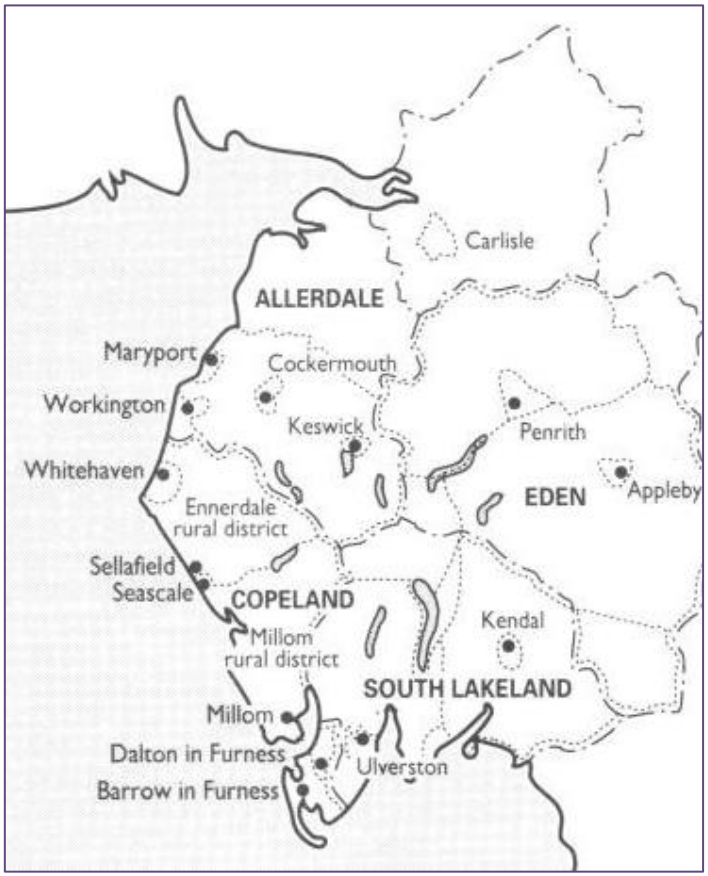
Linkage to spatial pollution data

Anna L Hansell

Assistant director of Small Area Health Statistics Unit
MRC-HPA Centre for Environment and Health
School of Public Health, Imperial College London

SAHSU

- **Established in 1987 as a recommendation of the Black enquiry into the incidence of leukaemia and lymphoma in children and young adults near the Windscale / Sellafield nuclear power plant**
- The main aim of SAHSU is to assess the risk to the health of the population of exposure to environmental pollutants, with an emphasis on the use and interpretation of routine health statistics



Why am I here?

SAHSU have a lot of experience in linkage between routine health data and spatially determined pollution estimates e.g.

Chlorination Disinfection By-Products and Risk of Congenital Anomalies in England and Wales

Mark J. Nieuwenhuijsen,[1,2] Mireille B. Toled
Cornelis de Hoogh,[1] Diana Wellesley,[3] Patric
David Briggs,[1] Lars Jarup,[1] and Paul Elliott[1]

BMJ

RESEARCH

Mobile phone base stations and early childhood cancers: case-control study

Paul Elliott, professor of epidemiology and public health medicine, head of department, director, MRC-HPA centre for environment and health, Mireille B Toledano, senior lecturer in epidemiology, J Bennett, research fellow, L Beale, research fellow, K de Hoogh, senior research officer, N Best, professor of statistics and epidemiology, D J Briggs, professor, chair in environmental and health sciences, school of public health

Risk of adverse birth outcomes landfill sites

Paul Elliott, David Briggs, Sara Morris, Cornelis de Hoogh, Christopher Hurt, Tina Kold Jensen, Ian Maitland, Sylvia Richardson, Jon Wakefield, Lars Jarup

Adult Cancers Near High-voltage Overhead Power Lines

Paul Elliott,[a] Gavin Shaddick,[b] Margaret Douglass,[a] Kees de Hoogh,[a] David J. Briggs,[a] and Mireille B. Toledano[a]

Background: Extremely low-frequency magnetic fields are designated as possibly carcinogenic in humans, based on an epidemiologic association with childhood leukemia. Evidence for associations with adult cancers is weaker and inconsistent.
Methods: We conducted a case-control study to investigate risks of adult cancers in relation to distance and extremely low-frequency

Conclusion: Our results do not support an epidemiologic association of adult cancers with residential magnetic fields in proximity to high-voltage overhead power lines.

(*Epidemiology* 2013;24: 184–190)

Small Area Statistics Un

r Environment &Health

Medical Research Council | Public Health England | KING'S College LONDON

Why am I here?

SAHSU staff are also involved in conducting linkage between cohorts and spatially determined pollution estimates

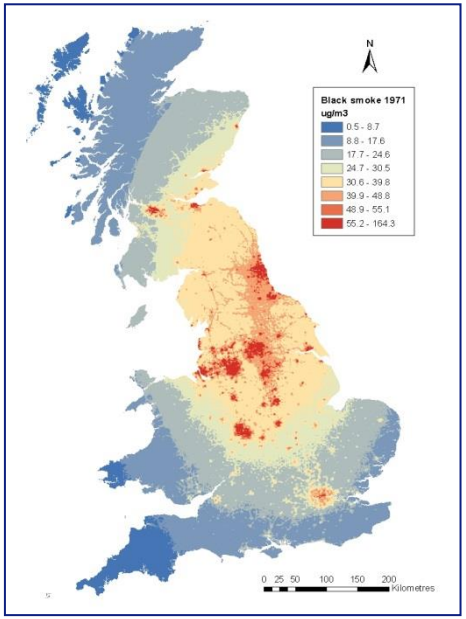
CHESS project – Longitudinal Study

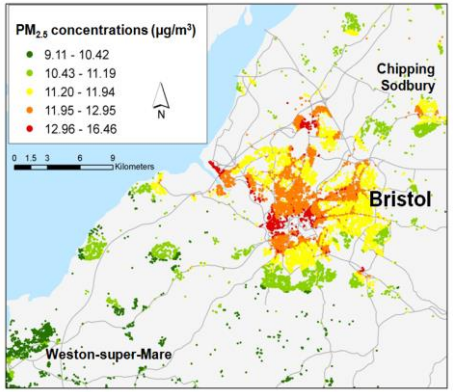
LEAP project – NSHD (1946 birth cohort) – air pollution estimates from four different projects (CHESS, Imperial, RGI, ESCAPE)

