# Pass it on: School attendance and public-facing employment predict COVID-19 transmission in England

Sara Begg, Supervisor: Dr Jonathan Read

*CHICAS, Lancaster Medical School, Lancaster University. Contact: s.begg@lancaster.ac.uk, jonathan.read@lancaster.ac.uk*

## Spatial variation to identify transmission factors

**The basic reproduction number ($R_0$)**
A useful measure of transmission, $R_0$:
- Indicates how many secondary cases arise from a primary case in a totally susceptible population.
- Is influenced by biological, environmental and socio-behavioural factors.

**Factors influencing transmission**
Understanding the relative importance of factors can help modellers:
- Predict how an epidemic might progress
- Identify appropriate policies to limit transmission

**Association between population characteristics and $R_0$**
This study:
- Estimates $R_0$ for each Lower Tier Local Authority (LTLA) in England, UK, prior to March 30th 2020.
- Investigates associations between these estimated $R_0$ and population characteristics

## Estimating $R_0$ and modelling association with population characteristics

**Data Sources**
Data were disaggregated by LTLA, and included:
- daily COVID-19 case counts
- deprivation - Index of Multiple Deprivation (IMD)
- population density
- household overcrowding (% of households)
- employment in sales, care and leisure (% of pop$^n$)
- commute by public transport (% of pop$^n$)
- school attendance (% of pop$^n$)
- black and minority ethnicities (% of pop$^n$)

All data are available via the ONS service "Nomis", or the UK government website[1,2].

**$R_0$ estimation**
R package R0[3] used to estimate $R_0$ for each LTLA
- Using daily COVID-19 case counts
- Selected the exponential growth method
- Links the exponential growth rate at the start of the pandemic to the initial reproduction ratio

**Stepwise Regression model**
A best fit model was constructed using R step function:
- Stepwise, forward addition and backward elimination process, to maximise AIC score
- Weighted by the inverse of the variance of estimated $R_0$, to account for variance heterogeneity
- Repeated with and without deprivation variable (IMD)

## Public-facing employment & school attendance positively associated with $R_0$. Clusters of high $R_0$ in North of England unexplained by model

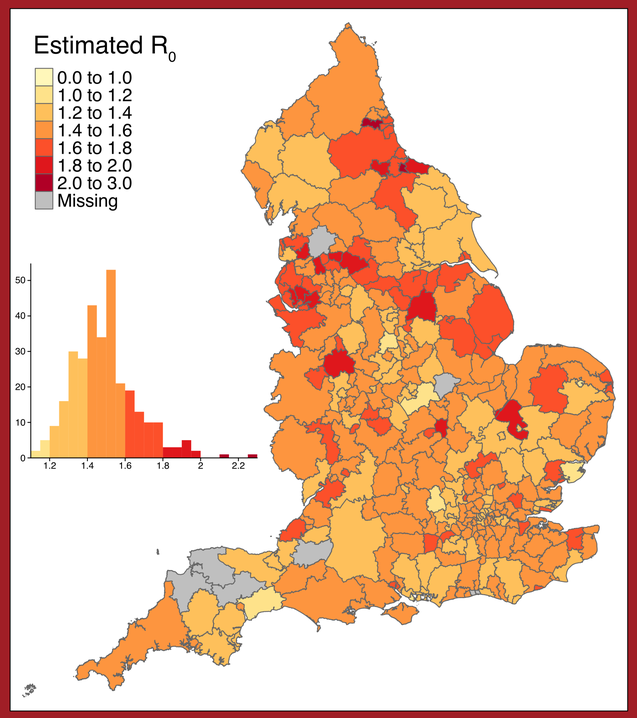


**Figure 1:** Spatial patterning of $R_0$ point estimates for LTLAs in England, using the Exponential Growth method. Histogram: distribution of $R_0$ point estimates.

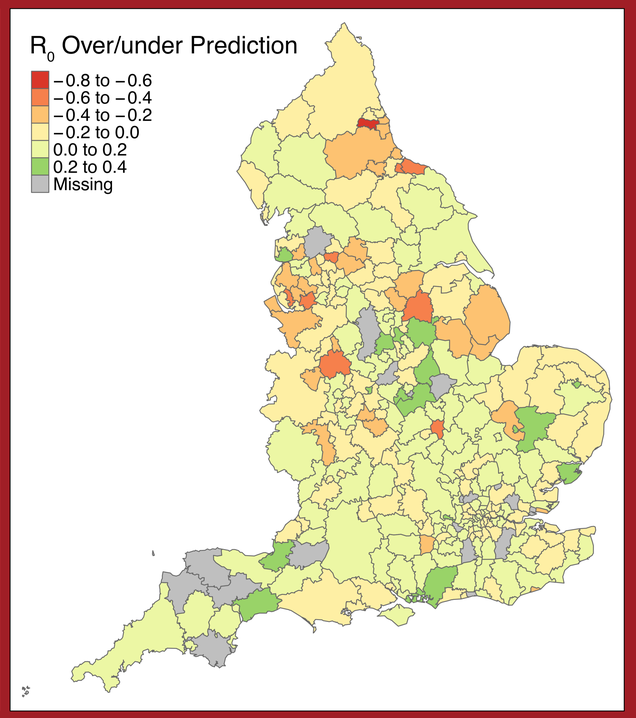


**Figure 2:** Difference in $R_0$ predicted by the models and estimated from case data for LTLAs in England, using the Exponential Growth method.

