



ELSEVIER

Contents lists available at ScienceDirect

## Data in brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)



### Data Article

# Data to support small area health impact modelling of air pollution in the United Kingdom



Andrew Ibbetson<sup>a, \*</sup>, Phil Symonds<sup>b</sup>, Emma Hutchinson<sup>a</sup>

<sup>a</sup> London School of Hygiene and Tropical Medicine, London, UK

<sup>b</sup> Institute of Environmental Design and Engineering, UCL, London, UK

#### ARTICLE INFO

##### Article history:

Received 28 November 2019

Received in revised form 8 January 2020

Accepted 9 January 2020

Available online 15 January 2020

##### Keywords:

Microsimulation

Health impact modelling

Environmental risks

Deprivation

Air pollution

#### ABSTRACT

The data presented in this article were used to estimate the impacts of air pollution policies on population health and health inequalities within a spatial microsimulation model, MicroEnv [1]. They provide a basis for comparison with similar models and allow researchers to integrate additional model components without duplication of effort. Relative risk estimates for the association between air pollution and rates of ischaemic heart disease (IHD) incidence, IHD case fatality and all-cause mortality were taken from a review of the epidemiological literature and meta-analyses [2]. Modelled small area air pollution data (PM<sub>2.5</sub>) for Greater London, UK were obtained from an environmental consultancy. All other data were collected from open source Governmental or Non-Government Organisation (NGO) data repositories. These include all-cause mortality rates; IHD incidence, prevalence and mortality rates; general fertility rates; small area socio-economic deprivation data; and relative risk estimates for the association between deprivation and all-cause mortality.

© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

\* Corresponding author.

E-mail address: [andrew.ibbetson1@lshtm.ac.uk](mailto:andrew.ibbetson1@lshtm.ac.uk) (A. Ibbetson).

## Specifications Table

Subject	Public Health and Health Policy.
Specific subject area	Health impact modelling of environmental risk factors at small area level.
Type of data	Graphs, Figures and Tables.
How data were acquired	All data were acquired from open access internet databases, except the modelled air pollution data and the relative risk estimates for the association between air pollution and rates of IHD incidence, IHD case fatality and all-cause mortality. The former were provided by Ricardo Energy and Environment and the latter were obtained from a review of the academic literature and meta-analyses.
Data format	Raw and Analyzed.
Parameters for data collection	All data were obtained at the highest possible spatial resolution, Lower Layer Super Output Area (LSOA), where possible. Population and health data were collected by gender and at the highest possible age resolution (single year of age if available, 5-year age groupings otherwise).
Description of data collection	Data were collected by downloading the relevant file from an open access internet database or by extraction from a review of epidemiological literature conducted by the authors.
Data source location	Greater London, United Kingdom.
Data accessibility	Repository name: MicroEnv Data identification number: <a href="https://doi.org/10.17632/dxrp4b656h.1">10.17632/dxrp4b656h.1</a> Direct URL to data: <a href="https://doi.org/10.17632/dxrp4b656h.1">https://doi.org/10.17632/dxrp4b656h.1</a>
Related research article	Phil Symonds, Emma Hutchinson, Andrew Ibbetson, Jonathon Taylor, James Milner, Zaid Chalabi, Michael Davies, Paul Wilkinson. MicroEnv: A microsimulation model for quantifying the impacts of environmental policies on population health and health inequalities. Science of The Total Environment <a href="https://doi.org/10.1016/j.scitotenv.2019.134105">https://doi.org/10.1016/j.scitotenv.2019.134105</a>

**Value of the Data**

- These data allow the estimation of impacts of air pollution and other environmental policies on population health and health inequalities.
- Policy makers and researchers may benefit from the insights that analysis of these data can provide, within the constraints of uncertainty.
- These data provide a basis for comparison with similar models and allow researchers to integrate additional model components without duplication of effort.

**1. Data**

The data described in this article form the basis of a spatial microsimulation model, MicroEnv, which aims to investigate the impact of environmental policies on population health [1]. All data are Greater London, UK specific and are outlined in Table 1.

**1.1. Health data****1.1.1. All-cause mortality rates**

All-cause mortality data were obtained from the Office for National Statistics (ONS) [3]. The ONS provides both period and cohort projections by single year of age and sex for the UK. Although the model is able to take these projections into account, fixed (2016) values for mortality rates were used within MicroEnv. Fig. 1 shows all-cause mortality rates used in the model by age and sex for 2016.