ESCAPE project - Born in Bradford, NSHD, ECRHS, EPIC Oxford

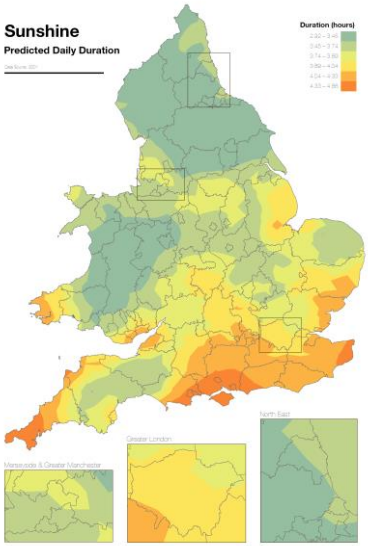
Bioshare project– HUNT, Lifelines, EPIC Oxford, UK Biobank

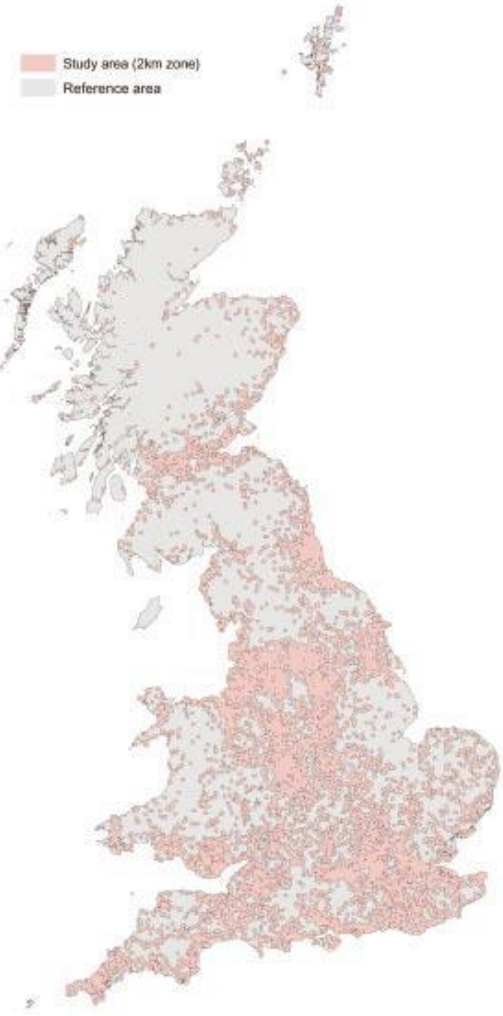
Others include…ALSPAC, Transplant Register....

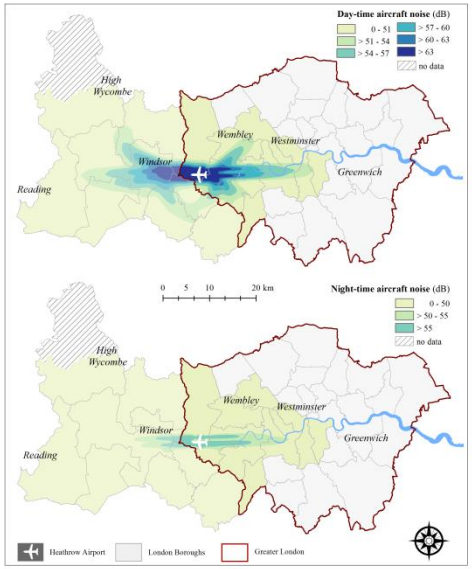


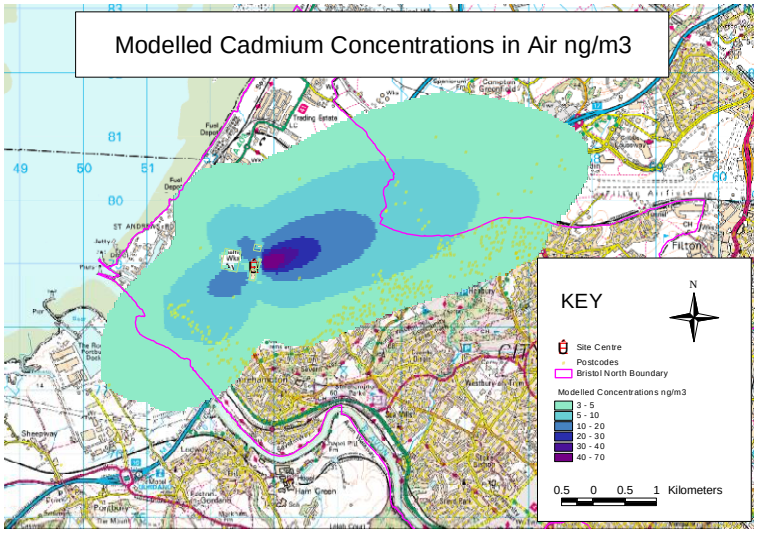


Examples of spatially determined exposures









Small Area Health Statistics Unit

MRC-PHE Centre For Environment &Health
Imperial College London
MRC Medical Research Council
Public Health England
King's College London

Examples of geographic data held by SAHSU

Theme	Description	Content	Geographic coverage	Data provider
Postcode Data	SAHSU2007v1	Postcode database covering 1974 to 2006	UK	Based on EDINA data
Address Data	AddressPoint	Addresses and co-ordinates	UK	Ordnance Survey
Topographic Data	AA map data	Transport networks, urban areas, rivers, coastline	Great Britain	AA
	Strategic tiles, 1:250000	Transport networks, urban areas, altitude (spot heights), water features	Great Britain	EDINA, OS
	MasterMap	Road Network, buildings, other topographic features	Great Britain	EDINA, OS
	tiles, 1:50000	Road network, water features, woods, place names	Great Britain	EDINA, OS
	road data, 000	Road network, speed, flow	Greater London	GLA
	data, 1:10,000	Road network, speed, flow	Greater London	GLA
	CIS version 6	Altitude, slope angle, road density, urban land	Great Britain	DEFRA
	Land-form PANORAMA	DTM	Great Britain	EDINA, OS
	Elevation Data	Gridded altitude data	Europe	LPDAAC
Boundary Data	1971 Census boundaries and centroids	Census boundaries and centroids	Great Britain	UKBORDERS
	1981 Census boundaries and centroids	Census boundaries and centroids	Great Britain	UKBORDERS
	1991 Census boundaries and centroids	Census boundaries and centroids	Great Britain	UKBORDERS
	2001 Census boundaries	Census boundaries	Great Britain	UKBORDERS
	Health boundaries	Boundaries for regional and local health authorities	Great Britain	UKBORDERS/ONS
	Administrative boundaries	NUTS 3	EU	GISCO

