

Linking administrative data from health and education

ECHILD: Education and Child Health Insights from Linked Data



HEALTH CARE



SCHOOLS



SOCIAL CARE



14.7 million children in England
Secure Research Service
Office of National Statistics

Safe-guarded at every step



- Data
- Access
- Users
- Projects
- Outputs



Ruth Gilbert r.gilbert@ucl.ac.uk <https://www.ucl.ac.uk/child-health/echild>



Department
for Education

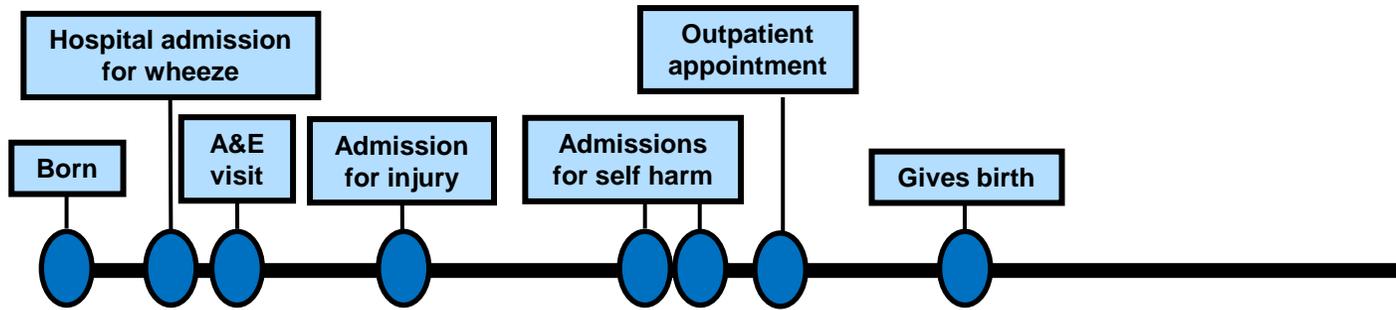


NIHR | Policy Research Unit
Children and Families

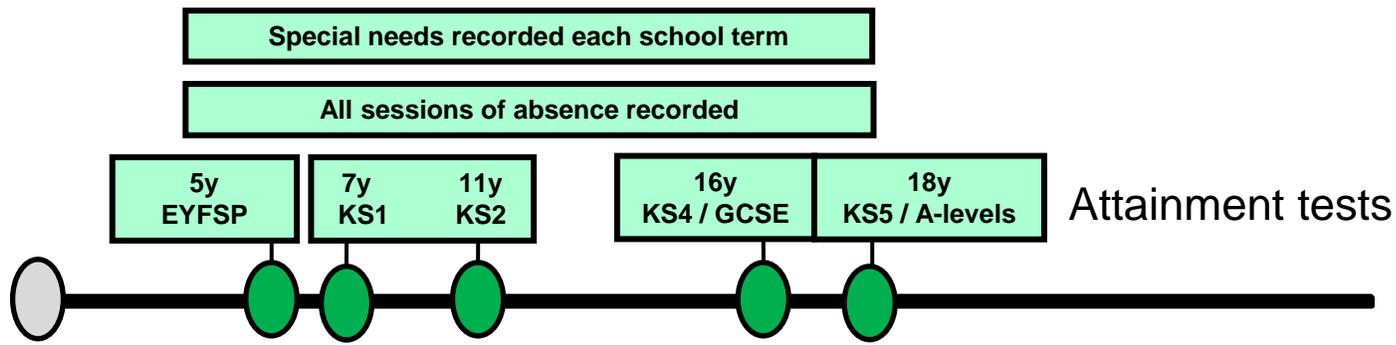


ECHILD trajectories

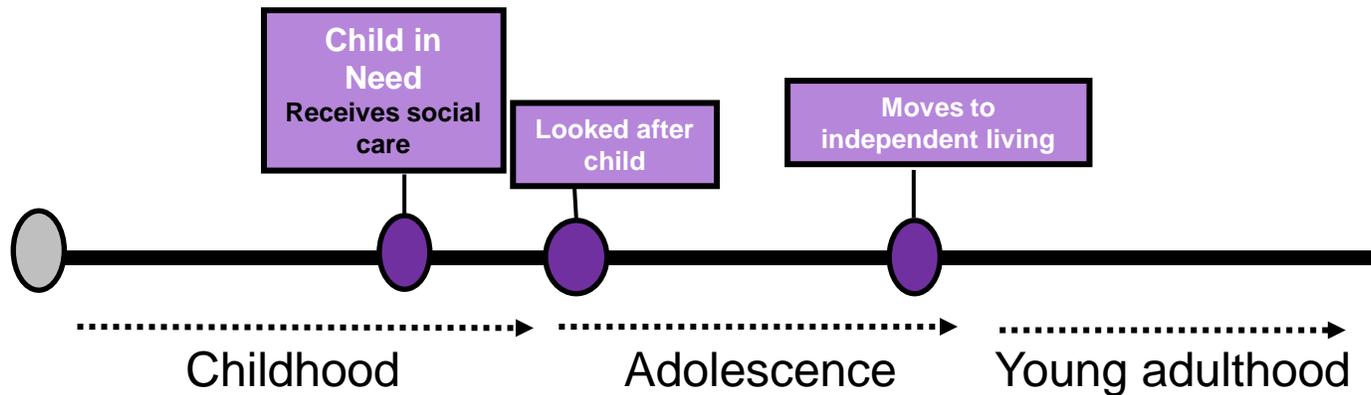
HEALTH CARE



SCHOOLS



SOCIAL CARE

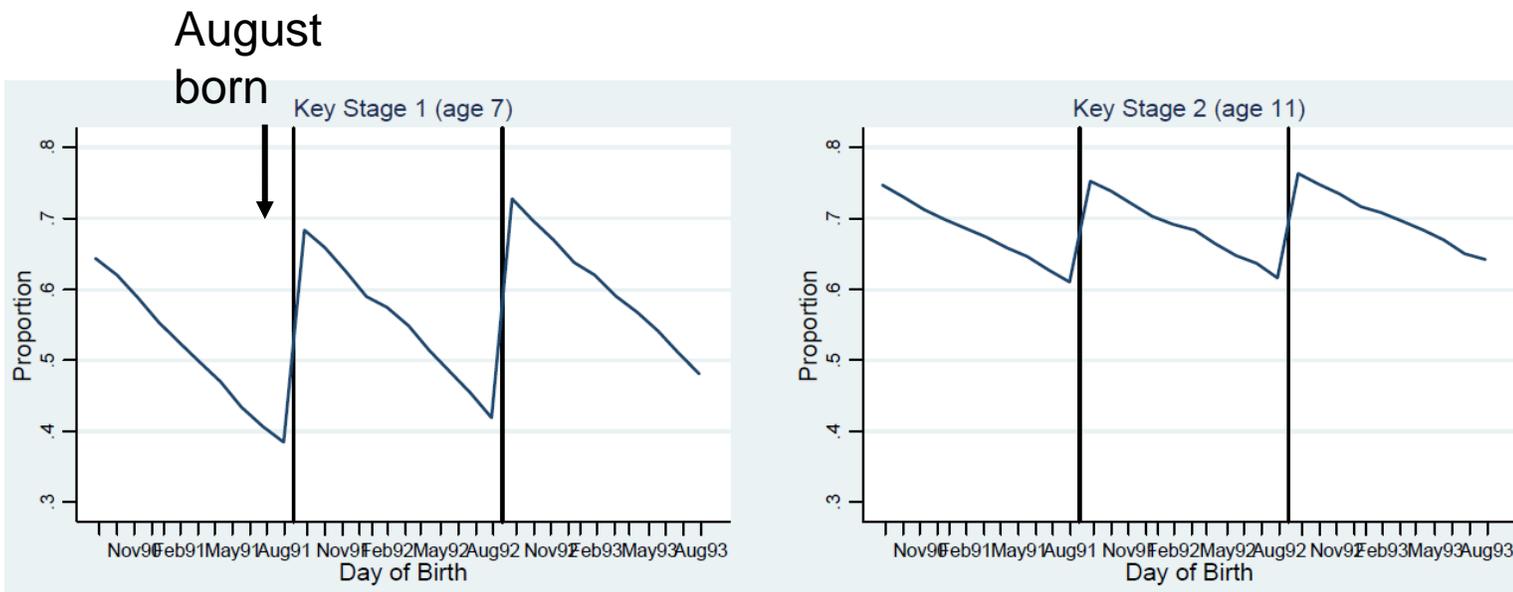


