Paper:
Rodgers, S., Johnson, R., Dearden, L., Al Sallakh, M., Mavrogianni, A., Davies, G., Lake, I., Akbari, A., Carruthers, D., et. al. (2018). Creating individual level air pollution exposures in an anonymised data safe haven: a platform for evaluating impact on educational attainment. *International Journal of Population Data Science, 3*(1)
http://dx.doi.org/10.23889/ijpds.v3i1.412

International Journal of Population Data Science

Journal Website: www.ijpds.org

Creating individual level air pollution exposures in an anonymised data safe haven: a platform for evaluating impact on educational attainment

Mizen, A[1], Lyons, J[1], Doherty, R[2], Berridge, D[1], Wilkinson, P[3], Milojevic, A[3], Carruthers, D[4], Akbari, A[1], Lake, I[5], Davies, GA[6], Sallakh, MA[1], Mavrogianni, A[7], Dearden, L[8], Johnson, R[1], and Rodgers, SE[1,9*]

Submission History

Submitted:	11/09/2017
Accepted:	16/02/2018
Published:	21/08/2018

[1]Health Data Research UK Wales and Northern Ireland, Swansea University Medical School, Wales, UK
[2]School of GeoSciences, The University of Edinburgh, Edinburgh, United Kingdom
[3]Department of Social and Environmental Health Research, London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK
[4]Cambridge Environmental Research Consultants, Cambridge, United Kingdom
[5]School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK
[6]Asthma UK Centre for Applied Research, Swansea University Medical School, Singleton Park, Swansea, UK
[7]UCL Energy Institute, University College London, Gower Street, London
[8]The Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE
[9]Department of Public Health and Policy, University of Liverpool, Liverpool, UK

Abstract

Introduction

There is a lack of evidence on the adverse effects of air pollution on cognition for people with air quality-related health conditions. We propose that educational attainment, as a proxy for cognition, may increase with improved air quality. This study will explore whether asthma and seasonal allergic rhinitis, when exacerbated by acute exposure to air pollution, is associated with educational attainment.

Objective

To describe the preparation of individual and household-level linked environmental and health data for analysis within an anonymised safe haven. Also to introduce our statistical analysis plan for our study: COgnition, Respiratory Tract illness and Effects of eXposure (CORTEX).

Methods

We imported daily air pollution and aeroallergen data, and individual level education data into the SAIL databank, an anonymised safe haven for person-based records. We linked individual-level education, socioeconomic and health data to air quality data for home and school locations, creating tailored exposures for individuals across a city. We developed daily exposure data for all pupils in repeated cross sectional exam cohorts (2009-2015).

Conclusion

We have used the SAIL databank, an innovative, data safe haven to create individual-level exposures to air pollution and pollen for multiple daily home and school locations. The analysis platform will allow us to evaluate retrospectively the impact of air quality on attainment for multiple cross-sectional cohorts of pupils. Our methods will allow us to distinguish between the pollution impacts on educational attainment for pupils with and without respiratory health conditions. The results from this study will further our understanding of the effects of air quality and respiratory-related health conditions on cognition.

Keywords

data linkage, cognition, asthma, seasonal allergic rhinitis, air pollution

Highlights

1. This city-wide study includes longitudinal routinely-recorded educational attainment data for all pupils taking exams over seven years;
2. High spatial resolution air pollution data were linked within a privacy protected databank to obtain individual exposure at multiple daily locations;
3. This study will use health data linked at the individual level to explore associations between air pollution, related morbidity, and educational attainment.

*Corresponding Author:
Email Address: Sarah.Rodgers@Liverpool.ac.uk (SE Rodgers)

http://dx.doi.org/10.23889/ijpds.v3i1.436

Introduction

Studies have investigated either the impact of air pollution on respiratory health, or air pollution on cognition (1–3). To our knowledge there are no existing studies investigating the combined effects of air pollution and respiratory health conditions on cognition. Health effects on cognitive behaviour can develop over a lifetime, but educational attainment is one measure of cognition that may have long-lasting effects into higher education choices and labour markets (4). More evidence is needed to quantify the impact of pollution on educational attainment for people from all backgrounds, particularly those who are sensitive to poor air quality. Only a small proportion of health is determined by genetic factors, with many health issues determined by social and environmental factors (5). It is particularly important that poor quality environments are not a burden for those who are most deprived.

Bringing together detailed data on air quality, health status and cognition for a large population is complex and to date, studies have been unable to model the complex relationship conclusively. Information governance also restricts the availability of appropriate data. This may be one reason why the causal role of education on ill-health remains underappreciated. Studies often conflate education with socioeconomic status (SES), or assume that education is only a proxy measure for SES (6). Causality between educational attainment and individual health has long been debated but there has yet to be a systematic assessment of the education effect on health across a full set of methodologically sophisticated studies, with controls for wealth and other dimensions of socioeconomic status (6). Using individual-level linked data we aim to provide insights into these explanatory link(s), enabling us to avoid overly reductionist approaches (7).

Air pollution in the UK is estimated to cost around £15 billion annually from human health effects alone (8). The main air pollutants of concern for human health are Particulate Matter (PM), Ozone (O_3) and Nitrogen dioxide (NO_2) (9). Exposure to air pollution, particularly PM, is thought to have both acute and irreversible effects on childhood cognition, as well as a damaging lifetime effect. There is a long-term risk to children exposed to poor air quality (O_3 and PM) during crucial periods of brain development (10). In the shorter term, cognitive ability is thought to be influenced in early childhood by exposure to air pollutants including O_3, PM, Nitrogen Oxides (11). Cognition may also be affected at older ages due to fine PM penetrating the lungs and inhibiting oxygen flow into the bloodstream and to the brain (12). Using distance from residence to road as a proxy for air pollution, increased incidence risk of dementia was found for people living less than 50 metres from a major traffic road HR=1.07 (95% CI 1.06–1.08) (13). A review concluded that even short-term exposure to air pollution, namely fine particulate matter $\leq$ 2.5 µg ($PM_{2.5}$) from traffic-related air pollution, may negatively impact cognitive ability (12).