Examples of environmental data held by SAHSU

Theme	Description	Content	Geographic Coverage	Years Covered	Data Provider
Landfills	Landfill sites	Name, co-ordinates, waste-type, operating status	Great Britain	1998-2001	EA/WRc
Incinerators	Municipal Solid Waste Incinerators	Location of 22 incinerators in UK ???	Great Britain	2003-2010	EA
Composting sites	Large and medium size composting facilities in the UK	Location of >200 facilities	England and Wales	2008	EA
Drinking water	Water supply zone	Boundaries and THM data	Water companies in England and Wales	1998-2001	Water Companies
Land cover	CIS version 6	Land cover (16 classes); designated areas	Great Britain	1999	DEFRA
	CORINE landcover	Land cover (44 classes)	EU (except) plus (in part)	Ca. 1991	GISCO
	Agricultural Census	Crop/agricultural land use data	England	2000	DEFRA
Light emissions	Modelled light emissions	Continuous integer scale from 0 to 65000 (no units)	EU	1999-2000	DMSP
Air pollution	Atmospheric Emissions Inventory	Annual emissions of So_x	Great Britain	1995	LRC
	Background air pollution	Concentrations of background levels of benzene, SO_2 , PM_{10} , No_x on 1km grids	Great Britain	1996 onwards	NETCEN
	Chemical Release Inventory	Point source emissions for all chemical emissions	England and Wales	1991-2000	Environment Agency
Powerlines					
Mobile phone masts	Mobile phone mast base station sites				

Exposures used in the SAHSU Environment and Health Atlas for England and Wales

- **Air Pollution (NO_2, PM_{10})** – Land Use Regression model (100m x100m grid) using data from National Air Quality Archive's
- **Radon Potential**- Modelled by the HPA and the British Geological Survey (BGS)
- **Sunshine Duration** - Meteorological Office, through MIDAS Land Surface Station Data
- **Metals (Cadmium, lead)** - Data collected and analysed by the Centre for Ecology and Hydrology as part of the 2000 Countryside Survey
- **Agricultural Pesticides** - Data from The Pesticides Usage Survey, conducted by The Food and Environment Research Agency
- **Chlorination Disinfection By-products** – Data from 10 Water companies

Air pollution data held at Imperial

Pollutant	Source	Geographical extent	Years	Resolution	Comments
Currently available					
Road traffic	CORINE	UK	1999-2009/10	1m	Assigned to address using GIS.
Distance from road	CORINE	UK	2001 &, 2009	1m	Assigned to address using GIS. Various metrics possible – distance to road, weighted road density, traffic on nearest road, defined buffer radii e.g. 50m, 100m,
$PM_{2.5}$	ESCAPE	400km of London	2009/10	Modelled to address	Land Use Regression (LUR) modelling based on an intensive monitoring campaign. Validity of PM models decreases with distance from London.
PMcoarse	ESCAPE	400km of London	2009/10	Modelled to address	LUR as above
Pmabsorb-ance (soot)	ESCAPE	400km of London	2009/10	Modelled to address	LUR as above
PM_{10}	ESCAPE	400km of London	1999-2009/10	Modelled to address	LUR as above, extrapolated back in time using national monitoring network data. Validity of PM models decreases with distance from London.
NO_2, NOx	ESCAPE	England & Wales	1999-2009/10	Modelled to address	LUR as above, extrapolated back in time using national monitoring network data.
NO_2	Imperial (JG)	England & Wales	1991 & 2009	100m grid	LUR based on national monitoring network.
NO_2, PM_{10}	Imperial (RGI)	England & Wales	2001	200m grid	LUR based on national monitoring network.
Possible improvements if funded…					
PM_{10}, NOx, NO_2	Imperial	UK	1999-2010	100m grid and/or modelled to address	6 months GIS RA time needed to develop the LUR models There are too few national monitoring sites for $PM_{2.5}$ to develop a stable model.

Small Area Health Statistics Unit

MRC-PHE Centre For Environment &Health
Imperial College London | MRC Medical Research Council | Public Health England | King's College London

Small areas

Postcode – on average ~42 people in England & Wales

Census Output Area (COA) – on average 304 people in England

Super Output Area – on average 1552 people in England

Ward – on average 6000 people in England

Addresses are points (except they're not! Centroid? Front door?)
http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/postal/index.html

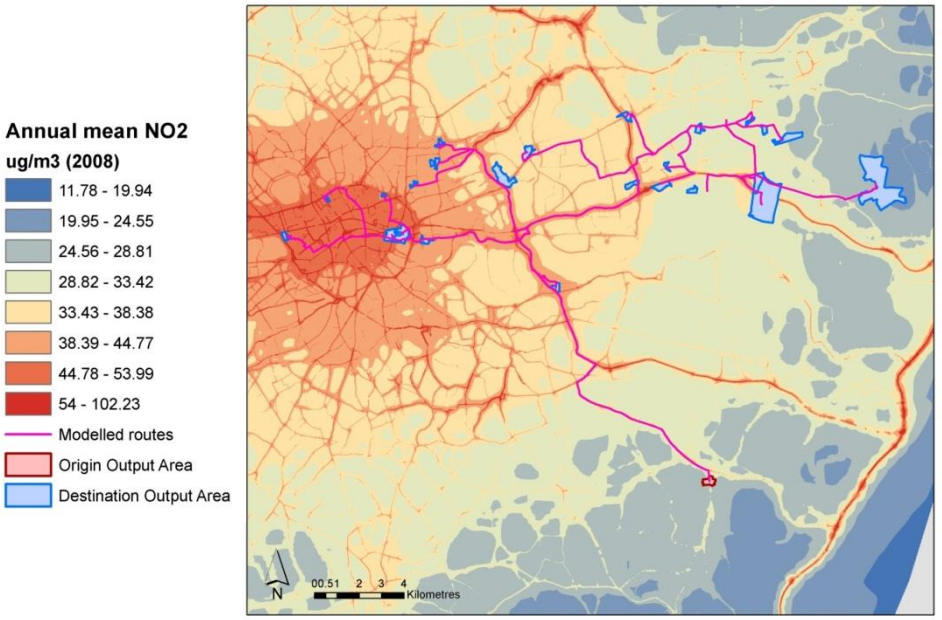
Assigning exposure to address

Usually to residential address

- Because that's what we usually have
- 60%+ of time spent indoors
- Children and elderly likely to spend more time at home
- Location of work information also useful

Does outside concentration = inside exposure = dose?

