Accounting for the Variation in the Confirmed Covid-19 Caseload across England:

An analysis of the role of multi-generation households, London and time

Peter Kenway and Josh Holden

New Policy Institute,
Can Mezzanine,
49-51 East Road,
London N1 6AH

11th April 2020

Introduction and summary

Introduction

Using data on older people who share a home with people of working age, we published a blog last week that linked this statistic with the number of confirmed cases of Covid-19 across English local authority areas. Such a link raises policy questions about the practicability of advice on social distancing for people whose housing conditions make self-isolation impossible.

This paper takes that initial finding forward in two ways. The first is to extend the statistical research by developing a model that goes beyond correlation to measure the size of the impact of various socio-economic variables on the confirmed Covid-19 caseload. Alongside the proportion of over-70s who share a home with people of working-age, these variables include overcrowding and local area deprivation.

The second is to draw out the policy questions raised by the model. Some of them relate to the policy response to the virus over the coming months. Others concern what have become norms of housing policy, especially the standards that determine what counts as adequate accommodation. These practical questions are what matter: the statistical model which highlights the underlying issue is a means to an end.

Summary of key findings

- The proportion of over-70s in a local authority area who share a household with people of working-age is confirmed to be a significant factor in accounting for the variation in the number of Covid-19 cases across England – even when levels of local deprivation, the time since the area first recorded five cases and an additional, non-specific, “London effect” are taken into account.
- In the short term, this finding poses question about the provision for people of working-age who, if they find themselves with mild symptoms, may not wish to self-isolate at home because they cannot maintain the necessary distance from other household members.
- In the medium term and beyond, the finding challenges both the norms which have come to define what constitute an adequate standard of housing, as well as local authorities’ and others’ strategic plans for the provision of homes (for example, the London Plan) which reflect those standards.

The statistical model

Data and variables

The statistical model aims to account for the variation in the number of confirmed cases of Covid-19 per head of population across 149 English upper tier local authorities (UTLAs – London boroughs, metropolitan and unitary districts and shire counties) on a single day. In this instance, that day is 5th April.

The 38,053 cases attributed to UTLAs on 5th April average 0.68 cases per 1,000 people. There is a more than tenfold variation between those with the lowest proportion (fewer than 0.2 cases per 1,000) and those with the highest (more than 2 cases per 1,000).

The preferred version of the model has four variables to account for the variation in cases between UTLAs. The four are:

- the proportion of over-70s in the UTLA living with people of working-age;
- an index of deprivation across the local area (the IMD average score);
- a variable set to 1 for London boroughs and 0 for others to pick up London specific effects;
- the number of days that have elapsed since the UTLA recorded its fifth case.

Each variable is “statistically significant”. This means that when it’s included in the model, the match between the actual and modelled Covid-19 caseloads improves by at least the minimum amount which statistical norms deem necessary.

A variable measuring the number of overcrowded households in the local area was assessed for inclusion in the model but discarded because it was not statistically significant. This does not mean it is unimportant or plays no part in the story – indeed, on its own overcrowding is statistically significant – but that it does not add enough once over-70s sharing is included.

The technical appendix provides more details.

