già presenti alla data di inizio del follow-up (1° gennaio 2011), la quale è antecedente quella a cui il modello di predizione è relativo. È infatti improbabile che l'esposizione valutata nel 2013 differisca in maniera significativa da quella dei due anni precedenti in termini di contrasto tra diverse classi di esposizione, seppure differenze assolute nelle stime tra i vari anni possano essere plausibili.

Saranno stimate le medie annuali delle concentrazioni di PM₁₀, PM_{2.5} e NO₂ alla massima risoluzione spaziale possibile compatibile con il metodo utilizzato e con le risorse di calcolo disponibili.

Al fine di sviluppare i modelli di predizione, si renderà necessaria la raccolta di un esteso set di parametri spaziali e spazio-temporali.

I predittori spaziali dovranno includere variabili alla più alta risoluzione disponibile per città.

La ricognizione effettuata ha permesso di individuare le seguenti variabili effettivamente utilizzabili:

- Imperviousness Density (Buffer 25,50,100 m)
 - Tutte le città: 2018 (10m), 2015, 2012, 2009, 2006 (20m)
- Urban Atlas Building Height (Buffer 25,50,100 m)
 - Tutte le città: 2012 (10m)
- Corine Land Cover (Buffer 25,50,100):
 - Civitavecchia (100m; 2018, 2012, 2006)
- Urban Atlas Land Cover/Land Use (Buffer 25,50,100 m)
 - Altre città: (10m; 2018, 2012, 2006)
- Normalised Difference Vegetation Index (NDVI)/Leaf area index (LAI) (Buffer 25,50,100 m)
 - Tutte le città: 2022-2017 (10m e 300m); 2016-2014 (300m); 2013-2006 (1000m)
- Popolazione (ISTAT) (Buffer 25,50,100 m):
 - sezioni censuarie 2021, 2011, 2001
- Strade (Open Street Map) (Buffer 25,50,100 m) – Unico Layer

Sono in corso le elaborazioni di variabili rappresentative del carico emissivo dei porti tenuto conto del fattore limitante legato al fatto che le stime ad alta risoluzione prodotte per il 2022 non sono disponibili per il 2013.

Per lo sviluppo dei modelli, i valori delle variabili spaziali saranno calcolati in buffer di diverso raggio (ad esempio 25m, 50m), a partire dai punti rappresentativi delle stime del modello nazionale di 1kmx1km, con l'obiettivo di rappresentare non solo ciò che avviene nel punto ma anche considerare l'influenza dell'intorno.

Per la scelta dei modelli andrà verificata l'esistenza di un legame lineare o più complesso tra la variabile risposta e le variabili esplicative; in caso di legami complessi, partendo da distribuzioni normali della variabile risposta, potranno essere utilizzati ad esempio

modelli additivi generalizzati (GAM; Ravindra et al, 2019; Clifford et al, 2011; Patton et al, 2014).

Nella scelta delle variabili saranno considerate le correlazioni tra le variabili (escludendo quindi variabili “strettamente” correlate attraverso le matrici di correlazione), l’analisi dei residui per verificare le ipotesi di base e la significatività statistica dei coefficienti delle variabili; potranno essere considerati ulteriormente, a seconda della scelta del modello, gli andamenti delle funzioni di smoothing e i parametri che consentono la convergenza dell’algoritmo di iterazione. Tutte le fasi di sviluppo dei modelli predittivi e di valutazione della loro performance, con il calcolo di alcuni dei più noti indici statistici di bontà dei modelli, saranno condotte applicando tecniche di cross-validazione (CV).

L’eventuale presenza di variabili statisticamente significative rappresentative in modo specifico dell’influenza del porto e delle attività correlate sulla variabilità spaziale delle concentrazioni, consentirà di valutare, per quota parte, l’“effetto porto” sull’esposizione della popolazione sulla base della varianza spiegata da queste variabili (e.g. Gaeta et al., 2017; Leone et al., 2023).

Esiti sanitari

Gli esiti sanitari attribuibili agli individui facenti parte della popolazione in studio possono riferirsi a esiti di mortalità o morbosità.