99% have a hospital record

> 95% of all children

25% of all children

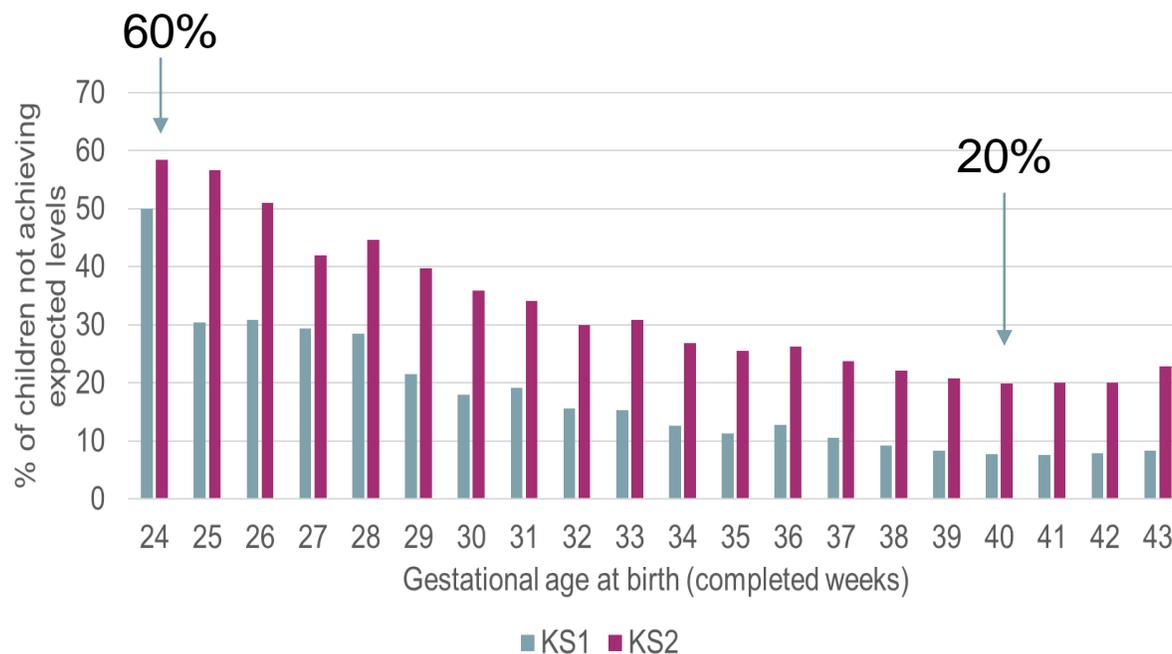
Month of birth



Pupils achieving expected levels

Crawford et al. When_you_are_born_matters_May_2013.pdf

Week of gestation



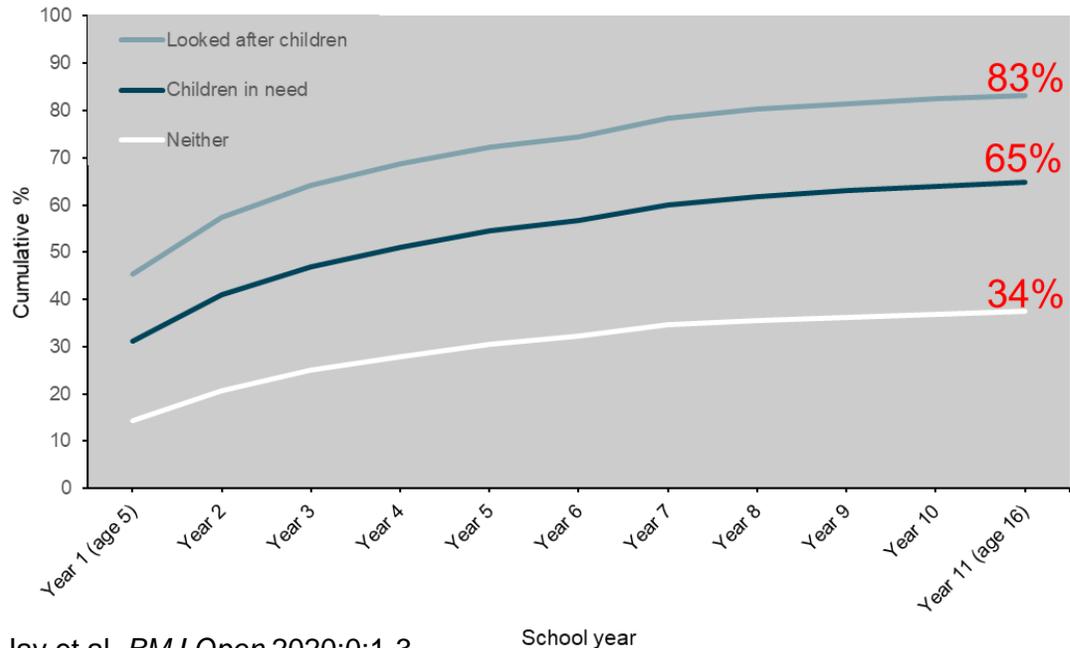
Pupils not achieving expected levels

(ECHILD - 452,013 pupils born 2004/5 linked to HES birth record)

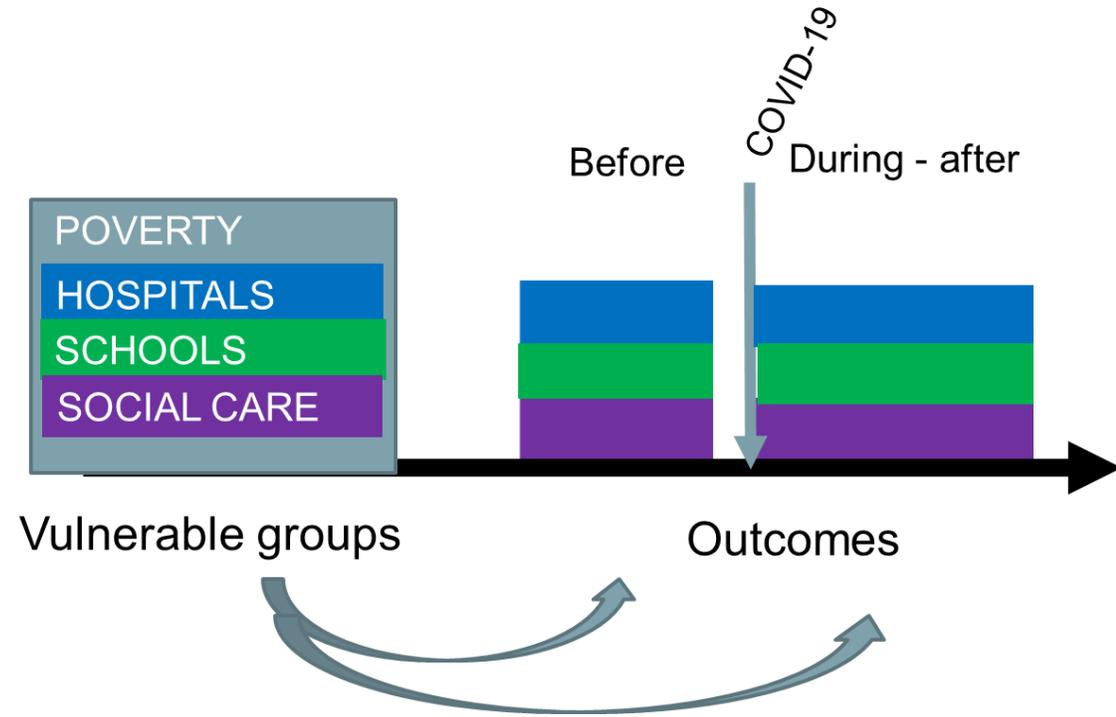
Multiple vulnerabilities

Special educational needs + social care

Cumulative incidence of special educational needs provision



Jay et al. *BMJ Open* 2020;0:1-3



Acknowledgements

Government: NHS Digital, Department for Education, Office of National Statistics, Department of Health and Social Care.