The importance of the model

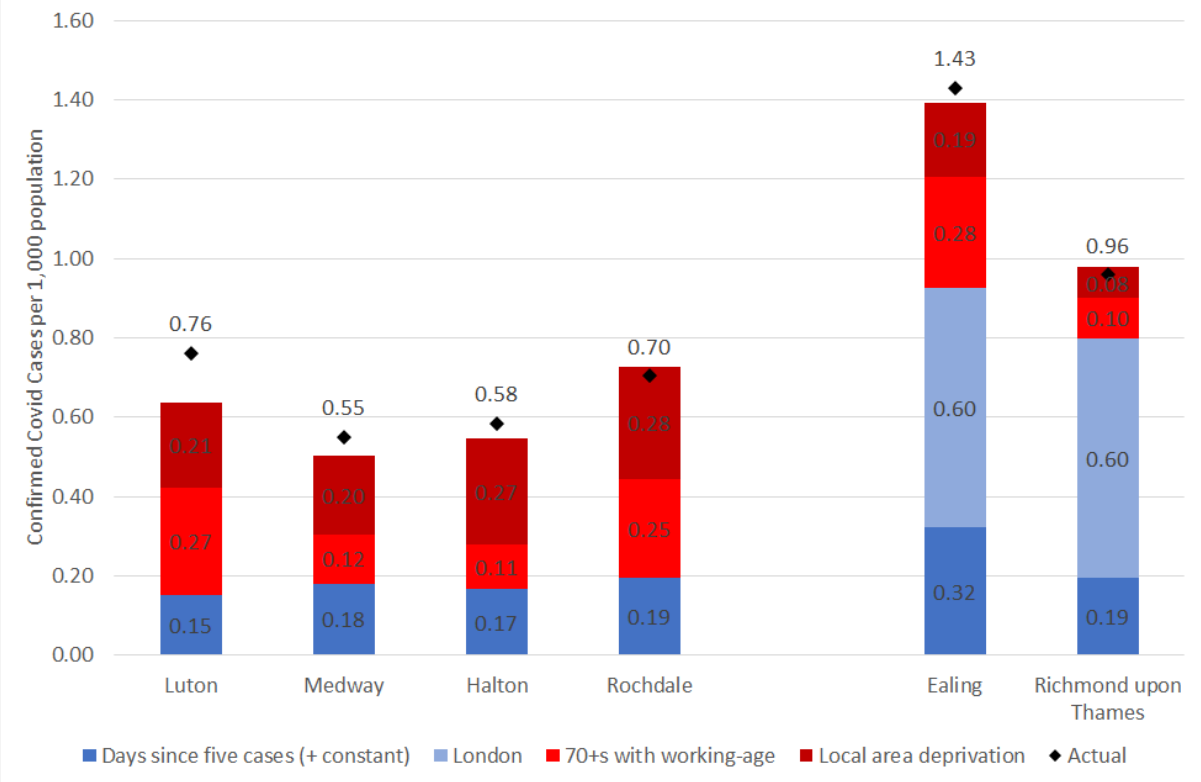
The model shows that the statistical link between the Covid-19 caseload across England and the proportion of over-70s sharing a home with people of working-age persists even after deprivation, the passage of time and other unspecified London effects have been taken into account. The simple correlation reported previously is not just a spurious one. The importance to Covid-19 of households where pensioners and working-age live together, especially in areas of high deprivation, is affirmed.

The results in detail

The scale of the effects: individual local authorities

Figure 1 shows the scale of the impact of each of the four variables using results from the model for six UTLAs – four outside of London and two London boroughs. The dot on the graph for each UTLA shows the actual Covid-19 caseload per 1,000 people per UTLA. The bar shows the caseload as predicted by the model, with the contribution from each of the four components (over-70s, deprivation, London and time) shown separately. There is always a gap between the actual and the predicted: here, the model underpredicts for four UTLAs and overpredicts for two.

Figure 1: actual and modelled impacts (six example local authority areas)



In turn:

- For Luton, the model predicts a total of 0.64 cases per 1,000, built up as the sum of: 0.15 due to the number of days since five cases were first recorded; 0.27 due to the proportion of over-70s living with working-age; and a further 0.21 due to local deprivation.
- For Medway, the model attributes a similar amount for deprivation (0.2) reflecting the fact that its underlying deprivation level is similar Luton. It shows a much smaller amount for the over-70s (0.12), reflecting the much lower proportion of over-70s living with working-age. The time component (0.18) is slightly larger because it reached five cases two days before Luton.

- Halton and Rochdale get similar amounts for deprivation (0.27 and 0.28), more than Luton and Medway because underlying deprivation is higher. The amounts for the over-70s are very different (0.11 and 0.25) reflecting Halton's low and Rochdale's high proportion of over-70s living with working-age.
- The two London boroughs show the size of the London effect (0.6). Ealing looks like Luton and Medway for deprivation and like Luton and Rochdale for over-70s living with working-age. Richmond is very low for both. Ealing has a large time component (0.32), having reached five cases as early as 9th March.