La scelta degli esiti sanitari di cui studiare l’associazione con le esposizioni ambientali dovrebbe tenere conto anche del tempo di latenza delle specifiche malattie, in rapporto al tempo di osservazione che si ha per i soggetti inclusi nella coorte. Volendo studiare malattie a lungo tempo di induzione usando un tempo di follow-up relativamente breve, si rischia infatti di non ottenere associazioni significative che invece si potrebbero eventualmente osservare in un intervallo di tempo maggiore.

Il rischio appena descritto è ancora più alto nel caso di malattie che producano un numero esiguo di casi osservati durante il periodo di studio. In questa eventualità va anche posta attenzione alla distribuzione degli esiti tra le diverse aree di esposizione,

qualora questa venga categorizzata, affinché ci sia un contrasto tra esposti e non esposti sufficiente a studiare eventuali associazioni.

Per quanto riguarda gli esiti di mortalità, l'attribuzione avviene tramite una procedura di record linkage con il Registro Regionale delle Cause di Morte se disponibile o sfruttando l'Indagine su Decessi e Cause di Morte dell'ISTAT.

Cause di morte (ICD 9 – ICD 10)
Mortalità naturale (001-799)
Tumori maligni (140-208 ICD9; C00-D48 ICD10)
Stomaco (151 ICD9; C16.0 ICD10)
Colon e retto (153-154; 159 ICD9; C18-C20 ICD10)
Fegato (155-156 ICD9; C22 ICD10)
Pancreas (157 ICD9; C25 ICD10)
Laringe (161 ICD9; C32 ICD10)
Trachea, bronchi e polmoni (162 ICD9; C33 ICD10)
Pleura (163 ICD9; C38.4 ICD10)
Vescica (188 ICD9; C67 ICD10)
Reni (189 ICD9; C64 ICD10)
Cervello e altre parti del SNC (191-192; 225 ICD9; C69-C72 ICD10)
Tessuti linfatici ed ematopoietici (200-208 ICD9; C81-C96 ICD10)
Leucemie (204-208 ICD9; C91-C95 ICD10)
Malattie neurologiche (330-349 ICD9; G00-G99 ICD10)
Malattie cardiovascolari (390-459 ICD9; I00-I99 ICD10)
Malattie cardiache (390-429 ICD9; I05-I09, I20-I5A ICD10)
Eventi coronarici acuti (410-411 ICD9; I24 ICD10)
Malattie cerebrovascolari (430-438 ICD9; I60-I69 ICD10)
Malattie respiratorie (460-519 ICD9; J00-J99 ICD10)
Infezioni delle vie respiratorie (460-466 ICD9; J00-J22 ICD10)
Bronco Pneumopatia Cronico Ostruttiva - BPCO (490-492, 494, 496 ICD9; J40-J47 ICD10)
Malattie renali (580-599 ICD9; N00-N29 ICD10)

Per attribuire esiti di morbosità la fonte è invece il Sistema Informativo Ospedaliero (SIO). Tramite il codice identificativo individuale si possono individuare le ospedalizzazioni subite dai soggetti di interesse nel periodo in studio e diversificare le cause di ricovero tramite l'apposita codifica ICD-9/ICD-10. Nel contesto di studi di coorte residenziale, le

operazioni di filtraggio sulle ospedalizzazioni comprendono di solito la sola inclusione di ricoveri urgenti/non programmati.