Academic team: Ruth Gilbert (PI), Ruth Blackburn, Louise Mc Grath-Lone, David Etoori, Nicolás Libuy, Erin Walker, Pia Hardelid, Jenny Woodman, Linda Wijlaars, Bianca De Stavola, Katie Harron, Matthew Jay, Ania Zylberzstejn (all UCL), & Christine Farquharson (IFS) Imran Tahir (IFS), Jan van der Meulen (LSHTM) and Min Hae Park (LSHTM).

Contact: ich.echild@ucl.ac.uk

Disclaimer: This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Wastewater monitoring for Covid-19 surveillance

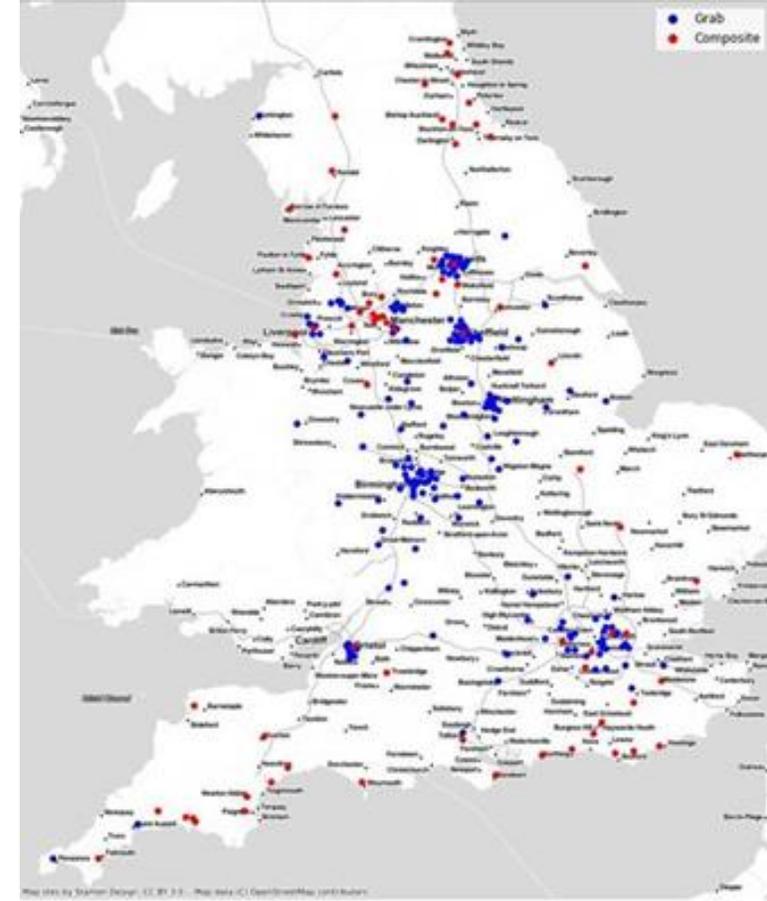
Dr Andrew Engeli, Deputy-Director EMHP

Joint Biosecurity Centre, UK Health Security Agency



What is wastewater testing and why does it matter?