- Exposure misclassification – expected direction is to under-estimate dose-response for e.g. air pollution



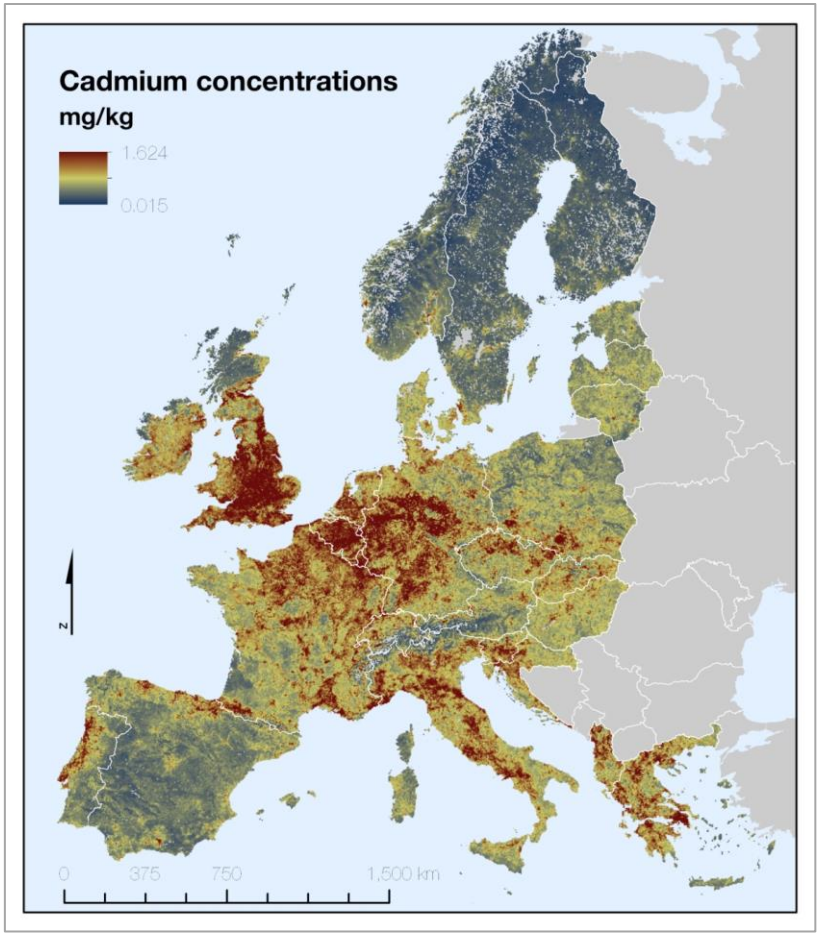
Example of routes modelled from one home-COA to multiple work-COAs

Technicalities

- Model exposure to address or postcode location (e.g. ESCAPE project)

- Drop address (or postcode or geocode or small area) onto a concentration surface or modelled grids
- Weighting can be conducted if small area>grid (e.g. SOA on top of 10m noise grids)

Estimated cadmium concentrations in European soils (Joint Research Centre, European Commission).



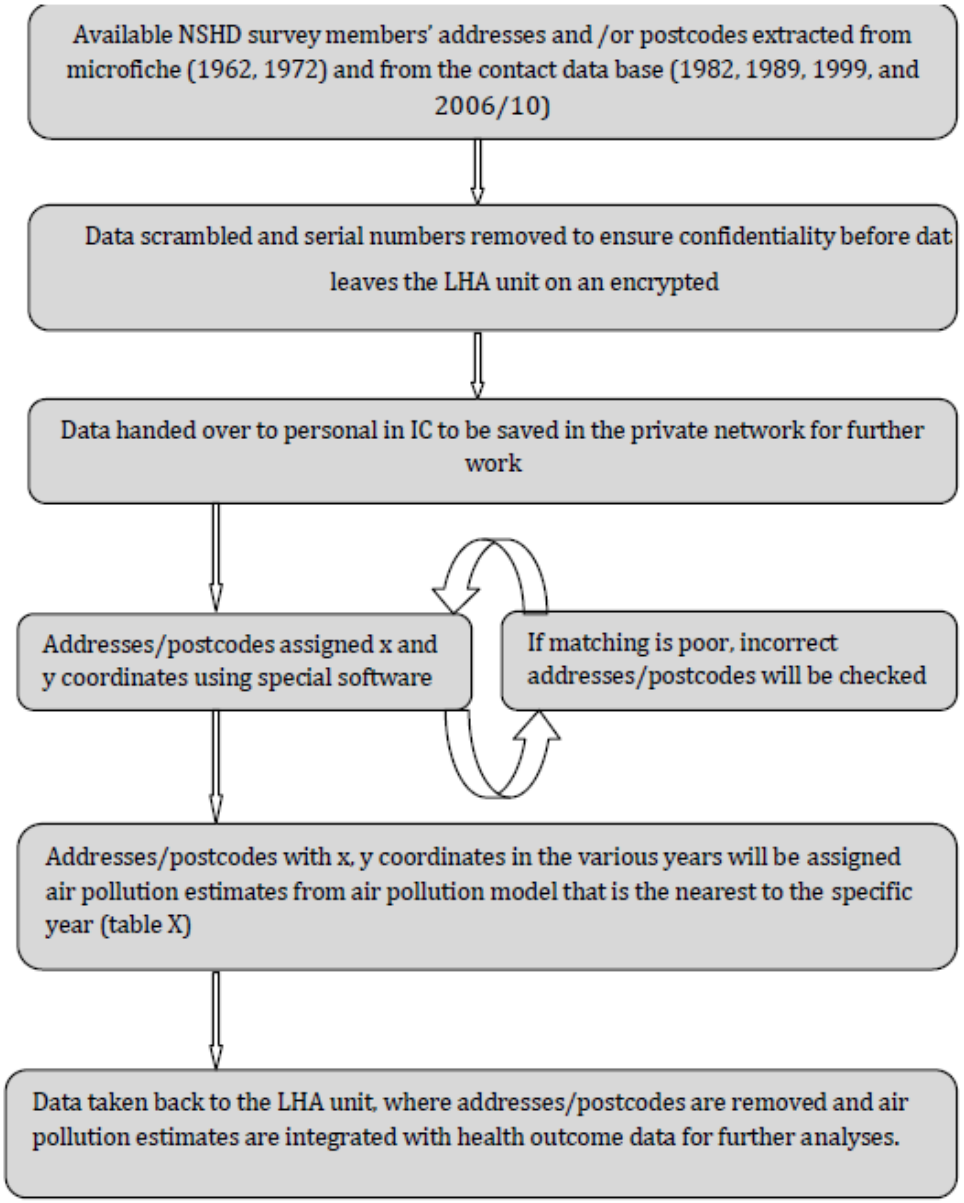
Constructing a residential history for follow-ups - NSHD