Cause di ospedalizzazione (ICD 9 - ICD 10)
Diabete (250 ICD9; E08-E13 ICD10)
Malattie neurologiche (330-349; G00-G99 ICD10)
Malattie cardiovascolari (390-459 ICD9; I00-I99 ICD10)
Malattie cardiache (390-429 ICD9; I05-I09, I20-I5A ICD10)
Malattie ischemiche (410-414 ICD9; I20-I25 ICD10)
Eventi coronarici acuti (410-411 ICD9; I24 ICD10)
Scompenso cardiaco (428 ICD9; I50 ICD10)
Malattie cerebrovascolari (430-438 ICD9; I60-I69 ICD10)
Malattie respiratorie (460-519 ICD9; J00-J99 ICD10)
Infezioni delle vie respiratorie (460-466 ICD9; J00-J22 ICD10)
Bronco Pneumopatia Cronico Ostruttiva - BPCO (490-492, 494, 496 ICD9; J40-J47 ICD10)
Asma (493 ICD9; J45 ICD10)
Polmonite (J12, J18 ICD10)
Malattie del sistema digerente (520-579 ICD9; K00-K95 ICD10)
Malattie renali (580-599 ICD9; N00-N29 ICD10)
Schizofrenia (F20-F20.9 ICD10)
Malattie della cute e del tessuto sottocutaneo (L00-L99 ICD10)

Revisione della letteratura

Le città portuali fungono da nodi essenziali per il commercio globale.. I benefici economici che apportano sono accompagnati da importanti sfide ambientali. Diverse forme di inquinamento derivanti da traffico marittimo, operazioni portuali e attività industriali che caratterizzano questo tipo di città rappresentano una minaccia multiforme per la salute e il benessere di residenti, lavoratori e visitatori di questi centri. Le evidenze relative ai potenziali rischi per la salute associati all'esposizione a tali emissioni sono ad oggi ancora incerte. Al fine di aggiornare le evidenze disponibili verranno condotte due diverse revisioni sistematiche della letteratura sugli effetti dell'esposizione residenziale all'inquinamento atmosferico e all'inquinamento acustico in città portuali.

Le revisioni saranno condotte seguendo le linee guida PRISMA (Page et al., 2021). La ricerca verrà effettuata su 3 database: Medline, Embase e Web of Science. La strategia di ricerca sarà costruita utilizzando una combinazione di termini liberi e sarà adattata a ciascun database seguendone le specifiche regole di sintassi. Non saranno applicate restrizioni temporali o di lingua.

Oggetto delle revisioni saranno studi relativi alla valutazione degli effetti a breve e lungo termine associati all'esposizione residenziale all'inquinamento atmosferico e acustico in città portuali. Sulla base dell'acronimo PEO (Popolazione, Esposizione, Outcome) sono stati definiti i criteri di inclusione. Verranno inclusi studi che considerino la popolazione generale residente in città portuali. Saranno esclusi studi che considerino popolazioni di lavoratori portuali.

Gli studi inclusi dovranno valutare l'esposizione all'inquinamento atmosferico e acustico. Saranno considerati tutti gli esiti associati a questo tipo di fattori di pressione ambientale. Malattie respiratorie, cardiovascolari e neurologiche, in aggiunta a incidenza di tumore sono tra gli esiti che in letteratura risultano più frequentemente associati (Viena 2020, Anderson 2011, Mueller 2023).

Titolo e abstract dei record derivanti dalla ricerca bibliografica verranno sottoposti a screening da parte di 2 revisori indipendenti, che procederanno poi ad estrarre e valutare il full-text dei record potenzialmente includibili. Per ogni record incluso verrà poi effettuata in doppio l'estrazione dei dati relativi a paese, periodo e disegno di studio, metodo di valutazione dell'esposizione, popolazione e outcome considerati e risultati. La qualità degli studi verrà valutata da 2 revisori indipendenti impiegando il tool Risk of bias assessment instrument for systematic reviews informing WHO Global Air Quality Guidelines.

Verranno considerate anche revisioni sistematiche già presenti in letteratura che rispecchino i criteri PEO precedentemente descritti.

Analisi statistica e misure di associazione

Metodi

Le caratteristiche dei soggetti partecipanti verranno riportate come numerosità totale e percentuale per le variabili categoriche e come media e deviazione standard o mediana e range interquartile per le variabili continue. Ogni residente della coorte contribuisce al computo degli anni persona a rischio, a partire dalla data di inizio del follow-up se già presente nell'area in studio a questa data, o a partire dalla data del suo ingresso nell'area studiata se entrato dopo la data di inizio del follow-up. Il contributo di ciascun residente al computo degli anni persona dura fino alla data di morte, emigrazione, o fine del follow-up. Gli anni persona saranno calcolati separatamente per sesso, classi di età e periodo di calendario. Gli anni-persona verranno utilizzati per il calcolo di prevalenza e incidenza come misure di occorrenza degli esiti sanitari di interesse.