Studies evaluating the relationship between air pollution and cognitive development are generally restricted to small samples of children in tailored prospective cohorts for whom there are detailed measures of cognition, for example the McCarthy Scale of Children's Abilities, for children aged 2-8 years old (14). Detailed cognitive test results for 210 children were analysed with Nitrogen Dioxide (NO_2) exposure using land use regression models (15). This study finding suggest that traffic-related air pollution (NO_2), whose levels vary greatly over short distances, may have an adverse effect on neurodevelopment (15). Recent studies conducted with participating pupils attending schools in Barcelona found that exposure to traffic-related air pollution (NO_2 and $PM_{2.5}$ and ultrafine particles (10-700nm)) may have potentially harmful effects on cognitive development causing inattention, forgetfulness, disorganization, and careless errors (16–18). These skills are essential for learning and exam performance. Studies in Israel using educational attainment to measure cognitive development found high levels of fine particulate matter pollution were negatively associated with exam results (19,20).

Existing literature on air quality-related diseases and educational attainment identifies associations between asthma or asthma symptoms and lower educational attainment (20,21), though the strongest association was with parental education (21). Another study showed there was no significant effect of asthma on educational attainment (22). However, these studies did not use air quality when examining the cause of asthma exacerbations. The role of environmental triggers for asthma has not yet been adequately explored in terms of potential impact on educational attainment. There are numerous complex interactions between outdoor pollutants, pollen aeroallergens and genetic susceptibility to asthma. There is increasing evidence of an interaction between air pollution and pollen. In particular, O_3 may enhance the allergenicity of pollen and impact cognitive development (23,24). Both air pollutants and pollen can cause exacerbations in asthma and there is growing evidence to show that air pollutants may increase the acute effects of allergens on health outcomes, as well as on cognition (25–27).

The rationale for our present study was to address the evidence gap on the environmental drivers in relation to cognition by combining datasets on air pollution, pollen, asthma, seasonal allergic rhinitis (SAR), and educational attainment. The small sample sizes in previous studies prompted us to design a city-wide study to include a larger number of children. Our study: COgnition, Respiratory Tract illness and Effects of eXposure (CORTEX) aims to explore the impact of air quality on educational attainment, while considering air quality related morbidity. This paper reports our data preparation methods according to the RECORD statement (28). It also describes our statistical analysis plan to evaluate the impact of air quality on educational attainment data for pupils aged 15-16 years old.

Methods

We designed a study with a large number of children with a wide variety of pollution exposures by using routinely collected educational attainment data as a proxy for cognition. We modelled pollution estimates for each home and school and obtained individual-level education data. Here we describe briefly our data linkage process, each source dataset, and the preparation of the dataset for analysis.

CORTEX is a repeated cross-sectional retrospective data linkage study with three principal research questions:

1. Is exam performance, negatively associated with asthma or SAR status?;
2. Is exam performance, negatively associated with pollution levels averaged at the school and home location over the term prior to the examination? And, if so;
3. Is the association explained by markers of respiratory health, including doctor-diagnosed respiratory illness (asthma)?;

Setting

Cardiff is the largest city in Wales, United Kingdom, with a population of more than 346,000 people (29). Within the city, air pollution levels and ill health vary spatially and temporally. Area-based health inequalities for Cardiff are growing, with life expectancy for men in the most deprived areas estimated to be 10 years shorter than for those in the most affluent areas (30). Inequalities in environmental exposure contribute to these differences from an early age (31,32). Exposure to four pollutants will be included in the analysis and exposure to pollen will be included as a confounder. The setting of Cardiff provides a sufficiently large population to investigate the natural experiment of exposure to air pollution for sub-populations of pupils with and without asthma and SAR.

Data Linkage

We used the Secure Anonymised Information Linkage (SAIL) databank, an anonymised safe haven, to create consistent data linkages for an individual among their environment, health and educational outcomes (Figure 1). Individuals within the SAIL databank have previously been successfully matched to the demographic data ($>$ 99.9%) (33,34).

The SAIL platform was designed to overcome data sharing issues and facilitate the evaluation of individual and household-level interventions and natural experiments to support public health decision-making (35). A globally unique anonymisation process within this database allows retrospective linkage among education, health outcomes and environmental exposures (36,37). The SAIL databank includes a number of core datasets where historical repeated measures of health and education data are routinely collected and available. The catalogue of core datasets includes: an anonymised but consistently linkable population register; mortality, inpatient, outpatient, and emergency department data across Wales; primary care data for 75% of GP practices in Wales; and educational attainment data for all schools in Wales.

Data in the SAIL databank are anonymised using a 'split-file' process. For this project the address identifiers were separated from the air pollution modelled estimates (Figure 1). The address component was sent to our trusted third party (TTP), to be anonymised and encrypted. Our TTP assigned a Residential Anonymous Linking Field (RALF) to each point location that referred to a place of residence or school location. A similar process operates for person-level data, with our TTP assigning an Anonymous Linking Field (ALF) to each individual record in the education dataset, replacing the unique pupil identifier with an ALF. Our detailed anonymisation methodology is reported elsewhere (38,39).

The Welsh Demographic Service dataset (WDS) contains historic addresses. We used residential and school level linkages, to take advantage of the high spatial resolution air pollution data. This meant that we were able to allocate accurate exposures to individuals. We linked daily air pollution estimates to the home address where the pupil was registered on the 1st of June in their exam year.

Data Sources

Three data sources were already available in the SAIL databank, and two additional datasets were obtained for this study (Table 1). We discuss the data linkage results of the air quality datasets in the results section.

General practice (WLGP)

The Welsh Longitudinal General Practice (WLGP) dataset contains individual-level health data including Read codes for all diagnoses, symptoms and treatments recorded for each person. A Read code is a coded thesaurus of clinical terms. Read codes have been used in the NHS since 1985 and provide a standardised vocabulary for clinicians to record patient results across primary and secondary care (40). Researchers extracted data to define individuals ever diagnosed with, or treated within the last year for, asthma or SAR (Appendix A). These data were available within the SAIL databank for the majority of practices in Cardiff (about 80% at the time of writing). The definition of who is considered an asthma patient varies greatly in the literature (41). These definitions are currently being refined to define patients in general practice with active asthma and SAR (42).