**1.1.2. Ischaemic heart disease (IHD) incidence, prevalence and mortality rates**

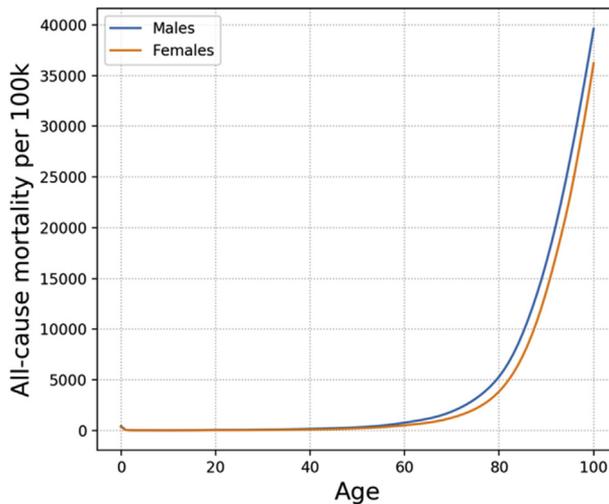
IHD incidence, prevalence and mortality data used within MicroEnv are obtained from the GBD Results Tool [4]. Data by 5-year age band and sex for the UK (2016) were used within our model. Whilst 2016 GBD data is used in the model, 2017 data is shown here for comparison to HSE data from 2017. GBD outputs are shown with the latest Health Survey for England (HSE) data [9] overlaid, the graphs in

**Table 1**

Data sources used in the microsimulation model.

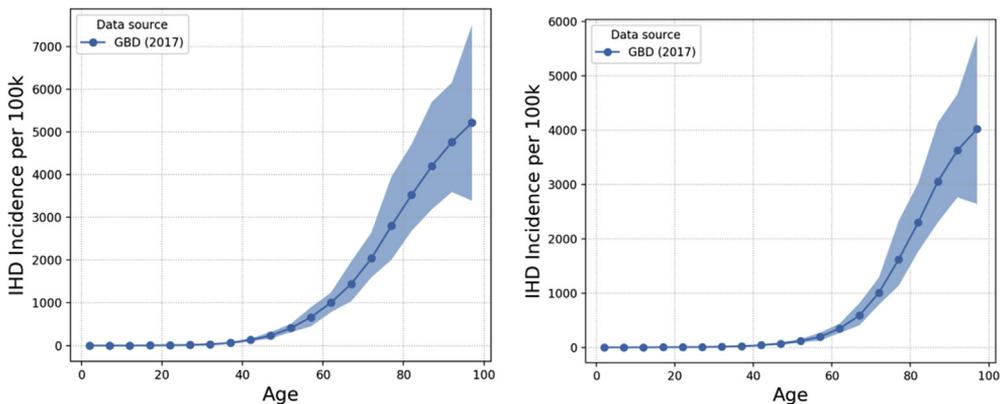
Data type	Year	Additional info	Reference
Mortality (all-cause)	2016	Period projections by year of age and gender	[3]
IHD mortality, incidence and prevalence	2016	By gender and 5-year age bands (UK)	[4]
Population	2015	Population by single year of age, gender and LSOA <sup>a</sup>	[5]
Socio-economic deprivation	2015	Decile of the Index of Multiple Deprivation (IMD) for each LSOA	[6]
Air pollution	2014	Annual averages of PM <sub>2.5</sub> at 1 × 1 grid (mapped to LSOA)	[7]
General fertility rates	2015	Number of live births per 1000 females aged 15–44 at local authority level. Applied to the LSOA-specific female population each year	[8]

<sup>a</sup> Lower Layer Super Output Area.