L'associazione tra le concentrazioni di inquinanti e gli esiti di salute, quali eventi di morte, ospedalizzazione o insorgenza di malattia, sarà valutata mediante il modello di regressione di Cox per l'analisi di sopravvivenza. Il tempo di sopravvivenza sarà misurato dalla data di ingresso nel follow-up fino alla data di conclusione prevista per follow-up se non si verifica l'evento o la data di realizzazione dell'evento entro il 2023. Il rischio stimato dal modello di Cox che, nello specifico, prende il nome di Hazard Ratio (HR) è calcolato come rapporto tra due rischi di sperimentare un evento ed esprime l'eccesso (o il difetto) di rischio del gruppo posto al numeratore rispetto al gruppo posto al denominatore.

La relazione tra l'esposizione ad un inquinante ad un certo livello di concentrazione e il rischio di un evento viene quindi rappresentata all'interno del termine esponenziale del modello di Cox e definita come funzione concentrazione-risposta (C-R). La stima della funzione C-R è fondamentale per la valutazione dell'impatto di interventi normativi volti a migliorare la qualità dell'aria. La loro efficacia, infatti, viene spesso valutata prevedendo i cambiamenti sulla salute pubblica derivanti dalla loro attuazione, tenendo conto dell'associazione tra inquinante e gli esiti di salute. Per catturare in modo più accurato l'effetto delle variabili che possono variare nel tempo, il modello di Cox sarà esteso per includere covariate tempo-dipendenti. Un modo per codificare covariate tempo-dipendenti è l'utilizzo di intervalli temporali per ciascun soggetto attraverso una rielaborazione formale dei dati, aggiornando i valori di esposizione a seconda del periodo di follow-up. Questo è particolarmente rilevante per variabili come le concentrazioni di

inquinanti, che possono fluttuare durante i periodi di follow-up. Includere variabili tempo-dipendenti consente di modellare l'effetto dinamico di tali esposizioni sugli esiti di salute, migliorando la precisione delle stime (Therneau et al. 2000, 2017). Verrà valutata l'evidenza sull'assunto di proporzionalità e considerata l'eventuale gestione di coefficienti tempo-dipendenti nel modello. Sarà esaminata l'associazione tra inquinanti e esiti tenendo conto della possibile relazione sopralineare tra le due (Burnett et al. 2018). Per la stima della funzione C-R saranno valutati differenti metodi statistici, tra cui la *Shape Constrained Health Impact Function* (SCHIF), la spline cubica ristretta (*restricted cubic spline*, RCS) e/o *monotonically increasing smoothing spline* (MISS, Nasari et al. 2016). L'HR sarà riportato con gli opportuni intervalli di confidenza al 95% (IC 95%). Nel modello di analisi l'età è considerata come asse temporale, usando il periodo di calendario, il livello socio-economico ed eventuali ulteriori variabili di contesto come variabili di aggiustamento.

Le analisi verranno condotte separatamente sulla popolazione adulta (30+) per le tre aree di Civitavecchia, Bari e Brindisi e congiuntamente nelle tre aree. Una unica analisi aggregata verrà condotta sulla popolazione pediatrica (0-14) sulla base di esiti specifici. Infine, i risultati dei modelli considerati saranno valutati sia graficamente sia attraverso l'utilizzo di opportuni test statistici. Saranno valutati ulteriori modelli per esaminare eventuali associazioni non lineari tra inquinamento ed esiti di salute.

Un valore di p-value minore di 0.05 sarà considerato statisticamente significativo.

L'analisi statistica verrà condotta utilizzando il software R (R Foundation for Statistical Computing, 2023).