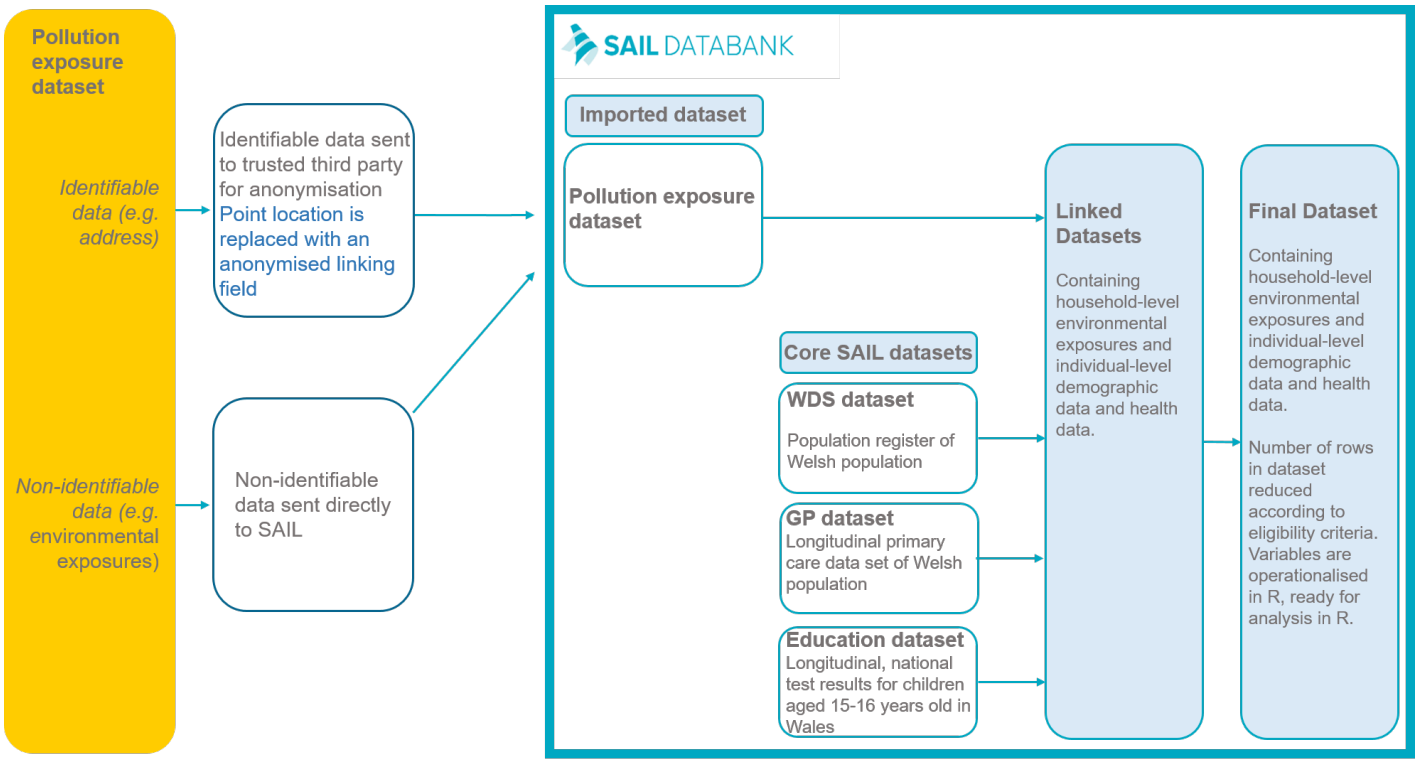
Welsh Demographic Service dataset (WDS)

The WDS is a population register that includes home addresses provided by patients attending primary care, a service that is free at the point of care in the UK. This dataset also allows researchers working in the SAIL databank to distinguish when an individual moves home. This is particularly important for exposure studies.

Education

Education data are held in the National Pupil Dataset (NPD) for Wales, which the Welsh Government provided to SAIL. This dataset contains exam scores and other education data relevant for each pupil in an anonymised format. Exam scores for the General Certificate of Secondary Education (GCSE) were used by the Welsh government to calculate a continuous measure of attainment as a proxy for cognition. Pupils are examined for their GCSE at 15 to 16 years old in their final year of mandatory schooling. Exam dates between 2009 and 2015 in the period 8th May to 29th June. The dataset included a number of school-level variables used to control for school characteristics.

Figure 1: Data linkage schematic illustrating the split file anonymisation process of the environmental exposures dataset, and the datasets linked and joined during CORTEX



Air pollution

Air pollutants $PM_{2.5}$ and PM_{10}, NO_2 and Ozone (O_3), were estimated for all home and school locations in Cardiff from 2006 to 2015 using the high-resolution Atmospheric Dispersion Modelling System (ADMS-Urban model) (43,44).

ADMS-Urban is a Gaussian-type plume model that simulates the dispersion of emissions from an urban area for each hour of the day. The main industrial and road traffic air pollution sources are modelled explicitly. This means that they are modelled to a very high spatial resolution along the road network or at the industrial site. The dispersion of pollutants in urban areas is also accounted for. Other smaller sources, including domestic and commercial emissions, are represented on a 1km grid, whilst pollutants transported into the urban area is treated as a spatially uniform but hourly varying background trend. Where concentration gradients are lowest, the spatial resolution is a minimum of 50 x 50 metres; close to main roads the resolution can be as high as a few metres.

Air pollution for all Cardiff locations were calculated. Only after the data were anonymised by our TTP and combined with the health and education data, could homes be selected for the relevant anonymised pupil cohort. Therefore, 157,361 school and home locations across Cardiff were anonymised and unique household level linkage fields were assigned to each location. Pollution estimates at the household and school locations were combined to create individual-level air pollution exposure, and subsequently joined to their health data.

