epa Research

Report No. 496

INHALE: Irish Nationwide Health and Air Quality Linkage

Authors: Éilis J. O'Reilly, Claire Buckley, Christina Dillon and Stig Hellebust

Lead organisation: University College Cork

What did this research aim to address?

Every day, every hour, air quality is monitored across Ireland. At the same time, thousands of routine health and social care interactions take place, each generating data that could be used to advance a programme of air quality and health research in Ireland. Safeguarding patient and service user privacy is the foremost concern, but barriers to accessing data prevent meaningful research on the health effects of pollution.

What this means in practice is that researchers with air quality data, curated and made accessible by the EPA, cannot link the relevant pollution exposure to specific individuals in health data. To protect privacy, health data are made available on aggregated units (hospital attended or county; monthly counts). Air quality varies substantially within these aggregations; therefore, this level of aggregation does not let us establish who is exposed to higher levels of air pollution.

The Irish Nationwide Health and Air Quality Linkage (INHALE) project reviewed the methods of linking air quality and health data in other jurisdictions and recommended ways to create capacity for world-leading environmental research in the Irish population.

What did this research find?

The INHALE project identified best-practice and novel approaches to data linkage to support research and policy-driven change. The report details two main methods for safely linking air quality and health data. The first recommendation is investment in trusted research environment (TRE) infrastructure that can fulfil requests to create linked, de-identified individual-level data for researchers or government departments through a secure platform.

In the absence of a TRE, the INHALE project recommends a second researcher-led solution where data custodians provide aggregated data on a unit relevant to the research question. First, the researcher assigns the relevant (for the condition being studied) air quality data to every residential Eircode (e.g. deciles of annual particulate matter of less than 10 micrometres or peak daily nitrogen dioxide) using EPA data. The researcher provides this file to the health data custodian to merge with the health data. Each Eircode (or other appropriate location identifier) occurring in the health data is assigned the corresponding level of pollution. The health data custodian then returns only the aggregated counts of individuals with the condition for each level of pollution.

How can the research findings be used?

The infrastructure and the data linkage expertise required to implement the recommendations provided by the INHALE project are already in place throughout Ireland, specifically within the Central Statistics Office (CSO), as is required by legislation. However, the CSO is not currently tasked with implementing this process, and in order to operationalise and implement the recommendations, additional human resources would need to be allocated to take ownership of the process.

The INHALE team strongly recommends that the EPA is at the table for future discussions of TREs and data linkage because of the organisation's wealth of experience with big data, and because air quality, environment and climate are going to remain leading causes of morbidity and mortality for the foreseeable future.

Project code: 2017-CCRP-MS.47