Controllo del confondimento

È opportuno raccogliere tutte le informazioni sui possibili confondenti della relazione tra inquinanti considerati e mortalità/morbosità della popolazione della coorte, quali età, periodo di calendario, posizione socio-economica (SEP), residenza in prossimità di strade principali, autostrade e industrie, ed eventualmente, occupazione e storia lavorativa comprendente attività portuali da dati INPS.

Il SEP verrà usato anche come stratificatore per tenere conto di possibili misclassificazioni dovute alla non presenza di alcune categorie di popolazione nell'anagrafe degli assistibili.

BIBLIOGRAFIA

- Ancona C, Mataloni F, Badaloni C, et al. Residential cohort approach in industrial contaminated sites: the ERAS Lazio project. *Epidemiol Prev.* 2014;38(2 Suppl 1):158-161
- Anderson JO, Thundiyil JG, Stolbach A. Clearing the air: a review of the effects of particulate matter air pollution on human health. *J Med Toxicol.* 2012 Jun;8(2):166-75. doi: 10.1007/s13181-011-0203-1. PMID: 22194192; PMCID: PMC3550231.
- Baker D, Nieuwenhuijsen MJ: *Environmental epidemiology: Study methods and application.* Oxford University Press 2008.
- Bauleo L, Bucci S, Antonucci C, et al. Long-term exposure to air pollutants from multiple sources and mortality in an industrial area: a cohort study. *Occup Environ Med.* 2019;76(1):48-57. doi:10.1136/oemed-2018-105059
- Bellander T, Berglind N, Gustavsson P, Jonson T, Nyberg F, Pershagen G, Järup L.: Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. *Environ Health Perspect* 2001 Jun; 109(6):633-9.
- Burnett R, Chen H, Szyszkowicz M, Fann N, Hubbell B, Pope CA 3rd, Apte JS, Brauer M, Cohen A, Weichenthal S, Coggins J, Di Q, Brunekreef B, Frostad J, Lim SS, Kan H, Walker KD, Thurston GD, Hayes RB, Lim CC, Turner MC, Jerrett M, Krewski D, Gapstur SM, Diver WR, Ostro B, Goldberg D, Crouse DL, Martin RV, Peters P, Pinault L, Tjepkema M, van Donkelaar A, Villeneuve PJ, Miller AB, Yin P, Zhou M, Wang L, Janssen NAH, Marra M, Atkinson RW, Tsang H, Quoc Thach T,

Cannon JB, Allen RT, Hart JE, Laden F, Cesaroni G, Forastiere F, Weinmayr G, Jaensch A, Nagel G, Concin H, Spadaro JV. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc Natl Acad Sci U S A*. 2018 Sep 18;115(38):9592-9597. doi: 10.1073/pnas.1803222115. Epub 2018 Sep 4. PMID: 30181279; PMCID: PMC6156628

- Caranci N., Biggeri A., Grisotto L., Pacelli B., Spadea T., Costa G. L'indice di deprivazione italiano a livello di sezione di censimento: definizione, descrizione e associazione con la mortalità. *Epidemiologia e prevenzione*, 34, 4, 167-176
- Carla Ancona, Chiara Antonucci, Lisa Bauleo, Simone Bucci, Marina Davoli e Francesco Forastiere, rapporto "Effetti delle esposizioni ambientali ed occupazionali sulla mortalità della popolazione residente nell'area di Civitavecchia", Maggio 2016
- Gaeta, A.; Leone, G.; Di Menno di Bucchianico, A.; Cusano, M.; Gaddi, R.; Pelliccioni, A.; Reatini, M.A.; Di Bernardino, A.; Cattani, G. Spatio-Temporal Modeling of Small-Scale Ultrafine Particle Variability Using Generalized Additive Models. *Sustainability* 2022, 14:313. <https://doi.org/10.3390/su14010313>.
- Landfill Gas Emissions Model www.epa.gov/ttn/catc/dir1/landgem-v302.xls
- Leone, G.; Cattani, G.; Cusano, M.; Gaeta, A.; Pellis G.; Vitullo, M.; Morelli, R. Wildfires Impact Assessment on PM Levels Using Generalized Additive Mixed Models. *Atmosphere* 2023, 2023, 14, 231. <https://doi.org/10.3390/atmos14020231>.
- Lisa Bauleo, Carla Ancona, Angela Morabito, Alessandra Nocioni, Roberto Giua, Giuseppe Spagnolo, Susi Epifani, Antonino Ardizzone, Stefano Spagnolo, Francesca Intini, Tiziano Pastore, Ilenia Schipa, Maria Serinelli, Annalisa Tanzarella, Vito Petrarolo, Giorgio Assennato, Lucia Bisceglia e Francesco Forastiere, rapporto "Studio di coorte sugli effetti delle esposizioni ambientali sulla mortalità e morbosità della popolazione residente a Brindisi e nei comuni limitrofi", Maggio 2017