- Wastewater testing involves taking samples from wastewater and testing it for traces of COVID-19
- Faecal shedding occurs in most people in the first 24 hours of viral acquisition for both symptomatic and asymptomatic cases
- This means COVID-19 is found in waste, and can be detected earlier than when symptoms emerge in people
- There are currently ~500 sites across England, testing wastewater at a community level covering 40 million people.



WBE has distinctive capabilities

Distinctive capabilities complementary to other approaches

- **Low cost and non-invasive**
- **Obviates community testing bias**
- **Detects asymptomatic and presymptomatic cases**
- **Unbiased estimator of overall prevalence**
- **Rapid insights into presence and spread of VOC/VUI**
- **Identifies other diseases circulating in the population (e.g., influenza, rotavirus)**

Five research themes



Covid-19 epidemiology

epidemiological insight for national Covid-19 response.



Wastewater genomics

use of existing and development of new genetic tools.



Quantification & detection

Optimisation of sample collection, handling, processing and analysis.



Beyond Covid-19

Extending to other health surveillance applications

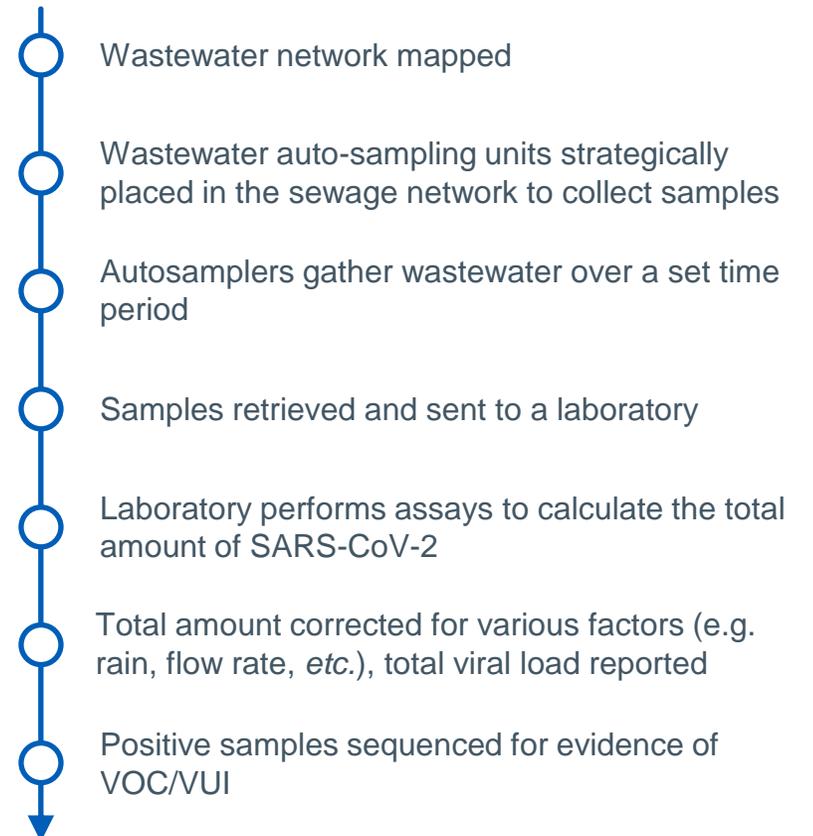


People, society, behaviour

Evaluating and improving the public health impact

Simple and rapid collect and analysis of data and insight

General COVID-19 wastewater-based epidemiology (WBE) process



WBE prevalence and incidence insights

Average WW virus concentration (29 Apr to 05 May 2021)



Figure 1: 7-day average concentration of SARS-CoV-2 RNA in wastewater at treatment works in England (254 sites). Data from 29 Apr to 05 May 2021

Average WW virus concentration change (29 Apr to 05 May 2021)