**$R_0$ estimation**
The median LTLA estimated $R_0$ was 1.48 and ranged from 1.10, 95% CI [0.88, 1.33] to 2.25, 95% CI [1.94, 2.64]
- LTLAs with a total case count of fewer than 10 by March 30th 2020 were excluded
- $R_0$ therefore estimated for 308 LTLAs (9 excluded)

Point estimates for $R_0$ for each LTLA were mapped to visualise the heterogeneity of point estimations spatially (Figure 1).
- Clustering of higher $R_0$ in the North East and North West
- Band of higher $R_0$ through South and West Yorkshire

| | | $R_0$ estimate, final model | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 1.0058 | 0.8184 – 1.1933 | **<0.001** |
| Percent BAME | -0.0014 | -0.0025 – -0.0002 | **0.018** |
| Sales, care or leisure | 0.0094 | 0.0034 – 0.0154 | **0.002** |
| School age | 0.0213 | 0.0104 – 0.0323 | **<0.001** |
| Observations | 292 | | |
| $R^2$ | 0.120 | | |

**Table 1:** Regression coefficients and p-values for final best fit model, excluding deprivation (IMD score) as a candidate variable.

**Model Construction and fitting**

*Deprivation variable included in the model:*
- Best fit model includes IMD score, % BAME, population density, public transport and school age
- 27% of $R_0$ variation could be explained by this model

*Deprivation excluded in the model (final model):*
- Significant relationship between $R_0$ and the percentages of the population that are BAME, that are employed in sales, care or leisure, and that are of school age (Table 1).
- % BAME negatively associated, % employed in sales, care or leisure, and % school age positively associated
- 12% of $R_0$ variation could be explained by the final model

*Spatial patterning of model fit:*
- Final model used to predict $R_0$ for each LTLA.
- The differences between predicted $R_0$ and estimated $R_0$ were calculated and mapped (Figure 2)
- Underpredicts from Lincolnshire through to South Yorkshire, and in the North East and North West
- Over-prediction in areas surrounding Leicester

## Occupational exposure and school-based transmission drive up $R_0$

*Occupation mix has an effect on $R_0$*
- An increase of $R_0$ by 0.01 for each percentage point increase of the population employed in sales, care or leisure
- Adds to the evidence that occupation has a role in COVID-19 transmission, even in the absence of controls
- A preliminary suggestion that this may be linked to the work environment, i.e., occupations that tend towards being public facing, requiring regular contact with many individuals[4]

*School attendance may contribute to a higher $R_0$*
- Percentage school age taken as a proxy for school attendance, suggests school-based transmission relevant
- Adds to evidence that secondary school students introduce COVID-19 to households and drive onward transmission[5]
- Study does not differentiate between primary and secondary schools, an area for future work (see below)

## Unexplained variation: deprivation in the North of England?

*Estimated $R_0$ was heterogenous*
- Higher $R_0$ in the North East and North West of England
- Higher $R_0$ in these regions not explained by employment in sales, care or leisure, or by population of school age

*Deprivation characteristics of regions of the north of England may be key*
- These regions are characterised by their relative deprivation and majority white populations
- Unexplained variation may be due to characteristics of deprivation in these regions: higher rates of unemployment and very low paid work compared to other areas of high deprivation[5] (urban South/London)
- May explain negative association with BAME: highest in London, reflecting differing character of deprivation.

## Next Steps: Contact Studies

- Studies investigating contact patterns related to occupational categories in England may prove a useful tool for modelling both this epidemic and future outbreaks.
- Exposure and transmission in the school environment merits significant attention in future work: potential role of overcrowding in primary schools and networking effect of secondary school pupils

## References

1. Nomis - Official Labour Market Statistics. (2021). Retrieved from https://www.nomisweb.co.uk
2. Official UK Coronavirus Dashboard. (2021). Retrieved from https://coronavirus.data.gov.uk/details/download
3. Obadia, T., R. Haneef, and P.-Y. Boëlle, The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Medical Informatics and Decision Making, 2012. 12(1): p. 147.
4. ONS. Which occupations have the highest potential exposure to the coronavirus (COVID-19)? - Office for National Statistics. 2020
5. Flasche, S. and W.J. Edmunds, The role of schools and school-aged children in SARS-CoV-2 transmission. The Lancet Infectious Diseases, 2021. 21(3): p. 298-299.
6. Hood, A. and T. Waters, *Living standards, poverty and inequality in the UK: 2017–18 to 2021–22*. 2017, The Institute for Fiscal Studies: London, UK.