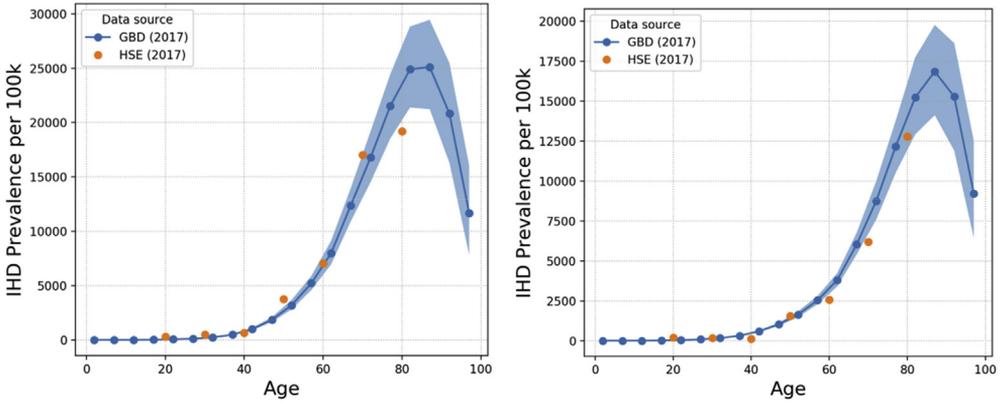


**Fig. 1.** UK all-cause mortality rates (per 100k) used in MicroEnv by single year of age and sex. Source: ONS.

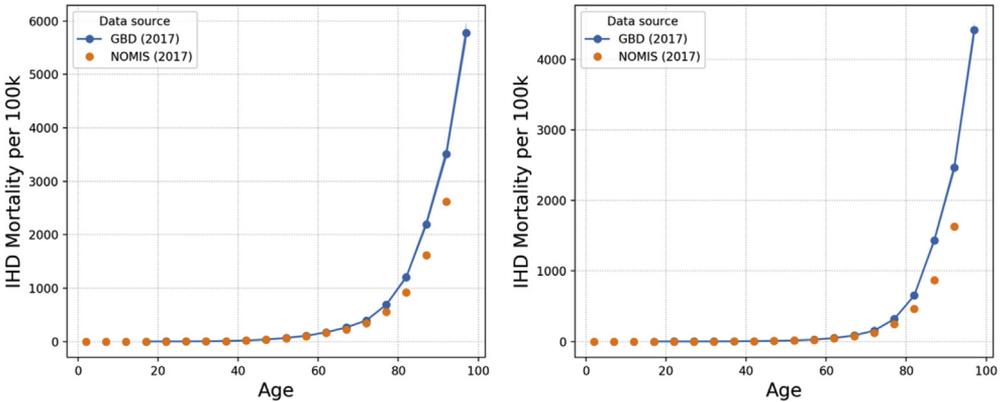
this paper show data for England (2017). Figs. 2–4 show incidence, prevalence and mortality data, respectively. For the incidence data, it is not possible to overlay HSE data, since HSE reports incidence in terms of hospital inpatient admissions rather than new cases of IHD.



**Fig. 2.** IHD incidence rates (per 100k) in England (2017) by age as output by the GBD Results Tool for males (left) and females (right). Error bands represent 95% confidence interval reported by GBD Results Tool.



**Fig. 3.** IHD prevalence rates (per 100k) in England (2017) by age as output by the GBD Results Tool compared to HSE data for males (left) and females (right). Error bands represent 95% confidence interval reported by GBD Results Tool.



**Fig. 4.** IHD mortality rates (per 100k) in England (2017) by age as output by the GBD Results Tool compared to nomis data from the ONS for males (left) and females (right).

1.1.3. Relative risks

Table 2 shows the relative risks used in the MicroEnv model and their source. These estimates modify the annual transition probabilities between disease states to account for the effects of air pollution and, separately, deprivation on health.

**Table 2**  
Relative risks used in the microsimulation model.

Relative risk	Coefficient used	Reference
IHD incidence	1.08 per 10 $\mu\text{g}/\text{m}^3$ ( $\text{PM}_{2.5}$ )	[2]
IHD case fatality	1.21 per 10 $\mu\text{g}/\text{m}^3$ ( $\text{PM}_{2.5}$ )	[2]
All-cause mortality	1.06 per 10 $\mu\text{g}/\text{m}^3$ ( $\text{PM}_{2.5}$ )	[10]
	1.7 between most and least deprived decile (males)	[11]
	1.5 between most and least deprived decile (females)	[11]

## 1.2. Population data

The population data used within MicroEnv to generate an estimate of the population of Greater London were from the ONS [5]. Data is available for the 2015 population by single year of age and by sex for each LSOA in Greater London. Fig. 5 maps the estimated population of each LSOA modelled.

## 1.3. Socio-economic deprivation

The socio-economic deprivation of individuals within each LSOA is defined using the 2015 decile of the Index of Multiple Deprivation (IMD). These are reported by the UK Department for Communities and Local Government (DCLG) [6]. Fig. 6 provides a map of IMD by LSOA for Greater London.

## 1.4. General fertility rates (GFRs)

The number of new births simulated in MicroEnv each year is calculated using local authority level General Fertility Rate (GFR) data (the number of live births per 1000 females aged 15–44 per year) [8]. Fig. 7 maps the GFRs by local authority as used within the model.