- M. Viana, V. Rizza, A. Tobías, E. Carr, J. Corbett, M. Sofiev, A. Karanasiou, G. Buonanno, N. Fann, Estimated health impacts from maritime transport in the Mediterranean region and benefits from the use of cleaner fuels, *Environment International*, Volume 138, 2020, 105670, ISSN 0160-4120, <https://doi.org/10.1016/j.envint.2020.105670>.
- Mataloni F, Bauleo L, Badaloni C, Nobile F, Savastano J, Noccioli F, Salatino CG, Balducci M, Cappai G, Rosa AC, Asta F, Fusco D, Michelozzi P, Davoli M. Geocodificare un milione di indirizzi mediante l'utilizzo di API: una procedura semiautomatica a più step [Geocoding one million of addresses using API: a semiautomatic multistep procedure]. *Epidemiol Prev.* 2022 May-Jun;46(3):160-167. Italian. doi: 10.19191/EP22.3.A463.031. PMID: 35443573.
- Nasari MM, Szyszkowicz M, Chen H, Crouse D, Turner MC, Jerrett M, Pope CA 3rd, Hubbell B, Fann N, Cohen A, Gapstur SM, Diver WR, Stieb D, Forouzanfar MH, Kim SY, Olives C, Krewski D, Burnett RT. A class of non-linear exposure-response models suitable for health impact assessment applicable to large cohort studies of ambient air pollution. *Air Qual Atmos Health.* 2016;9(8):961-972. doi: 10.1007/s11869-016-0398-z. Epub 2016 Mar 2. PMID: 27867428; PMCID: PMC5093184
- Natalie Mueller, Marie Westerby, Mark Nieuwenhuijsen, Health impact assessments of shipping and port-sourced air pollution on a global scale: A scoping literature review, *Environmental Research*, Volume 216, Part 1, 2023, 114460, ISSN 0013-9351, <https://doi.org/10.1016/j.envres.2022.114460>.
- Nyberg F, Gustavsson P, Järup L, Bellander T, Berglind N, Jakobsson R, Pershagen G.: Urban air pollution and lung cancer in Stockholm. *Epidemiology* 2000 Sep; 11(5):487-95.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: An updated guideline for reporting systematic

reviews. PLoS Med. 2021 Mar 29;18(3):e1003583. doi:

10.1371/journal.pmed.1003583. PMID: 33780438; PMCID: PMC8007028.

- R Foundation for Statistical Computing. (2023). *R: A Language and Environment for Statistical Computing*.
- RAMS <http://rams.atmos.colostate.edu/rams-description.html>
- Therneau, T., Crowson, C., & Atkinson, E. (2017). Using time dependent covariates and time dependent coefficients in the cox model. *Survival Vignettes*, 2(3), 1-25.
- Therneau, Terry M.; Grambsch, Patricia M. (2000). [Statistics for Biology and Health] *Modeling Survival Data: Extending the Cox Model*

ALLEGATO:

Confronto coorte da anagrafe regionale con anagrafe assistibili

Civitavecchia

Al fine di verificare l'efficacia della selezione della popolazione in studio tramite anagrafe assistibili, è stata fatta una valutazione riguardante i comuni inclusi in un precedente studio di Bauleo et al. (2019). È stata selezionata, sia da anagrafe comunale che da anagrafe assistibili, la popolazione residente nei comuni di Allumiere, Civitavecchia, Santa Marinella, Tarquinia e Tolfa. La selezione è avvenuta considerando la medesima finestra temporale per entrambe le fonti (a partire dal 31/12/2006, data di istituzione dell'anagrafe assistibili, e fino al 31/12/2013, data di ultimo aggiornamento della coorte da anagrafe comunale). Sono stati quindi esclusi da questa valutazione gli individui usciti dalla coorte comunale prima del 31/12/2006 o entrati in una delle due anagrafi dopo il 31/12/2013. Successivamente, sulla base del codice identificativo individuale attribuito nei sistemi informativi sanitari regionali, è stato fatto un linkage tra i due dataset. Si sono venuti a creare diversi sottogruppi: individui che si linkano tra entrambi i dataset, individui che sono presenti solo nella coorte comunale e individui che sono presenti solo nella popolazione selezionata da anagrafe assistibili. In seguito sono state eseguite analisi descrittive per età al baseline (ricalcolata al 31/12/2006 per gli individui iscritti alla coorte comunale prima di tale data), tempo-persona di permanenza nella coorte, sesso e stato socioeconomico di ciascuno di questi sottogruppi al fine di valutare eventuali divergenze che potrebbero introdurre bias in fase di successive analisi selezionando la popolazione sulla base dell'assistibilità sanitaria. Ulteriori controlli sul sottogruppo di popolazione presente solo nella coorte su base anagrafica comunale (fonte che viene considerata lo standard di riferimento) hanno evidenziato la presenza di una parte di individui presenti nell'anagrafe assistibili con un comune di residenza diverso da quello indicato nella coorte comunale, oppure con lo stesso comune ma con il primo contatto con l'anagrafe assistibili successivo alla chiusura della coorte comunale. D'altra parte sono presenti

individui che non vengono ritrovati nell'anagrafe degli assistibili, per la maggior parte nati all'estero ed entrati nella coorte comunale per immigrazione. Questo potrebbe in parte spiegare le differenze osservate per variabili individuali. Di seguito l'analisi descrittiva dei sottogruppi derivanti dalle operazioni di linkage.

La tabella descrittiva comprende anche una colonna riguardante la popolazione compresa in entrambe le fonti anagrafiche utilizzate per il confronto. Per descrivere la distribuzione della popolazione per età e stato socio-economico, è stato possibile quindi utilizzare le variabili relative a queste informazioni presenti in uno o nell'altro database. Mentre nell'anagrafe comunale ogni soggetto è rappresentato da un'unica osservazione, nell'anagrafe assistibili ciascun soggetto può essere registrato con osservazioni multiple in base a eventuali cambi di residenza. Lo stato socio-economico individuale viene quindi calcolato nell'anagrafe degli assistibili come la media di tale indicatore tra tutti gli indirizzi di residenza con cui uno stesso soggetto sia stato registrato. Per quanto riguarda le classi di età, eventuali discrepanze nella distribuzione tra i due database potrebbero essere dovute a differenze tra le date di iscrizione nell'anagrafe comunale e la data di registrazione nell'anagrafe degli assistibili.

È infine presente una tabella che specifica l'accuratezza della geocodifica nei vari sottogruppi di popolazione. Per recuperare tali informazioni sui soggetti presenti nell'anagrafe comunale, questi individui sono stati ricercati anche nell'anagrafe degli assistibili, e tale ricerca è stata in grado di attribuire coordinate geografiche a una parte di questi individui. Cercando nell'anagrafe assistibili gli individui residenti, da anagrafe comunale, nei 5 comuni in studio, è stato osservato come questi abbiano assegnato un comune di residenza compreso in una lista di circa 180 diverse municipalità.

Tabella 1. Analisi descrittiva dei sottogruppi derivanti dalle operazioni di linkage tra anagrafe comunale e anagrafe assistibili nella coorte di Civitavecchia e comuni limitrofi.