The scale of the effects: over-70s, deprivation and time

Another way to see the effect of the model is to compare the modelled caseload for UTLAs who have high values of a particular variable with the caseload for UTLAs who have low value for that variable. Table 1 does this using the top fifth and the bottom fifth of UTLAs (30 in each group) for each of three variables.

Table 1: modelled impacts for UTLAs with high and low values of the variables

	Local area deprivation	*Over-70s living with working-age*	*Days since five cases*
Average value of the variable for the 20% of UTLAs with the lowest values	12.4	10%	13
Average value of the for the 20% of UTLAs with the highest values	34.7	28%	27
Modelled caseload difference (highest minus lowest) per 1,000 population	0.184	0.203	0.193

The interpretation is as follows:

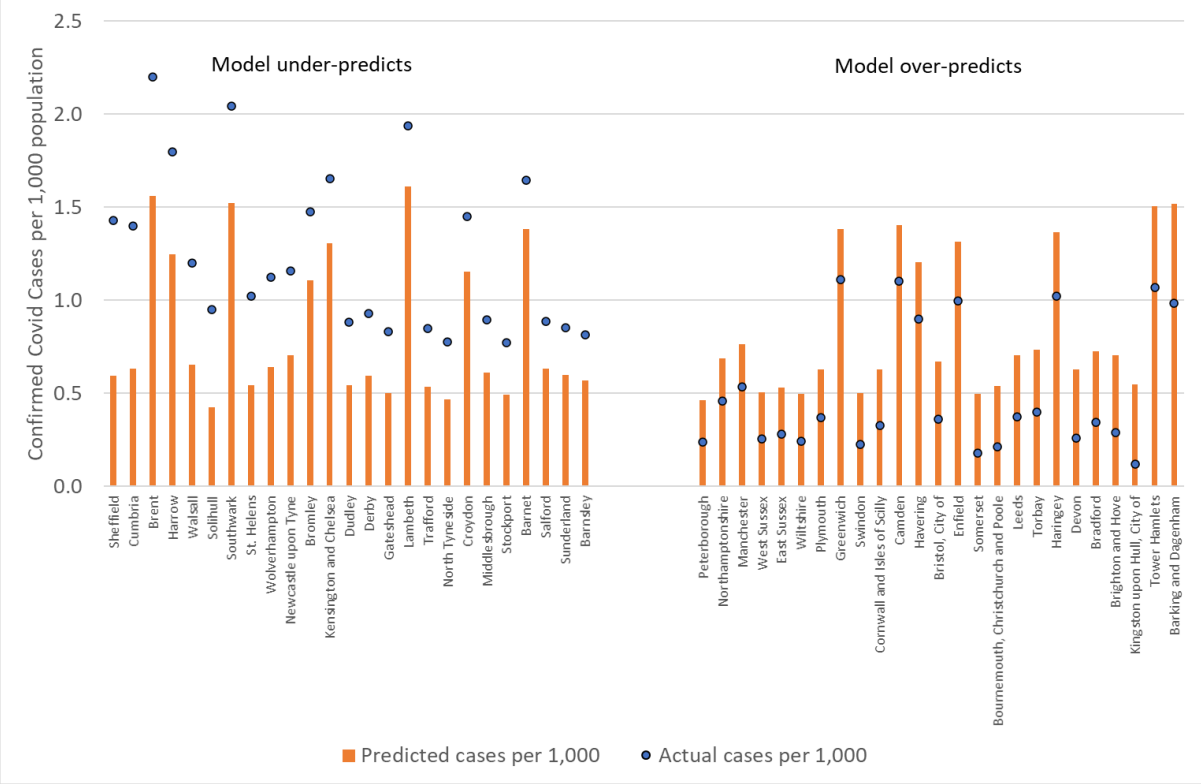
- Deprivation: the most deprived UTLAs are almost three times as deprived as the least on this measure (34.7 compared with 12.4). According to the model this difference in deprivation would add 0.18 extra cases per 1,000 population.
- Over-70s: UTLAs with the highest proportion of over-70s living with working age have almost three times as many as those with the lowest (28% compared with 10%). This difference adds 0.2 extra cases per 1,000.
- Days: UTLAs who've had five or more cases the longest reached that number two weeks before those who have had at least five cases for the least time. According to the model these two weeks add an extra 0.19 cases per 1,000.