## 2. Experimental design, materials, and methods

The majority of data compiled in this article for use in MicroEnv are from open source Governmental or Non-Government Organisation (NGO) data repositories. These websites provide annual updates to their data. ONS population data is estimated yearly by applying the ratio of change in consecutive GP Patient Register data to the population found by the previous census [12]. ONS all-cause mortality rates are based on deaths registered in England and Wales. The GBD Results Tool estimates incidence and prevalence rates of IHD using a variety of data sources including surveillance, survey data and claims data – inpatient and outpatient visits [13]. While IHD mortality rates are estimated using vital registration and verbal autopsy data [14]. The GBD employ a Bayesian meta-regression tool (DISMOD-MR 2.1), which combines data from various countries in an integrated manor to provide smooth outputs.

Relative risks for the association between air pollution and rates of IHD incidence, IHD case fatality and all-cause mortality were taken from a recent literature review and meta-analyses [2]. Small area air pollution data for the annual average ambient PM<sub>2.5</sub> were provided by an environmental consultancy, Ricardo Energy and Environment. This was based on air pollution modelled at 1 × 1 km grid square resolution for the UK using the air dispersion model ADMS and emissions estimates from the UK National Atmospheric Emissions Inventory 2014 [15]. The modelled air pollution was linked to provide

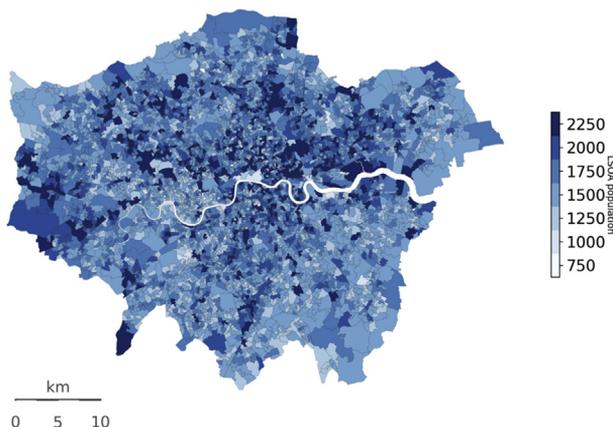
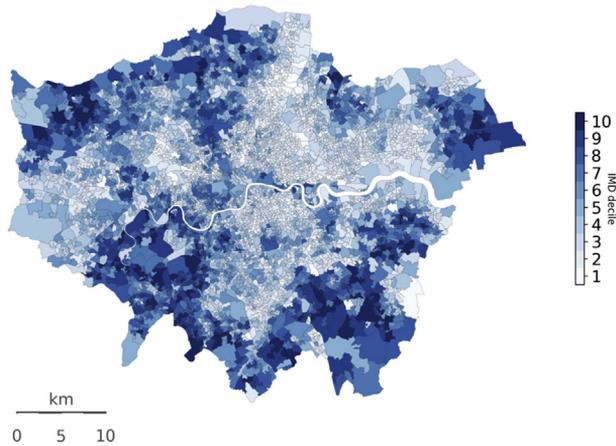
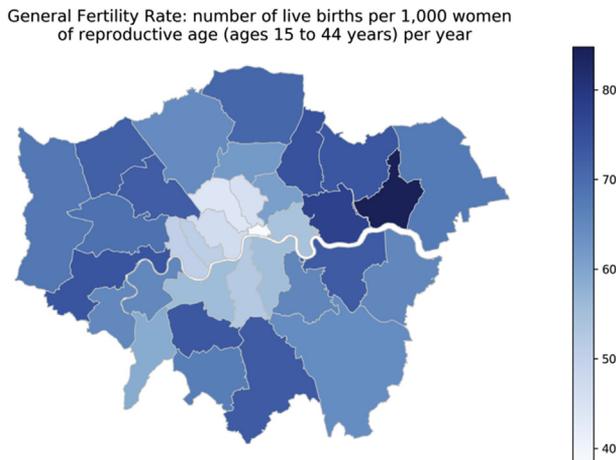


Fig. 5. Map of population (2015) by LSOA for Greater London. Source: ONS (?).



**Fig. 6.** Map of IMD (2015) by LSOA for Greater London. A lower deprivation decile represents a more deprived area. Source: DCLG.



**Fig. 7.** General Fertility Rate mapped at local authority level for Greater London. Source: ONS.

an average for each LSOA using GIS methods. Data analysis scripts to produce the graphs and maps in this article were written in Python V3.6 [16] using the geopandas, Basemap and matplotlib libraries. The shapely files used to generate the maps were obtained from the Greater London Authority, who produce GIS files with boundaries at various spatial resolutions [17].