¹ Mean (SD)	Anagrafe comunale	Anagrafe assistibili	Popolazione che si trova in entrambe le anagrafi		Popolazione che si trova solo nell'anagrafe comunale	Popolazione che si trova solo nell'anagrafe assistibili
			Da AC	Da AA		
	N = 115,544 ¹	N = 114,154 ¹	Overall, N = 109,248 ¹	Overall, N = 109,248 ¹	N = 6,296 ¹	N = 4,906 ¹
Età al baseline	40 (24)	40 (24)	40 (24)	40 (24)	38 (21)	42 (23)
Anni persona	5.97 (1.97)	11.6 (4.8)	6.10 (1.84)	11.9 (4.6)	3.85 (2.72)	5.2 (5.0)
Sesso						
Maschi	56,196 (49%)	55,218 (48%)	52,765 (48%)	52,765 (48%)	3,431 (54%)	2,453 (50%)
Femmine	59,348 (51%)	58,936 (52%)	56,483 (52%)	56,483 (52%)	2,865 (46%)	2,453 (50%)
Età in classi						
<18	23,550 (20%)	23,201 (20%)	22,654 (21%)	22,535 (21%)	896 (14%)	666 (14%)
18 - 34	26,059 (23%)	25,220 (22%)	23,948 (22%)	23,760 (22%)	2,111 (34%)	1,460 (30%)
35 - 49	26,007 (23%)	25,739 (23%)	24,404 (22%)	24,525 (22%)	1,603 (25%)	1,214 (25%)
50 - 64	20,268 (18%)	20,078 (18%)	19,290 (18%)	19,353 (18%)	978 (16%)	725 (15%)
65 - 79	14,538 (13%)	14,660 (13%)	14,090 (13%)	14,185 (13%)	448 (7.1%)	475 (9.7%)
80 +	5,122 (4.4%)	5,256 (4.6%)	4,862 (4.5%)	4,890 (4.5%)	260 (4.1%)	366 (7.5%)
1: Bassa deprivazione	20,023 (17%)	14,864 (13%)	18,644 (17%)	14,288 (13%)	1,379 (22%)	576 (12%)
2	20,527 (18%)	21,689 (19%)	19,362 (18%)	21,028 (19%)	1,165 (19%)	661 (13%)
3	32,447 (28%)	30,890 (27%)	30,802 (28%)	29,893 (27%)	1,645 (26%)	997 (20%)
4	18,503 (16%)	23,981 (21%)	17,566 (16%)	23,347 (21%)	937 (15%)	634 (13%)
5: Alta deprivazione	19,697 (17%)	14,342 (13%)	18,840 (17%)	13,860 (13%)	857 (14%)	482 (9.8%)
Missing	4,347 (3.8%)	8,388 (7.3%)	18,840 (17%)	13,860 (13%)	313 (5.0%)	1,556 (32%)

Tabella 2. Proporzione di individui presenti solo nella popolazione selezionata da anagrafe comunale, per comune di residenza.

Città	N = 6,296¹
ALLUMIERE	180 (2.9%)
CIVITAVECCHIA	2,219 (35%)
SANTA MARINELLA	2,804 (45%)
TARQUINIA	845 (13%)
TOLFA	248 (3.9%)
¹ n (%)	

Tabella 3. Accuratezza della geocodifica, da anagrafe assistibili, per i sottogruppi derivanti dalle operazioni di linkage

	ANAGRAFE COMUNALE	ANAGRAFE ASSISTIBILI	LINKATI	SOLO AC	SOLO AA
APPROXIMATE	834	6291	6063	18	228
CONFINE	1	55	54	0	1
GEOMETRIC_CENTER	3373	44151	43377	80	774
RANGE_INTERPOLATED	1556	40315	39412	42	903
RICERCA A MANO	275	6527	6453	19	74
RICERCA CON R	396	4088	4038	6	50
ROOFTOP	6,844	231021	226136	244	4885
MISSING	4,153 (183 persone)	61,185 (8,388 persone)	58,828 (6,832 persone)	142 (6 persone)	2,357 (1,556 persone)