Hourly air pollution data were summarised into daily time periods to enable extraction of school pollution estimates (9am - 3pm) and home pollution estimates (5pm – 9am) per pollution types $PM_{2.5}$ and PM_{10}, NO_2 and Ozone (O_3) and day of year. We also included a 3-5pm period for home or school in sensitivity analyses to assess the effect of this "uncertain location" period when pupils may have been at after school clubs. These were combined into tailored pollution exposures for each pupil.

Pollen (aeroallergen)

Data were obtained from the European Allergy Network (2006-2010) and from the UK Met Office (2011-2015). The decade (2006-2015) time series of daily pollen data were not spatially resolved for Cardiff; i.e. the pollen values did not vary by location. Data were summarised into annual total measures for several exam-related time periods.

The results of combining the two environmental datasets with the core SAIL datasets are discussed below.

Data Preparation

We joined linked datasets to individual-level health data within SAIL, and operationalised variables in preparation for analysis. We removed individuals who did not live at a Cardiff address on the census date or who moved between 1st April and their last exam.

Quantitative Variables

The educational outcome, environmental exposures and confounders are defined in Table 2. Educational attainment was assessed using the Capped Points Plus score (CPS). This is a

Table 1: Provincial data available for each indicator

Dataset name	Data Source	Derived Variables	Coverage
Welsh Longitudinal General Practice (WLGP) dataset*	Primary care records recorded by GP	Asthma and SAR treatments	~75% of practices in Wales provide data to SAIL
Welsh Demographic Service (WDS) dataset*	NHS Wales Informatics Service	Age, gender, week of birth, multiple move in and out of home dates	Total population of Wales who are registered with a General Practitioner, free at the point of service in the UK.
National Pupil Dataset*	Welsh Government	Capped points plus score, Gender, Free School Meal Eligibility, Special Educational Needs, Attendance, Percentage of Free School Meal Uptake, Number of Job Vacancies	All KS4 pupils in Wales between 2009-2015
ADMS-Urban**	Cambridge Environmental Research Consultants (CERC)	Nitrogen Dioxide (NO_2), Ozone (O_3), Particulate Matter ($PM_{2.5}$,PM_{10}), Nitrogen Oxides (NO_x)	Modelled pollution measurements for each pollutant, for four time periods per day. Mean, max and average values for 157,000 home and school locations in Cardiff
Pollen**	European Allergy Network (2006-2010) / Met Office (2011-2015)	Monthly total pollen count for tree pollen, grass and weed pollen	Daily pollen measurement for Cardiff between 2006-2015

*core SAIL dataset
** new dataset imported in to SAIL

continuous measure of attainment, calculated from grades for a pupil's best eight subjects which must include Mathematics and either English Language or Welsh as a first language. We standardised the CPS using z-scores.

Education is often used as a proxy for socio-economic status (SES) but education needs to be considered separately to understand its unique effects. Therefore, neighbourhood, school and household measures of SES will be included as confounders. Free School Meal Eligibility is a deprivation metric that we will use as a proxy for SES at both the household level (as a binary variable) and school level (as a percentage of students). The exposure variables were calculated as a weighted average for the pre and during exam periods.

For air pollution and pollen, we extracted total levels each year to correspond with the education data for two different time periods: i) Pre-examination period from the start of the summer teaching term (average start date 1st April) to the start of the exam period, and ii) during the exam period. It was anticipated that including the exam period would capture short-term exposure to pollen. The pre-examination period was selected so that it coincided with the start of the UK pollen season in late March. The examination period was defined using the Welsh Joint Education Committee (WJEC) examination timetable for the GCSE June series.

Results

Data linkage

A trained researcher had full access to the air pollution, pollen, WLGP and education datasets that were anonymised and imported into SAIL. Researchers were given restricted access to the demographic data, WDS, based on project geographic and temporal criteria.

Once in the SAIL databank, outdoor air pollution estimates were extracted at both the home and school location for each pupil. The pollution dataset contained 369 columns, 472,083 rows per year with one column per location, pollutant type, pollutant measurement, daily time-period, and day of year. The dataset was transformed to create a single date column to produce a five-column, 3,446,205,900-row matrix per year dataset. Pollution exposures were calculated for the revision and examination period per pupil, adjusting for weekends, school and bank holidays. We allowed the location to vary between 3pm-5pm on school days when pupil location was uncertain such as when traveling between school and home or engaged in after school activities.

We cleaned the dataset to remove poor quality linkages between the education dataset and the WDS (n=23). We defined poor quality matches as individuals who had fuzzy matching probability less than 0.9, or with no matching to WDS. The final study population only included pupils with high quality matches of demographic data to the WDS. We identified high quality matches as: matching surname, first

Table 2: Provincial data available for each indicator

Variable	Definition
Outcome	
Capped Point Score	A score calculated from the best eight GCSEs or equivalent, must include English or Welsh first language and Mathematics. Standardised using z-scores.
Exposures	
$PM_{2.5}$	Weighted average for the exposure periods (pre-exam and exam)
PM_{10}	Weighted average for the exposure periods (pre-exam and exam)
Ozone (O_3)	Weighted average for the exposure periods (pre-exam and exam)
Treated for asthma or not	Children treated for asthma identified from Read codes in GP records (Supplementary Appendix A details read codes used)
Treated for SAR or not	Children treated for SAR identified from Read codes in GP records (Supplementary Appendix A details read codes used)
Confounders	
Pollen	The average pollen count for the two exposure periods (pre-exam and exam) were calculated as the average maximum daily pollen value
Sex	Sex of pupil (Male/Female)
Free School Meal Eligibility	Pupil eligibility for a free school meal on the day of the Pupil Level Annual School Census (PLASC). This is a household indicator of socio-economic status
Special Educational Needs	Any special educational needs of a pupil on the day of the Pupil Level Annual School Census
Attendance	Number of half days that a pupil attended school for the academic year
Percentage of Free School Meal Uptake	The number of pupils taking a free school meal on the day of PLASC. This is a school level indicator of socio-economic status.
Number of Job Vacancies	Number of teaching vacancies in the school during the academic year
Welsh Index of Multiple Deprivation (WIMD)	Official measure of relative deprivation for small areas in Wales. This is a neighbourhood indicator of deprivation that is included in the WDS.

name, postcode, date of birth and gender; or fuzzy matching probability greater than or equal to 0.9. The matching process is documented in detail elsewhere (34).