For the average UTLA, with about 375,000 inhabitants, these numbers add around 70 extra cases for any of the three variables.

Some further questions

The model is good enough to establish that these three factors play a significant role in accounting for the variations in the caseload, but there is always scope for improvement. Figure 2 highlights several issues. The 25 UTLAs on the left of figure 2 are those where the model underpredicts the caseload by the largest amount. The 25 UTLAs on the right are those where it overpredicts by the largest amount.

Figure 2: UTLAs where the model under- or over- predicts the most



One feature of figure 2 is that London boroughs appear on both left and right, some with actual caseloads way above predicted levels (e.g. Brent and Southwark) and others with actual caseloads way below (e.g. Tower Hamlets and Barking and Dagenham). The London variable lifts the predicted values for all boroughs, but it cannot deal with the variation within London. There is clearly more to do here.

The model also falls short of accounting for the high caseloads in the West Midlands (Walsall, Wolverhampton, Solihull, Dudley) as well as parts of the North East and Greater Manchester. If there is a pattern here, it may point to a limitation of a model which treats UTLAs as isolated entities when they are part of an agglomeration.

The location of hospitals may matter. Confirmed cases are allocated to UTLAs according to the patient's postcode but hospitals themselves may "cause" cases by virtue of their testing regime. It has been suggested that this is a factor in Sheffield, whose Northern General hospital has a large testing capacity which has been used

to test its workers and patients, as well as those in the area known to have come home from infected areas. In any individual UTLA, a higher confirmed caseload may not reflect a higher incidence of Covid-19 in the population as whole.

The policy questions raised

This research confirms that there is a statistical link across local authority areas between the confirmed Covid-19 caseload and the proportion of households where pensioners and working-age live together, especially in areas of high deprivation. While research like this can always be further refined, there is enough evidence for policy makers to start addressing this link and considering the practical implications.

Short term implications, related to the management of Covid-19

There are several short-term implications. First, policy makers should consider what alternative accommodation could be made available for people with mild symptoms who feel they cannot safely self-isolate at home. Although in theory anyone could be in this position, it is especially likely to apply to someone who shares their home either with an elderly relative or someone in another high-risk group.

It should be stressed that we are not suggesting that some elderly and/or other vulnerable sharers should be moved out of their home. What we are suggesting is that policy makers should consider providing somewhere between the extremes of home and hospital for people with mild symptoms to self-isolate if that is necessary to protect other members of their household. Even if this isn't practical during the present lockdown, it can be considered for a second wave, or as part of a wider containment strategy. For example, French media have reported that such an idea has been proposed by the Mayor of Paris as part of a plan for exiting lockdown.

Second, responsibility for accommodation – housing – within English local government sits with lower tier rather than upper tier local authorities. London boroughs and metropolitan and unitary districts combine upper and lower tier functions but in the English counties, it is the shire districts who are the lower tier. Are lower tier functions, especially housing, being adequately funded by central government? It is a mistake to make national policy on the assumption that Covid-19 is mainly an urban problem. For example, one detailed finding here (figure 2) is that Cumbria has an unexpectedly high number of Covid-19 cases.

Third, general advice from the Government must take account of the variation in people's personal and household circumstances. Even if responses to Covid-19 can do little about this in the short term, policy makers, especially at the national level, must strive to take this variation into account and make provision for exceptions where necessary.

Longer term implications, related to housing standards

As commentators are now observing, the lockdown is revealing the inequality in our housing. As well as the importance of multi-generational households, this research shows there is a clear link between local area deprivation and Covid-19 cases.

The practical implications of a link to local area deprivation (an abstract concept) are unclear. Low income, poor health and poor housing are all associated with deprivation but details matter when designing concrete actions. Over the past 20 years, a large body of research has been built up into the causes and consequences of poverty and deprivation in the UK. The practical conclusions of this research now need to be considered anew by policy makers at every level of government and in every part of government.