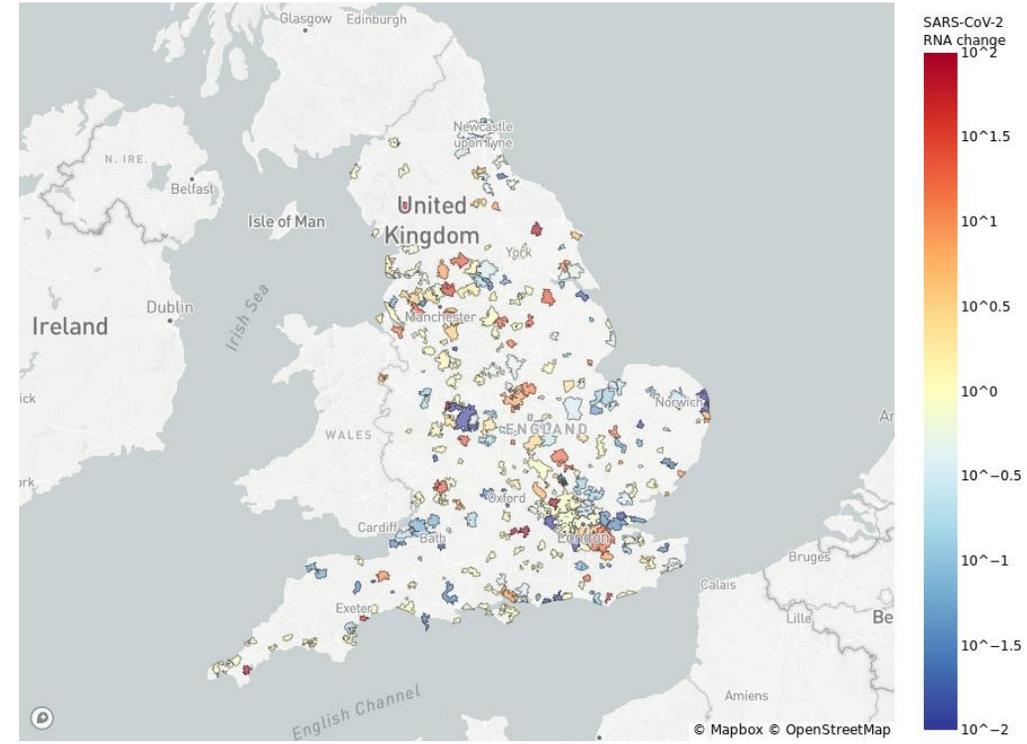
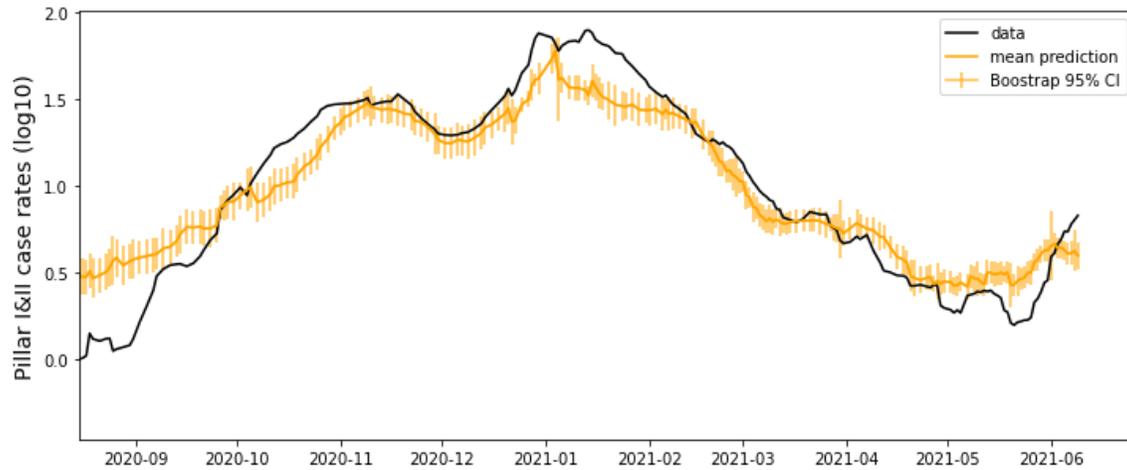


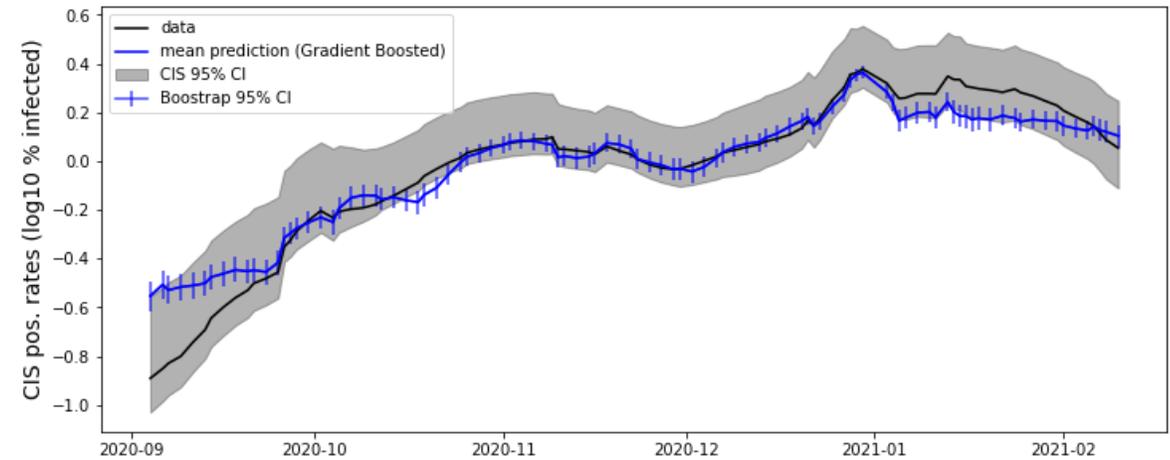
Figure 2: change in rolling 7-day average concentration of SARS-CoV-2 RNA in wastewater at treatment works in England (254 sites). Data from 29 Apr to 05 May 2021