## Acknowledgments

This work was supported by the Wellcome Trust 'Complex Urban Systems for Sustainability and Health' (CUSSH) project [205207/Z/16/Z and 209387/Z/17/Z]; the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) on Environmental Change and Health. The project is led by the London School of Hygiene and Tropical Medicine in partnership with Public Health England (PHE), and in collaboration with the University of Exeter, University College London, and the Met Office. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

The authors would like to thank the UK Committee on the Medical Effects of Air Pollutants (COMEAP) (sub-group on hazard identification and quantification of cardiovascular morbidity associated with long-term exposure to air pollution in the United Kingdom) for their valuable insight during the models' development. We also wish to thank Ricardo Energy and Environment for providing us with the modelled air pollution data for the UK.

## Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dib.2020.105148>.

## References

- [1] P. Symonds, E. Hutchinson, A. Ibbetson, J. Taylor, J. Milner, Z. Chalabi, M. Davies, P. Wilkinson, MicroEnv: a microsimulation model for quantifying the impacts of environmental policies on population health and health inequalities, *Sci. Total Environ.* 697 (2019) 134105, <https://doi.org/10.1016/j.scitotenv.2019.134105>. ISSN 0048-9697.
- [2] P. Wilkinson, P. Symonds, E. Hutchinson, A. Milojevic, J. Milner, B. Armstrong, A Model of Ischaemic Heart Disease in the UK that Suggests Reductions in Air Pollution Increase Overall Disease Prevalence, Joint Annual Meeting of the International Society of Exposure Science (ISES) and the International Society for Environmental Epidemiology (ISEE), 2018 (Ottawa, Canada).
- [3] ONS, Past and Projected Data from the Period and Cohort Life Tables, 2016-based, UK: 1981 to 2066 [WWW Document], 2017. Select Life tables, principal projection, United Kingdom 2016-based; UK (1951-2066); select Period, 2016, <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/lifetablesprincipalprojectionunitedkingdom>. (Accessed 3 July 2019).
- [4] Institute for Health Metrics and Evaluation, GHDx: GBD Results Tool [WWW Document], 2019. <http://ghdx.healthdata.org/gbd-results-tool>. (Accessed 3 July 2019).
- [5] ONS, Lower Layer Super Output Area Population Estimates [WWW Document], 2018. Select Small area population estimates, England and Wales: mid-2015; Mid-2015 Males and Mid-2015 Females, <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates>. (Accessed 3 July 2019).
- [6] Department for Communities and Local Government, English Indices of Deprivation 2015, 2015. Select File 1: index of multiple deprivation; IMD 2015, <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>. (Accessed 18 July 2019).
- [7] Ricardo Energy & Environment, Technical Report on UK Supplementary Assessment under the Air Quality Directive (2008/50/EC), the Air Quality Framework Directive (96/62/EC) and Fourth Daughter Directive (2004/107/EC) for 2015, 2017.
- [8] ONS, Births in England and Wales: Summary Tables [WWW Document], Office for National Statistics, 2018b. Select Births in England and Wales: 2017; Table 1: Summary of key birth statistics, 1838 to 2017, <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthsummarytables>. (Accessed 3 July 2019).
- [9] NHS Digital, Health Survey for England - Health, Social Care and Lifestyles, 2019. Select Health Survey for England, 2017: Cardiovascular diseases – tables; Table 2: Prevalence of ever having doctor-diagnosed IHD, stroke, IHD or stroke, by survey year, age and sex, <https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2017>. (Accessed 18 July 2019).
- [10] C.A. Pope III, R.T. Burnett, M.J. Thun, E.E. Calle, D. Krewski, K. Ito, G.D. Thurston, Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, *J. Am. Med. Assoc.* 287 (2002) 1132, <https://doi.org/10.1001/jama.287.9.1132>.
- [11] A. Baker, E. Romeri, C. Griffiths, Mortality by Deprivation and Cause of Death in England and Wales, Office for National Statistics, 2006.
- [12] ONS, Small Area Population Estimates QMI [WWW Document], 2018. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/smallareapopulationestimatesqmi>. (Accessed 24 September 2019).
- [13] Spencer L. James, et al., Appendix 1 of Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study, *The Lancet* 392 (10159) (2017) 1789–1858.
- [14] Gregory A. Roth, et al., Appendix 1 of Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study, *The Lancet* 392 (10159) (2017) 1736–1788.
- [15] NAEI, UK NAEI - National Atmospheric Emissions Inventory [WWW Document], 2014. <http://naei.beis.gov.uk/>. (Accessed 18 July 2019).
- [16] G. Van Rossum, F.L. Drake, *An Introduction to Python: for Python Version 3.2*, Network Theory Ltd, 2011.
- [17] GLA, Statistical GIS Boundary Files for London [WWW Document], 2019. <https://data.london.gov.uk/datasets/statistical-gis-boundary-files-london>. (Accessed 25 September 2019).