Unless we want to pretend that Covid-19 is a one-off event, remedying the conditions that encourage the spread of such diseases has to be a public priority.

Turning to housing, there is already enough evidence in this research to conclude that the standards which have come to determine what constitutes adequate housing now need to be abandoned and made anew.

At their heart lies the bedroom standard, a formula for the number of bedrooms a household is supposed to need based on the age, sex and adult relationships of the people in it. This formula influences everything from the amount of social security payable to support housing costs to the core calculations of the capital's long term (20 year) housing requirement in the London Plan. The Government's response to Covid-19 reveals the double inadequacy of this standard, first because it allows for too few bedrooms to allow people to self-isolate safely and secondly, because it is just about bedrooms when bathrooms, washing facilities and so on matter as well.

Policy around housing must also recognise the variety of housing arrangements that people choose to have. In identifying multi-generational families as a focus of concern around Covid-19, we are not decrying the arrangement but bemoaning the conditions under which it may happen. The pressure on adult social care services should push housing policy towards enabling older people to live comfortably with others – if they wish. Larger homes are more flexible. Policy should stop talking about housing in terms of "units" which reduces it to a simple game in which the maximum number of one and two bedroom "units" inevitably emerges as the answer.

Finally, policy makers (and researchers?) should stop treating work and housing as separate issues. Covid-19 points to the connection between key workers and poor housing. The current crisis has reinforced the crucial role played by key workers and it is vital that their housing needs are taken fully into account when considering their pay, status, training, progression and representation at work.

Technical appendix

The general model

The dependent variable is the number of confirmed Covid-19 cases per 1,000 people in each UTLA (via the daily dashboard, downloaded on 8th April 2020) . The specific model described in the paper emerges from a more general model containing six independent variables. They (along with their data sources) are:

- The proportion of people aged 70 or over in each UTLA who share their home with one or more people of working-age.
- A measure of deprivation in the UTLA (the IMD average score)
- A London dummy variable (set to 1 for London boroughs and zero for the rest)
- The UTLA population (via Nomis).
- The proportion of households in the UTLA defined as overcrowded in the 2011 Census (via Nomis).
- The number of days since the UTLA recorded at least 5 cases up to 5th April (computed from the daily dashboard above).

The model was estimated on the confirmed UTLA caseload data for 5th April 2020 as published on 8th April 2020.

Correlation coefficients between variables

	Cases per head	*Over-70s*	*London*	*Depri-vation*	*Over-crowding*	*Population*	*Days from 5th case*
Cases per head	1						
Over-70s	0.62	1					
London	0.75	0.59	1				
Deprivation	0.10	0.15	-0.10	1			
Overcrowding	0.70	0.81	0.76	0.22	1		
Population	-0.14	-0.03	-0.18	-0.21	-0.16	1	
Days from 5th case	0.40	0.32	0.35	-0.18	0.29	0.50	1

London (0.75), the over-70s share (0.62) and overcrowding (0.7) all have strong positive correlations with the number of cases. The over-70s share and overcrowding have similar relations to case numbers and also have a very strong correlation between them (0.81). Overcrowding is also strongly correlated with the London effect.

Test statistics

The general model was reduced from six to four independent variables by deleting the population and overcrowding variables (based on their individual t-statistics and with an F-test statistic for the joint hypothesis 0.16 against a critical value of 3.07).

The reduced model was tested for omitted variable bias and heteroskedasticity using RESET tests (F-test statistics of 0.95 and 0.69, against a critical value of 2.68).

Summary output for the reduced model

Dependent variable: the number of confirmed Covid-19 cases per 1,000 population

	Coefficients	*Standard Error*	*t-statistic*
Intercept	-0.072	0.121	-0.60
Ove-r70s sharing	1.128	0.383	2.95
London	0.603	0.065	9.24
Deprivation	0.008	0.003	3.00
Days since 5th case	0.014	0.005	2.98

Regression Statistics: TSS=26.028; RSS=9.270; multiple R=0.80; R^2=0.64; adjusted R^2=0.63; standard error=0.25; N=149.