Prevalence estimation

Test & Trace Pillar I&II



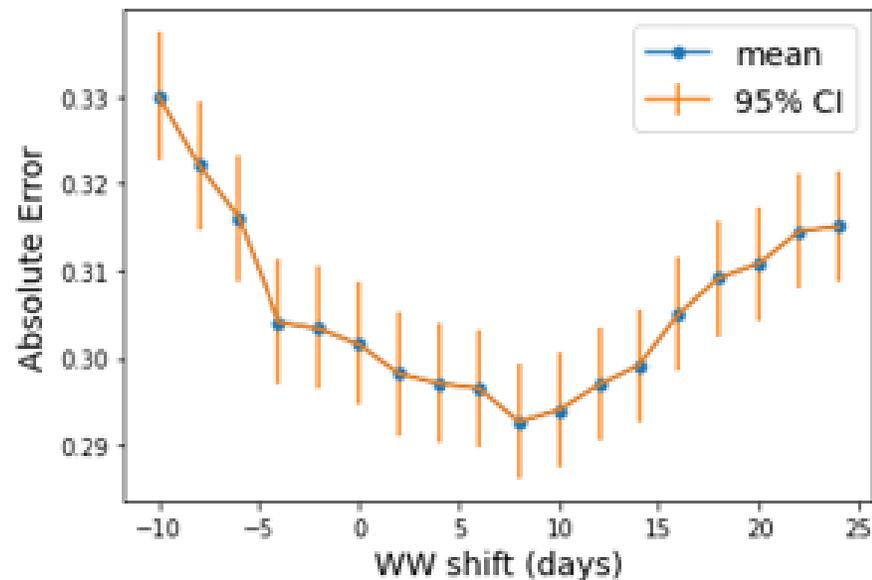
ONS CIS



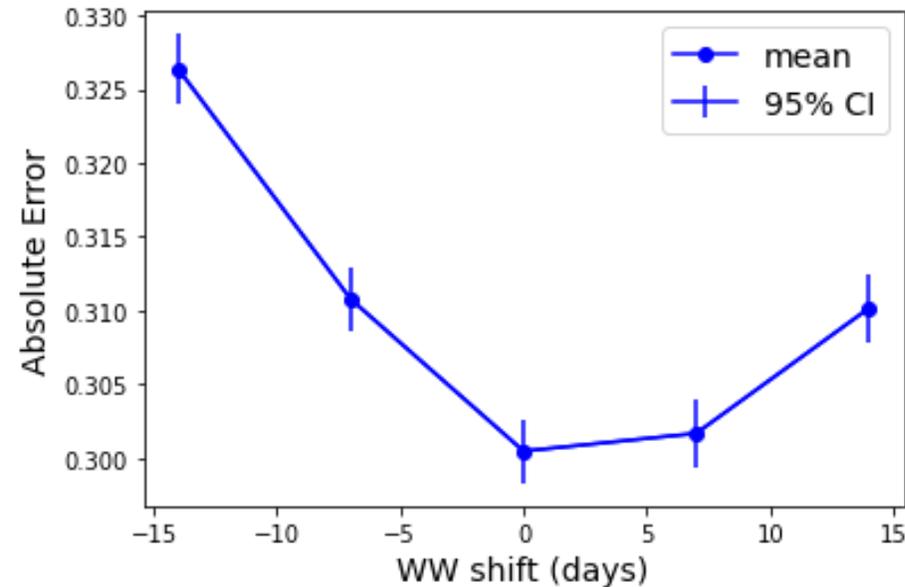
- Regression models are trained at subregional level to estimate **Pillar I&II new cases** from T&T or positivity rates from ONS **Coronavirus Infection Survey**
- Above plots show the nationally-averaged 7-day smoothed estimations using a gradient boosted regression trees model.
- Means and confidence intervals are estimated by bootstrapping by randomly drawing 50 independent 80% training/ 20% test sets

Lead and lag from regression models

Test & Trace Pillar I&II



ONS CIS



- These two figures show the average cross-validation prediction error (log10 space) when shifting the national Wastewater dataset backward and forward in time.
- When estimating Pillar I&II data (left panel), the error is minimal around 8 shifted days indicating a ~8 days lead from wastewater data over Test&Trace.
- When estimating ONS CIS (right panel), shifting the dates only decreases the models' performance, suggesting that no significant lead is present in this case between WW and CIS.
- The results are presented for the Gradient Boosted Regression Trees model but are similar for the linear and linear mixed models.

Challenges for WBE

1 Reliability

- Issues with **Levels of Detection (LOD)**
- High rate of false negatives

2 Insights

- Issues with **Levels of Quantification (LOQ)**
- Improving **probabilistic assessment** of VOC/VUI

3 Data and privacy

- Completely novel data source, **legal position** is that this is not disclosive data
- However, we know little about **public attitudes** towards WBE
- What are the major **ethical** considerations?

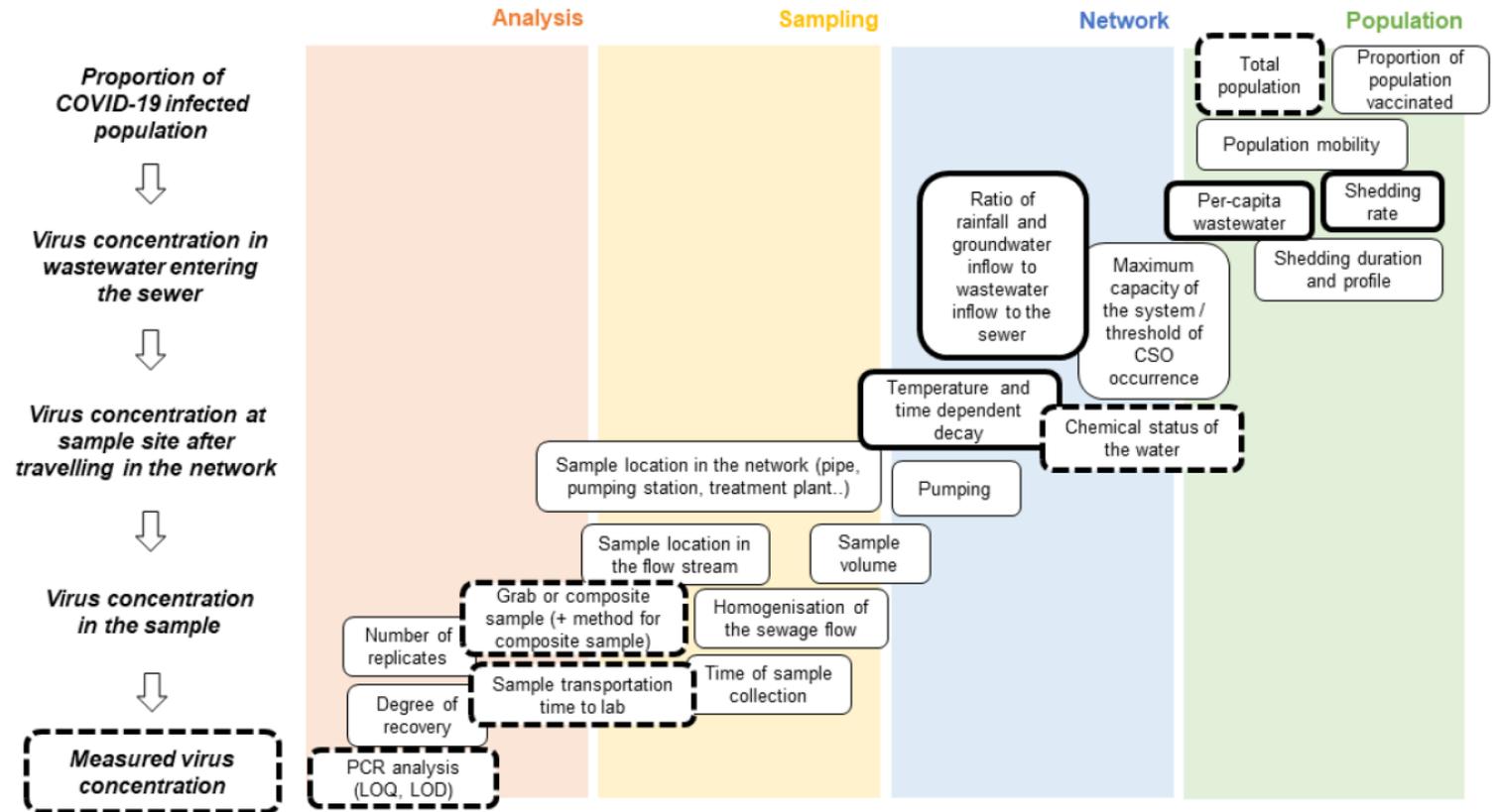


Figure 1: Factors hypothesised to influence measured concentrations of SARS-CoV-2 in wastewater.