- Cohort born in 1946
- Postcodes first introduced in 1961
- Records available in digital format from 1980s only
- 1961 and 1972 addresses available on microfiche

- Annual update via birthday card (but may not give date of moving)
- Most reliable contact details at follow-up dates
- Residential history information restricted to follow-ups

- Confidentiality issues regarding address release vs. postcode
- Geocoding (assigning x-y coordinate)
  - Only 60-95% successful on first pass and needed manual inspection.
  - 88%-95% of addresses/postcodes successfully geocoding

Constructing a residential history - NSHD



Constructing a residential history- CAPS

- Childhood Asthma Prevention Study, Sydney Australia
- Contact database maintained *overwritten* with new contact details up until age 8

- Question ‘Have you moved house since age 5 years?’ included at age 8 years follow-up

- A residential history is now being constructed from paper records
- Issues:
  - Date the new address was recorded (a visit) vs. moving date
  - Complex families (divorce, grandparents)
  - Frequent movers
  - Discrepancies between different sources of information

Constructing a full residential history with digital records - ALSPAC

SAHSU have been working with ALSPAC to conduct a study of health effects of air pollution exposure in early life on lung health

We have developed a cleaning algorithm that fixed gaps and overlaps in the address periods for ~14,000 ALSPAC participants included in the Imperial study with ~18,000 addresses.

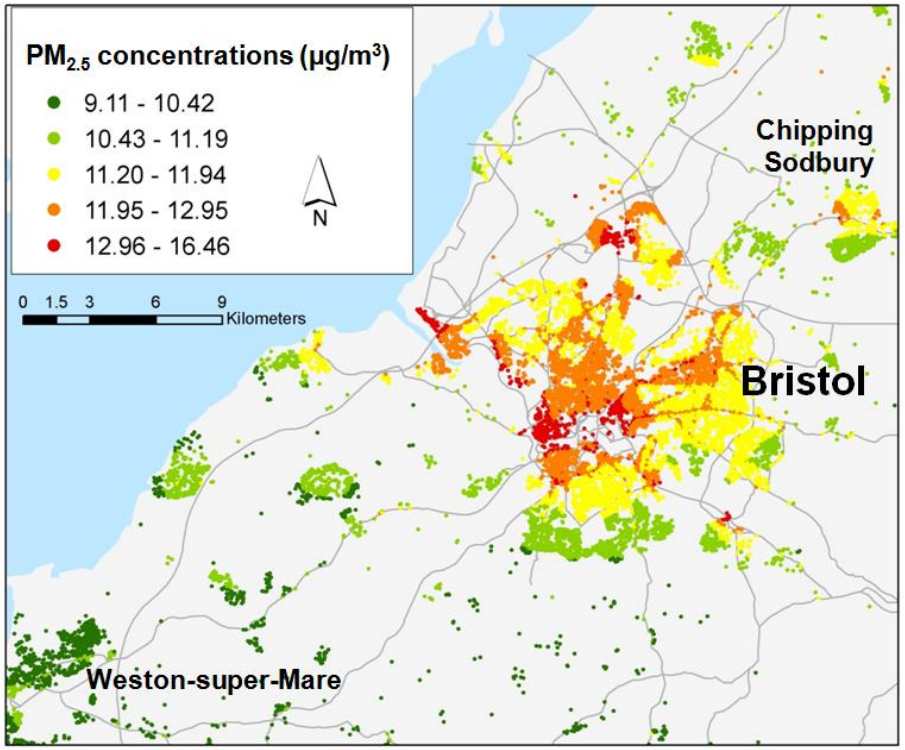
The cleaning algorithm makes an alteration if:

- at least one address record is regarded as a duplicate and deleted;
- if there is at least one record whose blank end date was imputed;
- or if at least one day in one address period was changed as a result of fixing a gap or an overlap

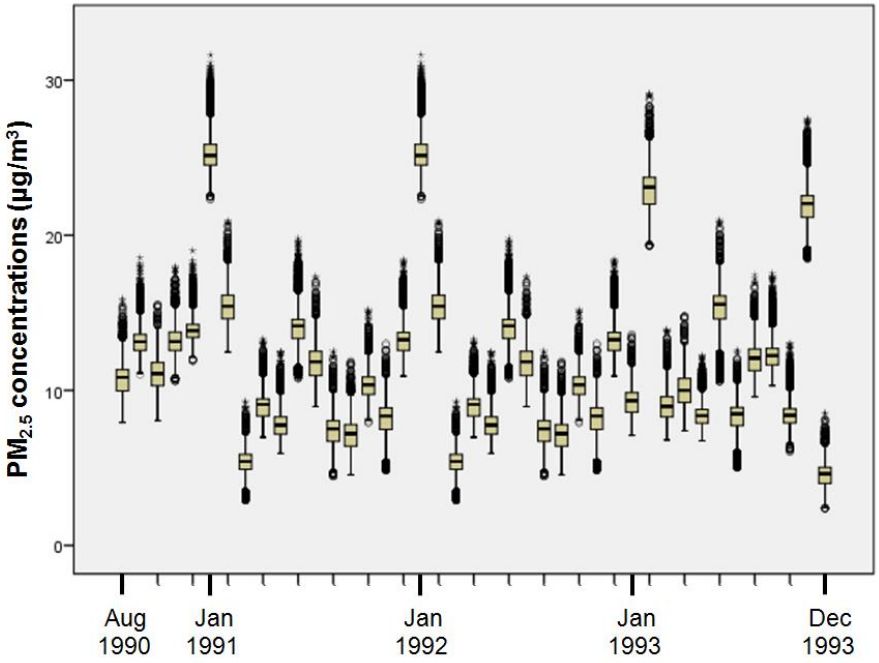
The algorithm alters 47% of records for the individuals in Imperial's study – or 18% if we don't count address records which had a blank end date imputed.

