



---

Supplement Article

# Using large and complex datasets for small-area environment-health studies: from theory to practice

Frédéric B Piel <sup>1,2\*</sup> and Samantha Cockings <sup>3</sup>

<sup>1</sup>UK Small Area Health Statistics Unit (SAHSU), Department of Epidemiology & Biostatistics, School of Public Health, Imperial College London, London, UK, <sup>2</sup>National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Health Impact of Environmental Hazards, Imperial College London, UK, and <sup>3</sup>Department of Geography and Environmental Science, University of Southampton, Southampton, UK

\*Corresponding author. Norfolk Place, St Mary's Campus, W2 1PG London, UK. E-mail: f.piel@imperial.ac.uk

Editorial decision 14 January 2020; Accepted 11 February 2020

---

Humans are exposed to a wide range of pollutants throughout their lifetime, many of which pose a potential risk to their health. Such hazards include features of the natural, human-modified, social and economic environments. In this supplement, we are primarily concerned with risks to human health resulting from hazards of the human-modified environment, although many of the concepts, methods and tools are equally applicable to investigations of the health impacts of other types of environmental hazards. Amongst human-modified environmental hazards, air pollution has been identified as the world's largest killer, being responsible for an estimated 6.4 million deaths per year (1 in 9 deaths).<sup>1</sup> According to the World Health Organization, two billion children live in areas where outdoor air pollution exceeds recommended international limits and 300 million children live in areas where outdoor air pollution exceeds six times those international limits. Other hazards of the human-modified environment include water pollutants, such as chemicals and microplastics; radiation from mobile phones, powerlines or nearby nuclear installations; and soil contaminants such as heavy metals.

When health risks are very high and localized, they tend to be rapidly identified by alert clinicians, public health professionals or members of the public. This can for example set

the launch of a cluster investigation for which Public Health England recently published guidance.<sup>2</sup> It is much harder to identify risks when they are less obvious and more ubiquitous. Small-area methods provide a powerful means to study health effects at local, regional or national level taking into account spatial heterogeneities in socio-demographic characteristics and environmental exposures.

In the UK, the UK Small Area Health Statistics Unit (SAHSU, [www.sahsu.org](http://www.sahsu.org)) was established in 1987 to investigate the potential health effects of environmental pollutants, following reports of excess risks of leukaemia and non-Hodgkin lymphomas in young people living near the Sellafield nuclear plant.<sup>3</sup> In recent years, SAHSU has undertaken a series of small-area studies to investigate *inter alia* potential health risks associated with emissions from municipal waste incinerators,<sup>4–6</sup> air pollution,<sup>7</sup> aircraft noise,<sup>8</sup> disinfection by-products<sup>9</sup> and exposure to non-ionizing radiation from living near mobile phone base stations or overhead powerlines.<sup>10–12</sup>

Although small-area studies have been used for a long time, their popularity seems to be increasing due to efforts to quantify local burden of diseases and risk factors. It is therefore timely to provide an overview of the strengths and limitations of this particular study design, and of recent methodological advances.

Building on over 30 years of expertise in SAHSU, the *Education Corner* manuscript accompanying this supplement<sup>13</sup> presents an overview of the methodological steps and challenges associated with designing and completing small-area studies, including data access, data linkage and data privacy. The supplement itself provides a more detailed critique into the accessibility of small-area data and dissemination of results;<sup>14</sup> novel methods to produce high-resolution population data;<sup>15</sup> recent developments of statistical models to monitor non-communicable disease in both space and time;<sup>16</sup> user-friendly tools to trace residential history when assigning environmental exposures,<sup>17</sup> mapping disease risk and conducting risk analysis;<sup>18</sup> and a practical example of a national small-area study investigating a specific source of exposure.<sup>19</sup>

Hodgson *et al.*<sup>14</sup> provide a detailed description of the challenges involved in accessing, analysing and disseminating small-area data. Substantive changes to data access regulations pose significant challenges, given that the size of geographical units used in small-area studies often necessitates the use of sensitive data. Growing concerns over data privacy and confidentiality, reflected both by the implementation of the EU General Data Protection Regulation (GDPR) in May 2018 and the launch of the National Data Opt-out Programme on the same day in the UK, mean that requirements to access sensitive data are becoming stricter over time. In parallel, developing studies in collaboration with members of the public and representatives of various stakeholders—or co-design—is becoming more common. Although both of these are welcome developments, they need to be carefully considered during the development and implementation of any small-area study.

Using examples from the UK and the US, Fecht *et al.*<sup>15</sup> describe novel methods to produce time-specific high resolution denominator data for small-area studies by combining new and emerging forms of data (such as sensed footfall or traffic data) with traditional data sources (such as census or surveys). They outline the challenges involved in using new data sources such as the American Community Survey to produce estimates of population at risk and describe openly available software (e.g. SurfaceBuilder247) that facilitates the creation of gridded population distribution models for specific times and dates.

Blangiardo *et al.*<sup>16</sup> present a review of recent advances in spatio-temporal disease surveillance for non-communicable diseases (NCDs) using Bayesian hierarchical methods. They discuss key challenges in dealing with NCD surveillance, particularly how to account for false detection and the modifiable areal unit problem. Traditional models focused on identifying either spatial or temporal patterns, whereas more recent methods, such as the hierarchical models described, enable dependencies between data sources to be

exploited in both space and time. Furthermore, the Bayesian framework allows uncertainties to be quantified and taken into account throughout the modelling approach.

Public health authorities often need to provide rapid assessment of potential health risks in a given area. This may follow e.g. the report of a suspected cluster of disease cases. Piel *et al.*<sup>18</sup> describe the functionalities of the latest version of the Rapid Inquiry Facility (RIF) that has been developed by SAHSU as an open access software for disease mapping and risk analysis at small-area level.