The future vision for WBE

A fully automated UK wide smart network for the detection of health indicators in WW

In Depth Sequencing and Analysis

- Understanding the risk through **metagenomic sequencing** to identify specific concerns e.g. variants.

Invoke Public Health Intervention

- Share data and insights with key stakeholders for decision making
- By presenting stakeholders outside of WW team with this new dataset they are informed to make decisions to **slow and reduce the spread of the threat**

1 Data Collect

- Automated continuous sensors are placed at high risk pumping locations **generating real time data**.
- National remote network of monitors – internet of things
- Detection of multiple pathogens providing a **pulse check** on public health status

2 Anomaly Alert

- Once the prevalence threshold is surpassed, it will trigger an alert to the JBC or local Public Health Response teams for **immediate intervention and/or prioritisation**
- Smart sensor network acting as **early warning indicator**

3 Begin Surge Sampling

- Targeted action to **begin detailed sampling and analysis** at the point of interest.
- This spatial deep dive will **generate specific local insight** for decision makers

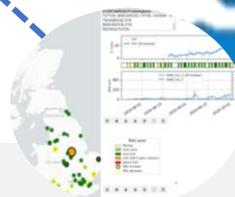
4

1

2

5

3



Local Data Spaces

Helping Local Authorities tackle COVID-19



Dr Mark Green mark.green@liverpool.ac.uk; @markalangreen



Economic
and Social
Research Council



UNIVERSITY OF
LIVERPOOL

- What is Local Data Spaces?
- Key insights of Local Authority needs
- Pilot outcomes



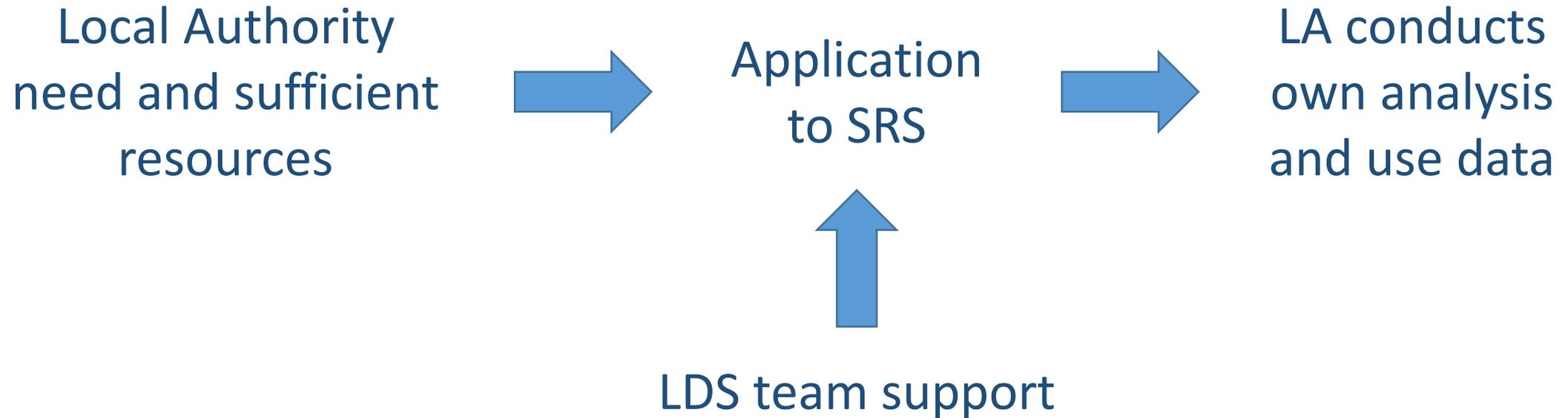
Introducing LDS

- Pilot Nov 2020 to April 2021
- Aim: To provide data and evidence support to Local Authorities in the response to COVID-19
- How can we best open up the data in the ONS Secure Research Service (SRS) to Local Authorities
- Three models of engagement:



Local Data Spaces

Model #1



E.g., Hackney and impact of COVID-19 on work



Local Data Spaces

Model #2



E.g., York and Tourism, Harborough and footfall, and Liverpool and 'Mass Testing'



UNIVERSITY OF
LIVERPOOL

Local Data Spaces

Model #3



E.g., Hull and occupational inequalities in COVID-19



UNIVERSITY OF
LIVERPOOL

- Co-produced outputs with ~25 Local Authorities
- 10 short reports for all 314 Local Authorities in England (<https://data.cdrc.ac.uk/geodata-packs>)
- 6 bespoke reports, including submission of evidence to SAGE with ONS
- R code for cleaning key SRS data products



Geospatial Inequalities in COVID-19 Risk : Leeds (Yorkshire and The Humber)

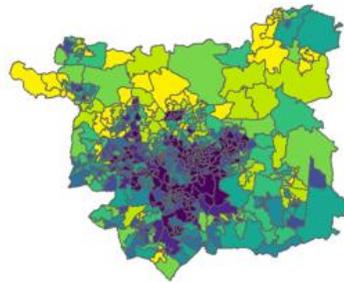
Local Data Spaces ADR UK Research Team

Index of Multiple Deprivation (IMD)

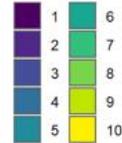
The map below shows IMD and Positivity Rates of COVID-19, for the First and Second Wave. The maps are presented on the same scale, with the same breaks, so can be directly compared. Index of Multiple Deprivation 1 means the LSOA is within the 10% most deprived in England. For the bivariate maps the deprivation has been swapped, so that a higher decile means a higher deprivation. To interpret these maps:

- Look at the Decile by LSOA map to provide context on deprivation (IMD 1 is most deprived).
- Look out for areas in the First Wave that are pink (high decile (IMD 1) and low positivity) that change to dark blue (high decile and high positivity).
- Look out for areas in the First Wave that are white (low decile (IMD 10) and low positivity) that change to turquoise (low decile and high positivity).