Educational and health outcomes in the final cohort were representative of the target population (GCSE pupils in Cardiff). The cohort with air pollution exposures were successfully linked to their corresponding health and education data for 95% of pupils.

Participants

Eligible participants sat their GCSE Maths and English examinations between 2009 and 2015 in Cardiff. We created the study population based on agreed eligibility criteria (Figure 2).

Statistical Analysis Plan

Multilevel linear regression will be used to account for individuals who are nested within an exam year group, nested within schools. This will cluster pupils within a yearly cohort, from 2009-2015. Cohort and individual-level confounders will be controlled for. The model will therefore take into account differences between exam standards year-to-year for pupils in different exam cohorts. We will also complete an additional sensitivity analysis of the models by changing the after school hours "uncertain location" pollutants from school to home exposure locations.

Using the spatially resolved air pollution data, we will investigate whether different levels of outdoor air pollution are associated with differences in CPS. We will also explore whether levels of outdoor air pollution at school explain the variation in mean CPS between schools. Using the time series of spatially unresolved pollen data we will investigate whether different levels of pollen between cohorts explain the variation in mean CPS between cohorts taking exams in different years.

Clustering within household will be taken into account at the cohort level because siblings will appear in different cohorts. Twins will cluster within a household, however the number of twins within each cohort was so small that we decided that it was not necessary to account for clustering within households.

It is possible that taking over-the-counter antihistamines could influence examination outcomes, however it is not possible to account for in the study. Furthermore, dust mites are a common trigger of year-round allergies and asthma. However, data were not available to include this as a confounder.

Discussion

This paper has demonstrated that it is possible to link high spatial resolution data for individuals to fill the important evidence gap. Analysis of this linked dataset will allow researchers to explore the role of environmental triggers for pupils with and without asthma in terms of potential impact on educational attainment as a proxy for cognition. The SAIL databank was an appropriate secure data linkage platform for both health and non-health datasets and is world leading in its capabilities for household level data linkage (33). Multiple yearly cohorts for the pupil population of a city with health and education data were linked to high spatial and temporal resolution air pollution data allowing novel dual home and school exposures to be allocated retrospectively.

We have created multiple cross sectional cohorts in this initial study; requiring the availability of educational qualifications in previous years to create a longitudinal dataset would have reduced the number of pupils in each exam cohort resulting in underpowered analyses. Modelled pollution data are not available routinely for all urban areas, however, we anticipate that future CORTEX study results will prove these to be valuable, encouraging additional urban area modelling and increasing the initial sample size to make a viable longitudinal study. This expansion may also allow us to include data on emergency hospital admissions for respiratory conditions, which occur too infrequently in the current single city study sample.

Exploring the interaction of air pollution with pollen is novel and we anticipate the investigation of interactions between these different aspects of air quality with attainment for pupils with and without GP diagnosed asthma and/or SAR will be particularly informative. The results generated by CORTEX will demonstrate the feasibility and need to model air pollution at a high spatial resolution on a national scale.

A limitation of the routinely collected clinical data is that only those with a GP diagnosis or prescription will be captured. Allergic rhinitis may be self-diagnosed in up to 50% of the population, and over-the-counter medications are commonly used, although this is mitigated by the availability of free prescriptions in Wales since the 1st April 2007. However, those with more severe disease, and higher susceptibility to pollen and pollution, are more likely to be identified. Additionally, potentially confounding factors such as parental education may not be available routinely.