Estimation of exposure to environmental pollutants often depends on information on residential address. Accurate reconstruction of residential histories can greatly reduce bias in exposure assignment in such studies, particularly in those spanning long time periods. Fecht *et al.*<sup>17</sup> present an algorithm for undertaking this complex and time-consuming task. They illustrate its use in constructing prenatal and early-life air pollution exposure for 14 541 pregnant women participating in the Avon Longitudinal Study of Parents and Children (ALSPAC) in the South West of England.

The final paper in the supplement, a study by Toledano *et al.*,<sup>19</sup> investigates the hypothesis that air ion density or electric fields in the vicinity of high-voltage overhead power lines may be associated with cancer risk in adults.

We hope that the expertise shared in this supplement and the examples provided highlight the ongoing usefulness of the small-area study design and will help readers conduct their own rigorous small-area studies to assess the health impacts of environmental hazards.

## Funding

The UK Small Area Health Statistics Unit (SAHSU) is funded by Public Health England (PHE). The MRC-PHE Centre for Environment and Health is supported by the Medical Research Council (MR/L01341X/1). Part of this work was supported by a Wellcome Trust Seed Award in Science to F.B.P. (204535/Z/16/Z). F.B.P. also acknowledges acknowledge support from the NIHR Health Protection Research Unit in Health Impact of Environmental Hazards (HPRU-2012-10141).

## References

1. Landrigan PJ, Fuller R, Acosta NJR *et al.* The Lancet Commission on pollution and health. *Lancet* 2018;391: 462–512.
2. Fletcher T, Crabbe H, Close R *et al.* *Guidance for Investigating Non-Infectious Disease Clusters from Potential Environmental Causes*. In: England PH (ed). London: Public Health England, 2019, p. 45.
3. Black D. *Investigation of the Possible Increased Incidence of Cancer in West Cumbria: Report of the Independent Advisory Group*. London: H.M.S.O., 1984.
4. Freni-Sterrantino A, Ghosh RE, Fecht D *et al.* Bayesian spatial modelling for quasi-experimental designs: An interrupted time series

- study of the opening of Municipal Waste Incinerators in relation to infant mortality and sex ratio. *Environ Int* 2019;128:109–15.
5. Ghosh RE, Freni-Sterrantino A, Douglas P *et al.* Fetal growth, stillbirth, infant mortality and other birth outcomes near UK municipal waste incinerators; retrospective population based cohort and case-control study. *Environ Int* 2019;122:151–8.
  6. Parkes B, Hansell AL, Ghosh RE *et al.* Risk of congenital anomalies near municipal waste incinerators in England and Scotland: retrospective population-based cohort study. *Environ Int* 2019;134:104845.
  7. Smith RB, Fecht D, Gulliver J *et al.* Impact of London's road traffic air and noise pollution on birth weight: retrospective population based cohort study. *BMJ* 2017;359:j5299.
  8. Hansell AL, Blangiardo M, Fortunato L *et al.* Aircraft noise and cardiovascular disease near Heathrow airport in London: small area study. *BMJ: Br Med J* 2013;347:f5432.
  9. Nieuwenhuijsen MJ, Toledano MB, Bennett J *et al.* Chlorination disinfection by-products and risk of congenital anomalies in England and Wales. *Environ Health Perspect* 2008;116:216–22.
  10. Elliott P, Toledano MB, Bennett J *et al.* Mobile phone base stations and early childhood cancers: case-control study. *BMJ* 2010;340:c3077.
  11. Elliott P, Shaddick G, Douglass M, de Hoogh K, Briggs DJ, Toledano MB. Adult cancers near high-voltage overhead power lines. *Epidemiology* 2013;24:184–90.
  12. Gao H, Aresu M, Vergnaud A-C *et al.* Personal radio use and cancer risks among 48,518 British police officers and staff from the Airwave Health Monitoring Study. *Br J Cancer* 2019;120:375–8.
  13. Piel FB, Fecht D, Hodgson S *et al.* Small-area methods for investigation of environment and health. *Int J Epidemiol* 2020;49. doi: 10.1093/ije/dyaa006.
  14. Hodgson S, Fecht D, Gulliver J *et al.* Availability, access, analysis and dissemination of small area data. *Int J Epidemiol* 2020;49(Suppl 1):i4–i14.
  15. Fecht D, Cockings S, Hodgson S, Piel FB, Martin D, Waller L. Advances in mapping population and demographic characteristics at small area levels. *Int J Epidemiol* 2020;49(Suppl 1):i15–i25.
  16. Blangiardo M, Boulieri A, Diggle P, Piel FB, Shaddick G, Elliott P. Advances in spatio-temporal methods for non-communicable disease surveillance. *Int J Epidemiol* 2020;49(Suppl 1):i26–i37.
  17. Fecht D, Garwood K, Butters O, Henderson J, Elliott P, Hansell A. Automation of cleaning and reconstructing residential address histories to assign environmental exposures in longitudinal studies. *Int J Epidemiol* 2020;49(Suppl 1):i49–i56.
  18. Piel FB, Parkes B, Hambly P *et al.* The rapid inquiry facility 4.0: an open access tool for environmental public health tracking. *Int J Epidemiol* 2020;49(Suppl 1):i38–i48.
  19. Toledano M, Shaddick G, de Hoogh C *et al.* Electric field and air ion exposures near high voltage overhead power lines and adult cancers: a case control study across England and Wales. *Int J Epidemiol* 2020;49(Suppl 1):i57–i66.