MODELLING HISTORICAL EXPOSURES TO FINE PARTICULATES ($PM_{2.5}$) IN BRISTOL AND AVONMOUTH: THE ALSPAC BIRTH COHORT STUDY

Modelled long-term $PM_{2.5}$ exposures (1990-1993) at cohort address.



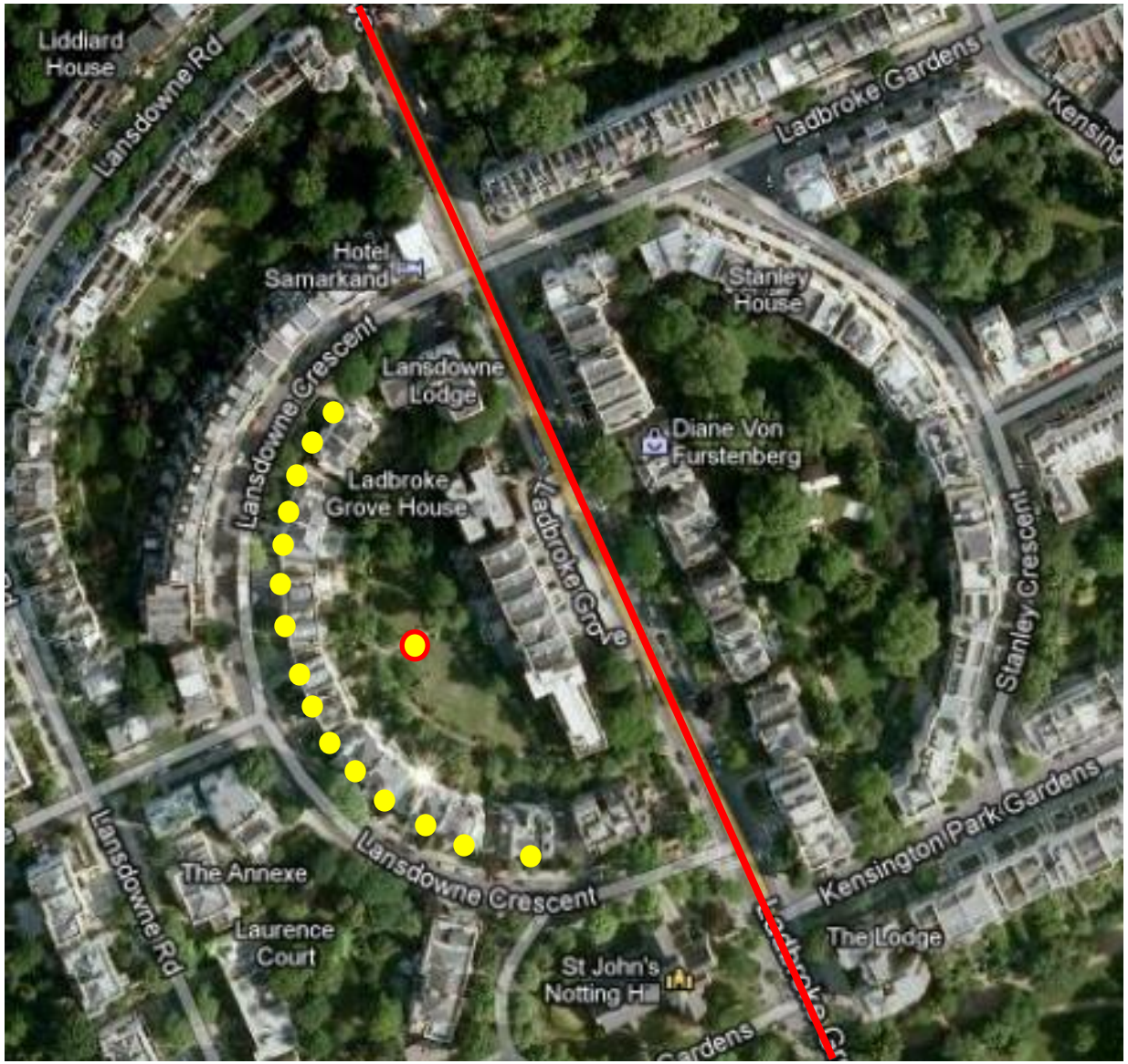
Modelled $PM_{2.5}$ exposures by calendar month (1990-1993)



Exposures modelled for c. 18,000 address locat

MRC-PHE Centre For Environment &Health
Imperial College London
MRC Medical Research Council
Public Health England
KING'S College LONDON