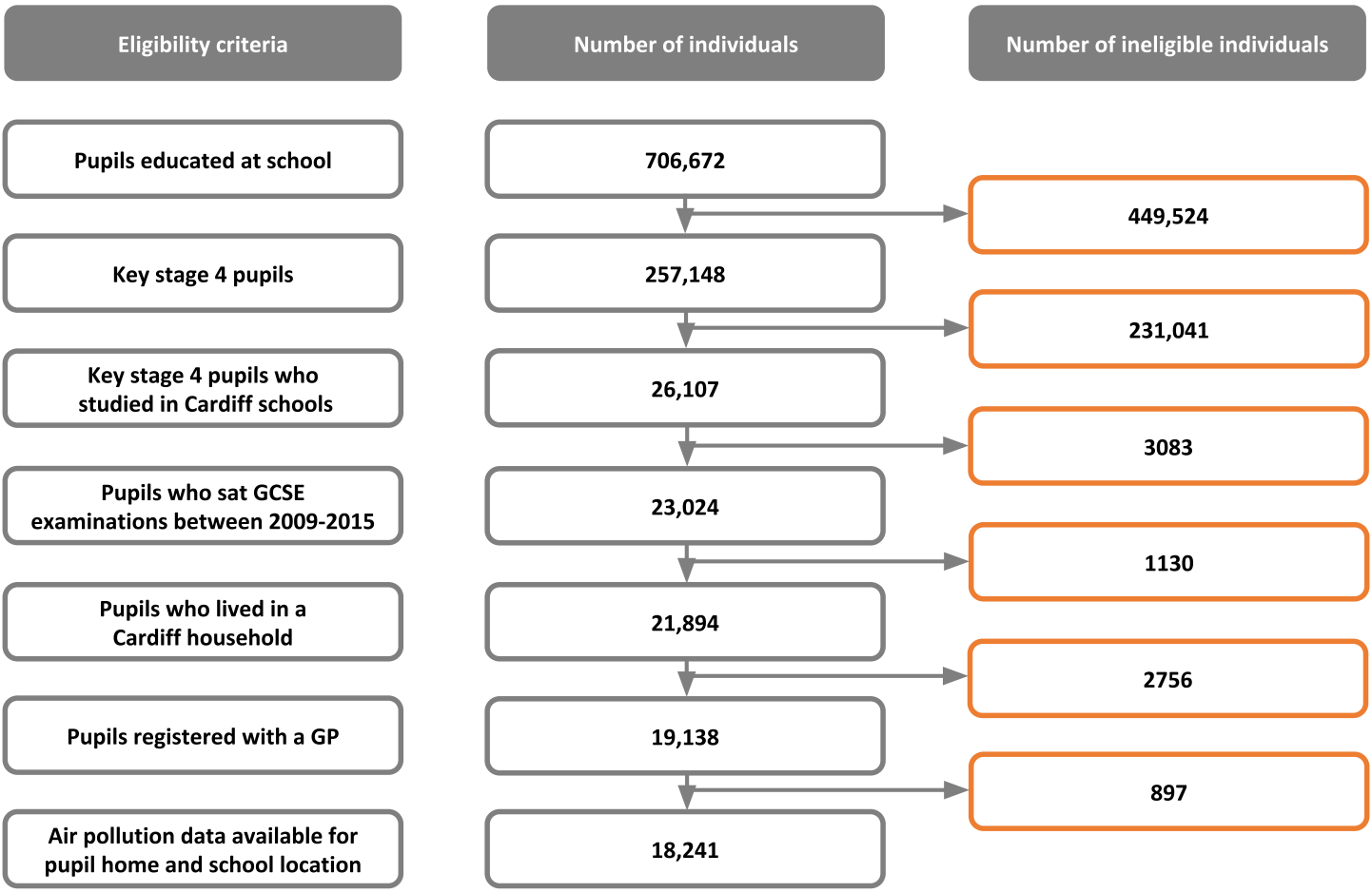
We have not thoroughly investigated the matching quality of our linkage method, like Harron and colleagues have (45), as this validation work is still to be undertaken. However, work undertaken by the Office for National Statistics (ONS) suggests that a large proportion of the population are likely to be correctly matched (46). Furthermore, children are likely to be registered to the correct GP as they have to have vaccinations, health checks and are usually taken to seek medical help when they are ill.

This study design shares some similarities to a recent Norwegian study that used routine education data linked to environmental data, however, they used pollen only (4). Our study has several new advances: we included health data to explore the direct and indirect effects of air quality related morbidity on attainment; we have used high spatial resolution data to enable us to separate the effects of home and school air pollution. In addition, our use of routine educational data means that it will not be necessary to ask pupils to complete tests, thus, our study is not limited to an actively participating small sample. We have used assessments of educational attainment retrospectively and on a city-wide scale, allocating daily exposures combined from both home and school locations.

Conclusion

We know little about the large-scale impact of air quality on cognition, particularly the interactive effects of air pollution and pollen, because studies have generally used small samples.

Figure 2: Eligibility criteria used to extract study population from raw linked dataset



Furthermore, the short-term effects of pollution on cognition are not well documented. Data linkage systems using routinely collected administrative data have opened up novel opportunities to analyse all pupils sitting exams in multiple years across an entire city. We hope that using individual-level multi-location daily exposure assessment will help to clarify the role of traffic and prevent potential community-level confounding (12).

Knowledge of associations between air pollution and aeroallergens on cognition could contribute to evidence-based policy-making. This feasibility study will demonstrate whether larger scale studies are possible, the results of which could have far-reaching policy implications. This may include raising awareness of the impact of pollution on cognition to a sufficient level to provide the impetus needed for government to further regulate air quality to benefit the UK population. We anticipate that this feasibility study will contribute to the evidence base which can inform policies that encourage stricter regulation on traffic flow near schools and locating new school sites in areas of less congestion and a better air quality could be considered. It is important to consider that lower educational attainment may lead to lower health literacy, which is associated with poorer adherence with asthma medication and thus poorer outcomes (18). The greatest benefit is likely for those vulnerable to air pollution and aeroallergens, including school children sitting their final examinations.

Declarations

Ethics approval and consent to participate

CORTEX uses anonymised, individual-level data held within the privacy-protected SAIL databank. CORTEX was approved by an independent Information Governance Review Panel (IGRP), which reviews all project proposals intending to use the SAIL databank. Due to the use of anonymised data, SAIL databank projects are exempt from undergoing a full ethics review.

Consent for publication

Not applicable.

Availability of data and material

The datasets generated and/or analysed during the current study are held in the SAIL databank repository, www.saildatabank.com, and may be available following project review by an independent Information Governance Review Panel.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by Chief Scientist Office of Scotland, the Medical Research Council and the Natural Environment Research Council (R8/H12/83/NE/P010660/1).

We acknowledge the support from The Farr Institute @ CIPHER : MR/K006525/1, which is supported by a 10-funder consortium: Arthritis Research UK, the British Heart Foundation, Cancer Research UK, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, the Medical Research Council, the National Institute of Health Research, the National Institute for Social Care and Health Research (Welsh Assembly Government), the Chief Scientist Office (Scottish Government Health Directorates), the Wellcome Trust, (MRC Grant No: MR/K006525/1)

Authors' contributions

All the authors have contributed to the CORTEX study design and methods. Amy Mizen led the writing of this manuscript. All other authors have been involved in revising it critically for important intellectual content and have approved the final version.