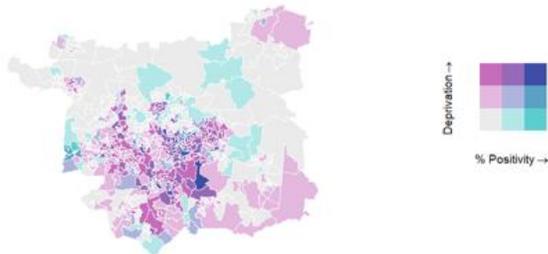
Index of Deprivation Decile



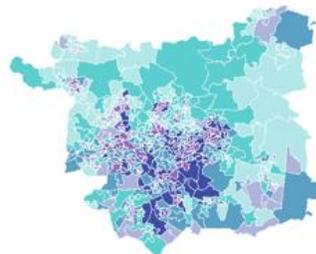
Index of Multiple Deprivation Decile



COVID-19 Positivity Rate and IMD in First Wave



COVID-19 Positivity Rate and IMD in Second Wave



Local Data Spaces

Outputs

Occupational Inequalities in COVID-19: Leeds

Local Data Spaces ADR UK Team

- Introduction
- Data and Sources
- Summary
- Employment Density Profiles and Ranking
- Workforce Distribution
- Workforce Growth Rates
- Workforce Breakdown by Sectors
- Individual Sectors
- Contact
- Partners

Introduction

The following report was prepared to preserve the aim of the descriptive statistics from 2020 to 31st January 2021. This work was endorsed by the research team which may not be endorsed by the research team.

Data and Sources

Local profiling is based primarily on two datasets available through the Office for National Statistics (ONS) Secure Research Service (SRS) catalogue:

Human Mobility Report for Leeds (Yorkshire and The Humber)

Local Data Spaces ADR UK Team

- Introduction
- Data and Sources
- Mobility

Introduction

This report highlights benchmarks patterns of distribution and density for local business sectors and workforces in the Local Authority of Leeds (LAD: E08000035). Making use of the most recent comprehensive national surveys and registries with a sector-based focus, this profile breaks down the concentration and relative comparison of different industry sectors as compared to regional (Yorkshire and The Humber - E12000003) and national averages.

Data and Sources

This report is based primarily on two datasets available through the Office for National Statistics (ONS) Secure Research Service (SRS) catalogue:

Sector-Based Densities and Distributions: Leeds (Yorkshire and The Humber)

Local Data Spaces ADR UK Research Team

April 14, 2021

- Introduction
- Data and Sources
- Broad Picture of Business Activity
- Sector-Based Businesses and Workforce Distribution
- Five-Year Total Growth Rates
- Business Breakdown by Sectors
- Employment Density Profiles and Ranking
- Research Team
- Contact
- Partners

Introduction

This report highlights benchmarks patterns of distribution and density for local business sectors and workforces in the Local Authority of Leeds (LAD: E08000035). Making use of the most recent comprehensive national surveys and registries with a sector-based focus, this profile breaks down the concentration and relative comparison of different industry sectors as compared to regional (Yorkshire and The Humber - E12000003) and national averages.

This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

All counts from secured datasets less than 10 have been suppressed for security and data aggregated to the appropriate spatial or temporal dimension to ensure non-disclosivity.

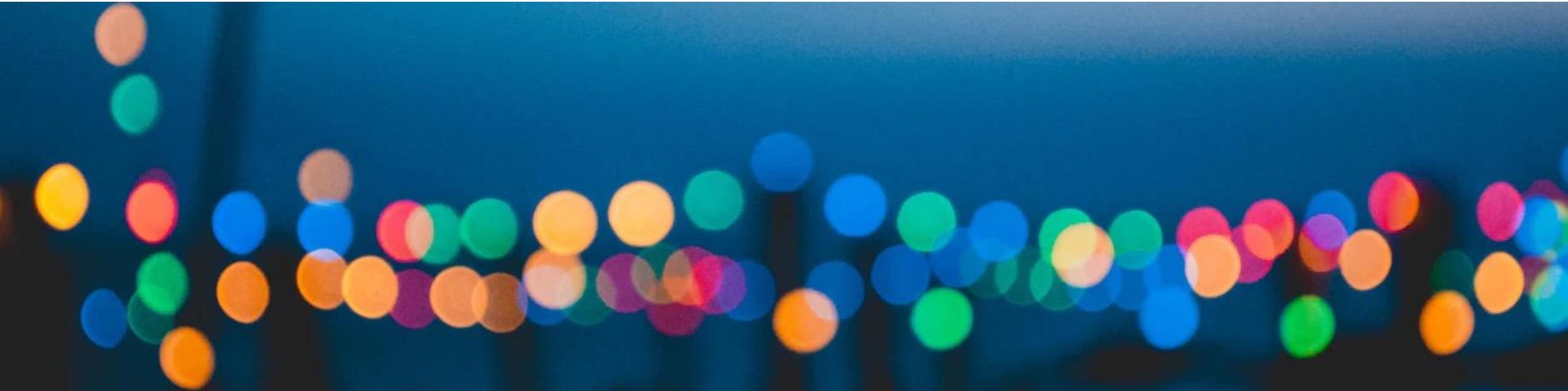
Data and Sources

Local profiling is based primarily on two datasets available through the Office for National Statistics (ONS) Secure Research Service (SRS) catalogue:

- The Business Structure Dataset (BSD): doi.org/10.5255/UKDA-SN-6697-10
- Business Registry and Employment Survey (BRES): [10.5255/UKDA-SN-7463-9](https://doi.org/10.5255/UKDA-SN-7463-9)

Local Data Spaces

Helping Local Authorities tackle COVID-19



Dr Mark Green mark.green@liverpool.ac.uk; @markalangreen



UNIVERSITY OF
LIVERPOOL