Improving exposure estimates by address vs. postcode



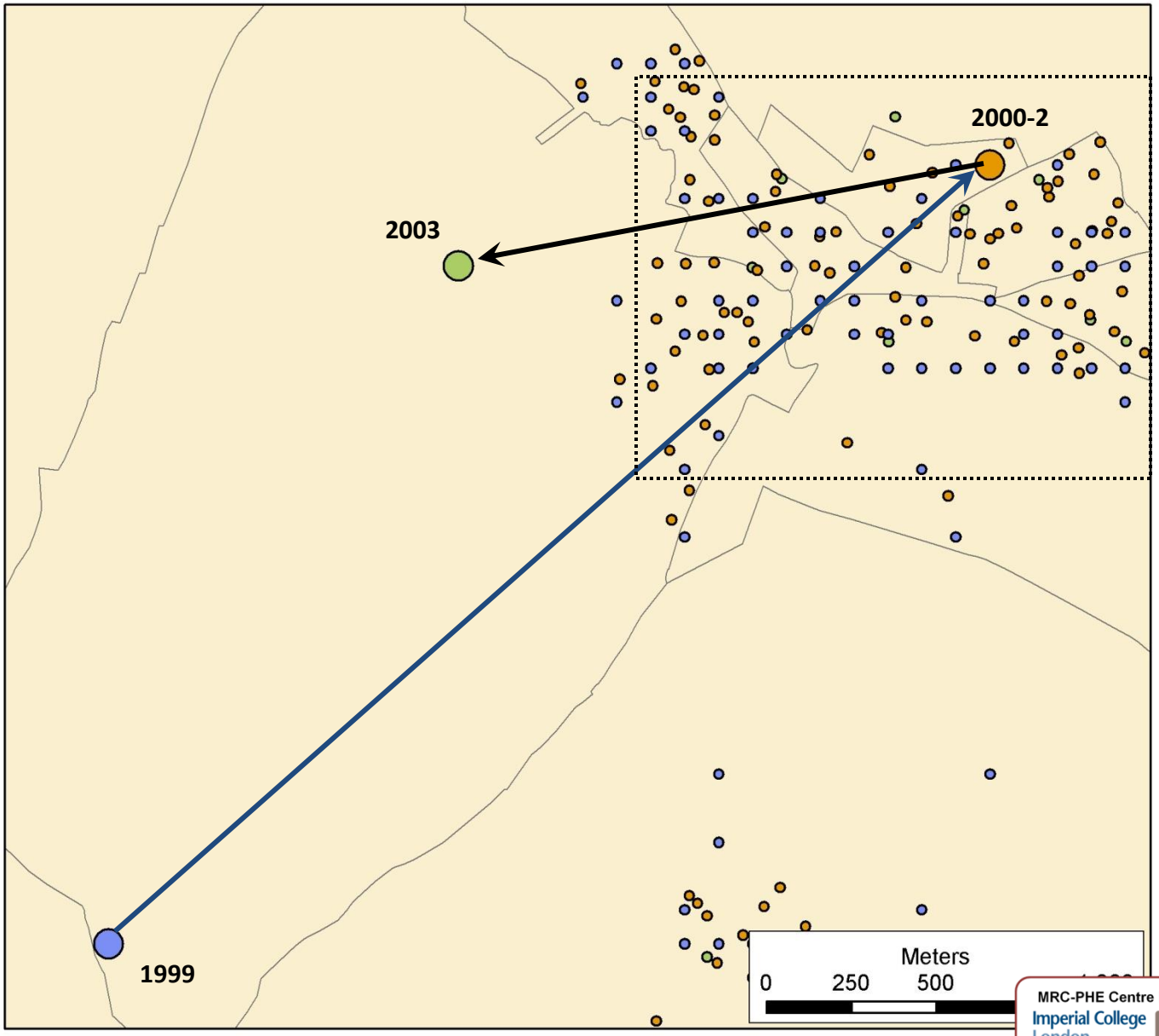
Address point

Postcode centroid

Major road

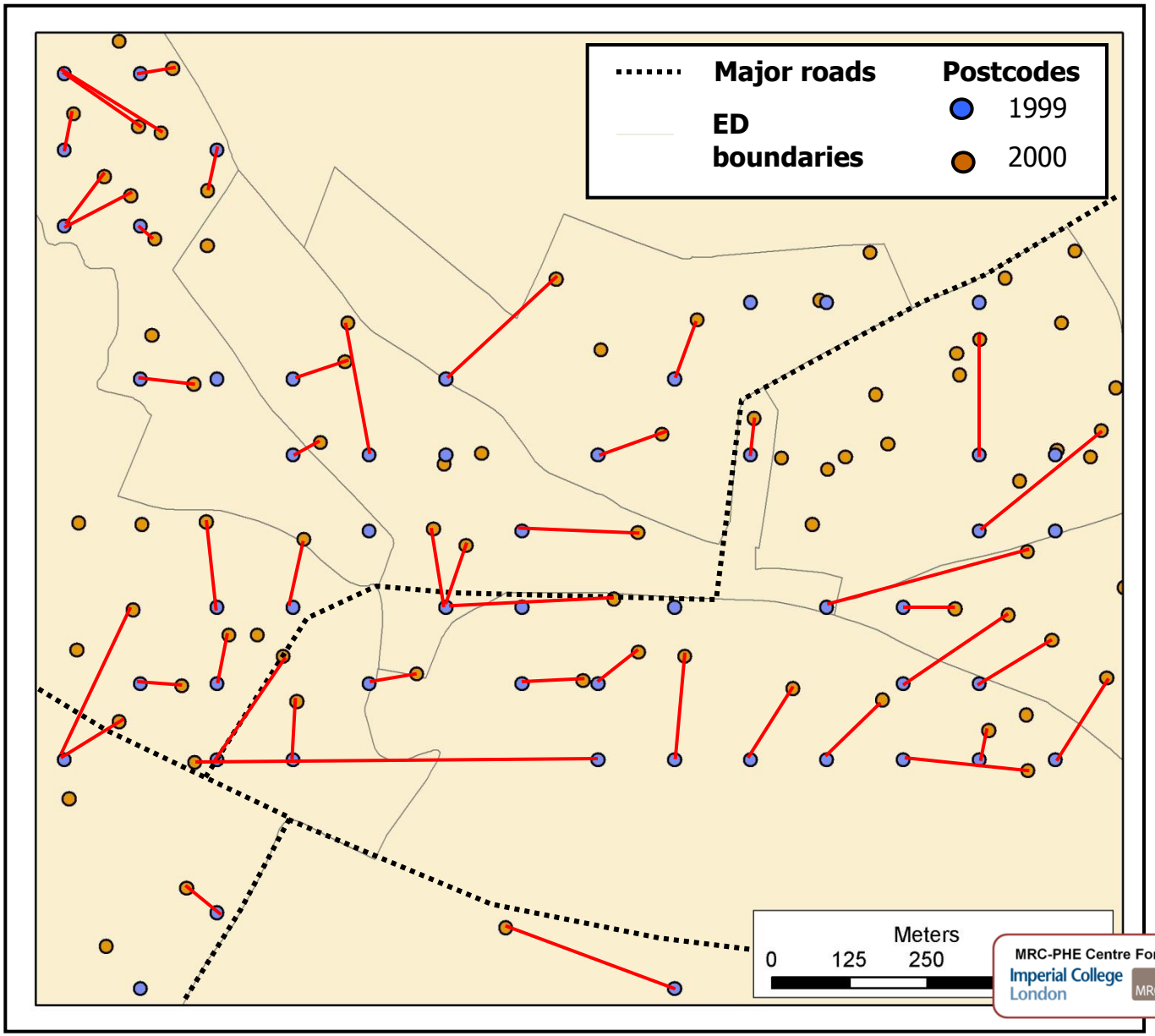
Geometric centroid of postcode
not representative of address exposures