References

1. Guxens M, Sunyer J. A review of epidemiological studies on neuropsychological effects of air pollution. Swiss Med Wkly. 2012 Jan;141:w13322.
2. Clifford A, Lang L, Chen R, Anstey KJ, Seaton A. Exposure to air pollution and cognitive functioning across the life course - A systematic literature review. Environ Res [Internet]. 2016;147:383–98. Available from: https://doi.org/10.1016/j.envres.2016.01.018
3. Xu X, Ha SU, Basnet R. A Review of Epidemiological Research on Adverse Neurological Effects of Exposure to Ambient Air Pollution. Front Public Heal [Internet]. 2016;4(August). Available from: https://doi.org/10.3389/fpubh.2016.00157
4. Bensnes SS. You sneeze, you lose The impact of pollen exposure on cognitive performance during high-stakes high school exams. J Health Econ [Internet]. 2016;49:1–13. Available from: https://doi.org/10.1016/j.jhealeco.2016.05.005
5. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539–42.
6. Baker DP, Leon J, Smith Greenaway EG, Collins J, Movit M. The Education Effect on Population Health: A Reassessment. Popul Dev Rev. 2011 Jun;37(2):307–32. https://doi.org/10.1111/j.1728-4457.2011.00412.x
7. Gatrell AC. Complexity theory and geographies of health: A critical assessment. Soc Sci Med. 2005;60(12):2661–71. https://doi.org/10.1016/j.socscimed.2004.11.002
8. Department for Environment Food and Rural Affairs (DEFRA). Air Pollution: Action in a Changing Climate. Environment [Internet]. 2010;(March):24. Available from: http://www.defra.gov.uk/publications/files/pb13378-air-pollution.pdf

9. Guerreiro C, Gonzalez Ortiz A, de Leeuw F, Viana M, Horalek J. Air quality in Europe — 2016 report [Internet]. 2016. Available from: https://www.eea.europa.eu/publications/air-quality-in-europe-2016

10. Calderón-Garcidueñas L, Reed W, Maronpot RR, Henríquez-Roldán C, Delgado-Chavez R, Calderón-Garcidueñas A, et al. Brain inflammation and Alzheimer's-like pathology in individuals exposed to severe air pollution. Toxicol Pathol. 2004 Dec;32(6):650–8. https://doi.org/10.1080/01926230490520232

11. Currie J, Zivin JG, Mullins J, Neidell M. What Do We Know About Short- and Long-Term Effects of Early-Life Exposure to Pollution? Annu Rev Resour Econ. 2014;6(1):217–47.

12. Götschi T, Heinrich J, Sunyer J, Künzli N. Long-Term Effects of Ambient Air Pollution on Lung Function. Epidemiology [Internet]. 2008;19(5):690–701. Available from: https://doi.org/10.1097/EDE.0b013e318181650f

13. Chen H, Kwong JC, Copes R, Tu K, Villeneuve PJ, van Donkelaar A, et al. Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study. Lancet. 2017;389(10070):718–26. https://doi.org/10.1016/S0140-6736(16)32399-6

14. Goldman J, Stein CL, Guerry S. Psychological Methods of Child Assessment. Psychology Press; 1983. 416 p.

15. Freire C, Ramos R, Puertas R, Lopez-Espinosa M-J, Julvez J, Aguilera I, et al. Association of traffic-related air pollution with cognitive development in children. J Epidemiol Community Health. 2010 Mar;64(3):223–8. https://doi.org/10.1136/jech.2008.084574

16. Sunyer J, Suades-González E, García-Esteban R, Rivas I, Pujol J, Alvarez-Pedrerol M, et al. Traffic-related Air Pollution and Attention in Primary School Children. Epidemiology [Internet]. 2017;28(2):181–9. Available from: https://doi.org/10.1097/EDE.0000000000000603

17. Sunyer J, Esnaola M, Alvarez-Pedrerol M, Forns J, Rivas I, López-Vicente M, et al. Association between Traffic-Related Air Pollution in Schools and Cognitive Development in Primary School Children: A Prospective Cohort Study. PLOS Med. 2015 Mar;12(3):e1001792. https://doi.org/10.1371/journal.pmed.1001792

18. Apter AJ, Wan F, Reisine S, Bender B, Rand C, Bogen DK, et al. The association of health literacy with adherence and outcomes in moderate-severe asthma. J Allergy Clin Immunol. 2013 Aug;132(2):321–7. https://doi.org/10.1016/j.jaci.2013.02.014

19. Lavy V, Roth S. The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation. Natl Bur Econ Res. 2014;1–40.

20. Champaloux SW, Young DR. Childhood Chronic Health Conditions and Educational Attainment: A Social Ecological Approach. J Adolesc Heal. 2015 Jan;56(1):98–105. https://doi.org/10.1016/j.jadohealth.2014.07.016

21. Ruijsbroek A, Wijga AH, Gehring U, Kerkhof M, Droomers M. School Performance: A Matter of Health or Socio-Economic Background? Findings from the PIAMA Birth Cohort Study. PLoS One. 2015 Aug;10(8):e0134780. https://doi.org/10.1371/journal.pone.0134780

22. Sturdy P, Bremner S, Harper G, Mayhew L, Eldridge S, Eversley J, et al. Impact of Asthma on Educational Attainment in a Socioeconomically Deprived Population: A Study Linking Health, Education and Social Care Datasets. PLoS One. 2012;7(11):1–8. https://doi.org/10.1371/journal.pone.0043977

23. Beck I, Jochner S, Gilles S, McIntyre M, Buters JTM, Schmidt-Weber C, et al. High environmental ozone levels lead to enhanced allergenicity of birch pollen. PLoS One. 2013;8(11). https://doi.org/10.1371/journal.pone.0080147

24. Pasqualini S, Tedeschini E, Frenguelli G, Wopfner N, Ferreira F, D'Amato G, et al. Ozone affects pollen viability and NAD(P)H oxidase release from Ambrosia artemisiifolia pollen. Environ Pollut. 2011 Oct;159(10):2823–30. https://doi.org/10.1016/j.envpol.2011.05.003

25. Brandt EB, Biagini Myers JM, Acciani TH, Ryan PH, Sivaprasad U, Ruff B, et al. Exposure to allergen and diesel exhaust particles potentiates secondary allergen-specific memory responses, promoting asthma susceptibility. J Allergy Clin Immunol. 2015 Aug;136(2):295–303.e7. https://doi.org/10.1016/j.jaci.2014.11.043

26. Perzanowski MS, Chew GL, Divjan A, Jung KH, Ridder R, Tang D, et al. Early-life cockroach allergen and polycyclic aromatic hydrocarbon exposures predict cockroach sensitization among inner-city children. J Allergy Clin Immunol. 2013;131(3):886–93. https://doi.org/10.1016/j.jaci.2012.12.666