Postcode CB1 6LN, 1999-2003



MRC-PHE Centre For Environment &Health
Imperial College London
MRC Medical Research Council
Public Health England
KINGS College LONDON

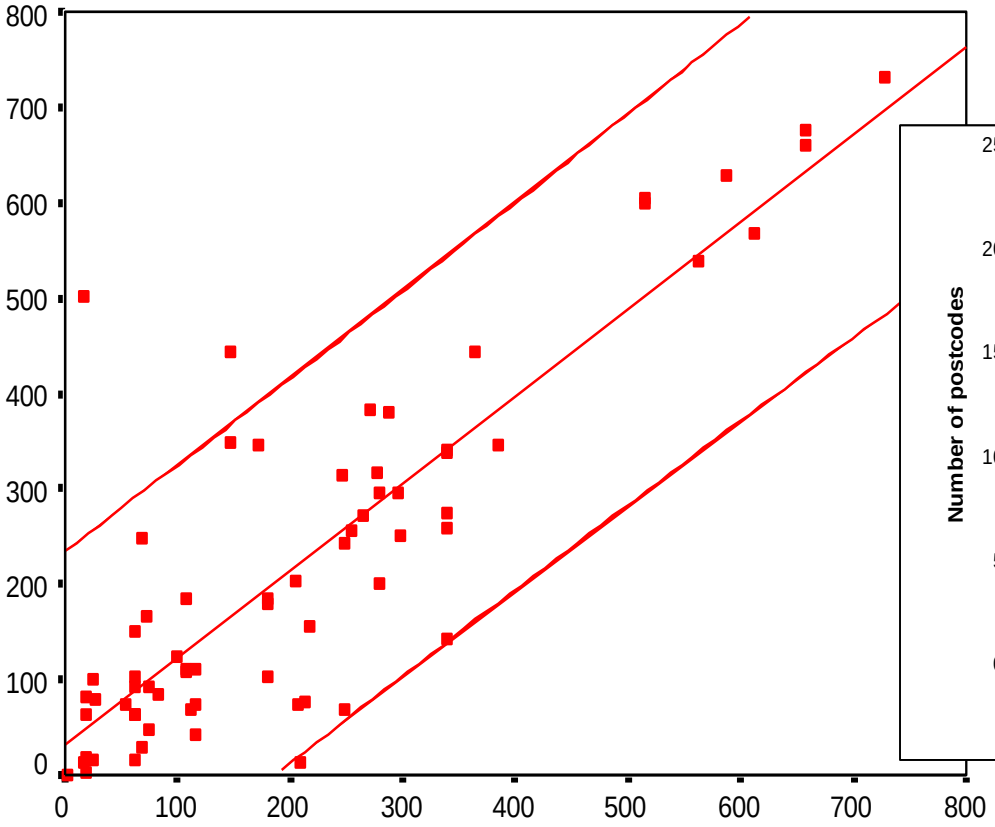
Postcode shifts, Cambridge (CB1), 1999-2000



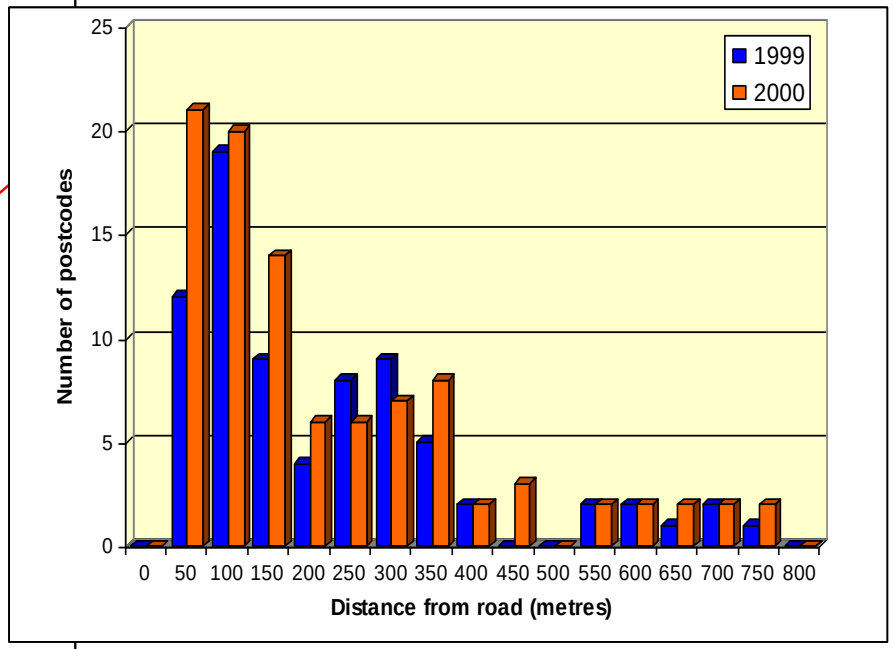
MRC-PHE Centre For Environment &Health
Imperial College London
MRC Medical Research Council
Public Health England
King's College London

Implications of uncertainties in postcode location

Distance from nearest main road, 1999 and 2000



Distance to road, 1999 (metres)



MRC-PHE Centre For Environment &Health
Imperial College London
MRC Medical Research Council
Public Health England
KING'S College LONDON

QUESTIONS FROM KEV

Do you record any date information for when case study members live at a given address?
Do you associate a start date when case study members live at a given address? Do you associate an end date?

lease tick any choices which describe features your software has for ensuring that correct date values are entered.
Ensures date fields are not blank
Ensures date fields are in a date format
Advertises preferred date format to users *(example: dd/mm/yyyy)*
Ensures start date is not greater than the end date
Ensures that there are no gaps between successive addresses.
Ensures users enter date in ways that do not require them to type *(example, drop down lists for months or graphically displayed calendar)*.
What date format does your software record date values? *Tick <u>one</u> choice*.

How do you validate addresses? How do you handle duplicates?

Thanks