27. Sedghy F, Sankian M, Moghadam M, Ghasemi Z, Mahmoudi M, Varasteh AR. Impact of traffic-related air pollution on the expression of Platanus orientalis pollen allergens. Int J Biometeorol [Internet]. 2017;61(1):1–9. Available from: https://doi.org/10.1007/s00484-016-1186-z

28. RECORD Reporting Guidelines [Internet]. [cited 2017 Dec 19]. Available from: http://www.record-statement.org/

29. White E. 2011 Census - Office for National Statistics [Internet]. 2012 [cited 2017 Dec 19]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/

bulletins/2011censuskeystatisticsforwales/2012-12-11#tab---Proficiency-in-Welsh

30. Public Health Wales. Measuring inequalities 2011: Trends in mortality and life expectancy in Wales [Internet]. 2011. Available from: http://www.publichealthwalesobservatory.wales.nhs.uk/inequalities-2011

31. Guarnieri M, Balmes JR. Outdoor air pollution and asthma. Lancet (London, England). 2014 May;383(9928):1581–92. https://doi.org/10.1016/S0140-6736(14)60617-6

32. Perera FP. Multiple Threats to Child Health from Fossil Fuel Combustion: Impacts of Air Pollution and Climate Change. Environ Health Perspect. 2017 Feb;125(2):141–8. https://doi.org/10.1289/ehp299

33. Rodgers SE, Demmler JC, Dsilva R, Lyons RA. Protecting health data privacy while using residence-based environment and demographic data. Heal place. 2012;18(2):209–17. https://doi.org/10.1016/j.healthplace.2011.09.006

34. Lyons R A, Jones KH, John G, Brooks CJ, Verplancke J-P, Ford D V, et al. The SAIL databank: linking multiple health and social care datasets. BMC Med Inform Decis Mak [Internet]. 2009 Jan [cited 2014 Jan 14];9:3. Available from: https://doi.org/10.1186/1472-6947-9-3

35. Jones KH, Ford D V., Jones C, Dsilva R, Thompson S, Brooks CJ, et al. A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation. J Biomed Inform. 2014 Aug;50(100):196–204. https://doi.org/10.1016/j.jbi.2014.01.003

36. Rodgers SE, Lyons RA, Dsilva R, Jones KH, Brooks CJ, Ford D V., et al. Residential Anonymous Linking Fields (RALFs): A Novel Information Infrastructure to Study the Interaction between the Environment and Individuals' Health. J Public Health (Bangkok). 2009;10:1–7. https://doi.org/10.1093/pubmed/fdp041

37. Lyons RA, Ford D V., Moore L, Rodgers SE. Use of data linkage to measure the population health effect of non-health-care interventions. Lancet. 2014;383(9927):1517–9. https://doi.org/10.1016/S0140-6736(13)61750-X

38. SAIL Databank - The Secure Anonymised Information Linkage Databank [Internet]. [cited 2017 Oct 3]. Available from: https://saildatabank.com/saildata/data-privacy-security/#protecting-identities

39. Jones KH, Ford D V., Jones C, Dsilva R, Thompson S, Brooks CJ, et al. A case study of the secure anonymous information linkage (SAIL) gateway: A privacy-protecting remote access system for health-related research and evaluation. J Biomed Inform [Internet]. 2014;50:196–204. Available from: https://doi.org/10.1016/j.jbi.2014.01.003

40. Read Codes - NHS Digital [Internet]. [cited 2017 Dec 19]. Available from: https://digital.nhs.uk/article/1104/Read-Codes

41. Al Sallakh MA, Vasileiou E, Rodgers SE, Lyons RA, Sheikh A, Davies GA. Defining asthma and assessing asthma outcomes using electronic health record data: a systematic scoping review. Eur Respir J. 2017 Jun;49(6). https://doi.org/10.1183/13993003.00204-2017

42. Hammersley V, Flint R, Pinnock H, Sheikh A. Developing and testing search strategies to identify patients with active seasonal allergic rhinitis in general practice. Prim Care Respir J J Gen Pract Airways Gr. 2011 Mar;20(1):71–4. https://doi.org/10.4104/pcrj.2010.00086

43. Carruthers DJ, Holroyd RJ, Hunt JCR, Weng WS, Robins AG, Apsley DD, et al. UK-ADMS: A new approach to modelling dispersion in the earth's atmospheric boundary layer. J Wind Eng Ind Aerodyn. 1994 May;52:139–53. https://doi.org/10.1016/0167-6105(94)90044-2

44. Stocker J, Hood C, Carruthers D, McHugh C. ADMS-Urban: developments in modelling dispersion from the city scale to the local scale. Int J Environ Pollut. 2012;50(1/2/3/4):308. https://doi.org/10.1504/ijep.2012.051202

45. Harron KL, Doidge JC, Knight HE, Gilbert RE, Goldstein H, Cromwell DA, et al. A guide to evaluating linkage quality for the analysis of linked data. Int J Epidemiol [Internet]. 2017;1699–710. Available from: https://doi.org/10.1093/ije/dyx177

46. Ralphs M, Elkin M, Whitworth S, Wall J, Rasulo D. Beyond 2011: Matching Anonymous Data [Internet]. 2014. Available from: https://www.ons.gov.uk/census/censustransformationprogramme/beyond2011censustransformationprogramme/reportsandpublications

Abbreviations

ADMS	Atmospheric Dispersion Modelling System
AQ	Air Quality (pollen and air pollution)
CORTEX	Cognition, respiratory tract illness and effects of exposure
GP	General practitioner; primary care physician
NO_2	Nitrogen Dioxide
O_3	Ozone
PM	Particulate Matter
SAIL	Secure Anonymised Information Linkage
SAR	Seasonal allergic rhinitis
SES	Socioeconomic status
STROBE	Strengthening the reporting of observational studies in epidemiology
WDS	Welsh Demographic Service Dataset