Preliminary analysis of PHE Care Home data

Ian Hall and Hugo Lewkowicz, University of Manchester

April 2020

Summary

Using data within PHE's HPZone database the number of new reported outbreaks of COVID19 in care homes rose rapidly in early to mid-March. Since late-March the number reported has plateaued (but not showing evidence of reduction), Figures 1 and 2. This data is reported by Health Protection Teams.

The early doubling time is consistent with the doubling time seen in other measures of disease surveillance at that time.

Using an 'SIS' model for care home outbreaks (Figure 3 and Table 1) we find

1. a natural conclusion is that with no change in disease transmission in future we might expect at least 90% of care homes to report at least one case eventually if current reporting trends are maintained (currently about 20% have reported such). Given social distancing (blue lines on plots) timing it does not appear to affect these outbreaks.

2. However, it may be that the vehicle of connecting care homes is the staff and staff seem to be suffering disease at similar number to residents (though reason for staff absence is unclear in the data and may be that staff are absent for precautionary reasons). If staff work in multiple care homes then these high attack rates may lead to depletion of susceptible staff and so reduce transmission in time. Moreover, staff interact with households and community and so infection can be passed to and from care homes in this manner. This is uncertain and so should not be factored into planning yet (on the basis of this email alone that is).

The early growth in the data that drives these results may be an artefact of surveillance (a relatively new scheme having improved uptake in usage over time) rather than epidemiological.

This data and analysis says nothing about the size of outbreaks within care homes. There is no data on ongoing outbreak size and management. The source HPZone files shows some care homes are suffering relatively high attack rates when they are reported, 4 (and note that these are likely still under progression with care home). At present previous estimates of expecting high within care home attack rate remains highly likely.

Data

Two data sets have been made available.

Firstly data showing the number of new care homes reporting an outbreak of at least one case each day by PHE Centre.

Further data is available for some care homes giving the outbreak status at the time of report (number of case etc.) but there is no simple mechanism for updating this information at present so a dynamical model within care homes is of limited use.

Secondly we have data from CQC on all cause mortality. This is not COVID19 specific.

Neither dataset is specific to individual

GAM analysis

Using a GAM to look at trend over time (using mgcv with quasipoisson link) in both the NW, Figure 1, and England as a whole, Figure 2, shows signs of plateau in reported outbreaks in past 10 days. For the first 20 days or so there was rapid growth with a 3-4 day doubling time.

England as a whole seems to be more erratic, perhaps a sign of dips in reporting at weekends. At smaller spatial units such as regions these weekend reporting artefacts are lost in the noise associated with signal.

Running the GAM approach on the CQC all cause death data shows no strong evidence of plateau in West Midlands, North East, North West and London. Other regions have weak evidence of having zero growth rate.

Modelling

We thus assume care homes are in two states, either showing no cases of infectious disease or with some cases detected. Given the number of care homes N is fixed we can consider just the number of care homes reporting at lease one case.

The number of care homes with at least one case, I say, will increase by further cases being detecting in previously unaffected care homes and will decrease when laboratory confirmation arrives that the infected cases are not COVID-19 (fast timescale of a couple of days) or when the care home outbreak is declared over (if outbreak is over a couple of generations and allowing for enhanced surveillance after final case an outbreak may last 3-4 weeks). Once declared infection free a care home may be reinfected later in pandemic.

Assuming an average rate γ can be constructed as a composite of the fast acting timescale of laboratory confirmation and slower timescale of outbreak cessation then $I \to I - 1$ at some rate γI. The increase in I is harder to model accurately. We assume there is a between care home infection rate β so that $I \to I + 1$ with rate $\Lambda(N - I)$ where N is the number of care homes in total. A key question is what is an appropriate form of Λ

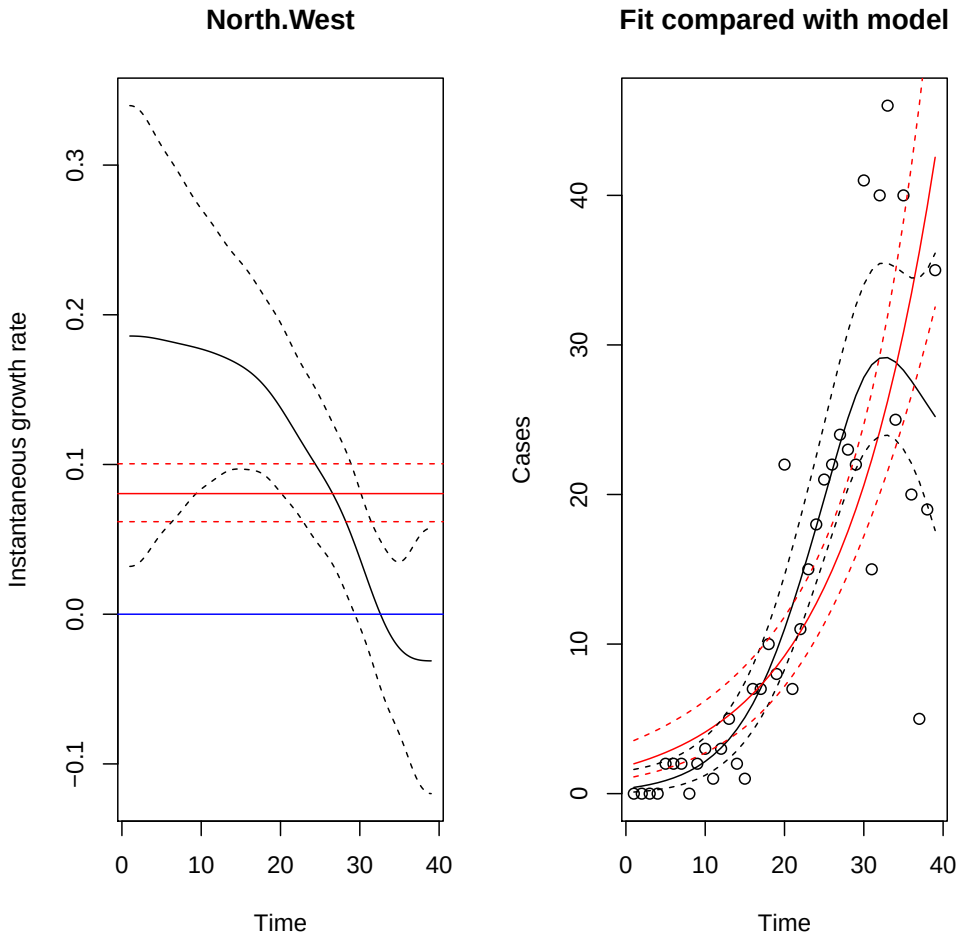

Figure 1: Left panel: Instantaneous growth rate in number of reported care homes in North West from GAM (black) and GLM (red); Right panel: reported number of care home outbreaks (dots), GAM best fit curve (black line) and CI (dashed lines), and comparison to GLM (red). Time zero is 07/04/2020.

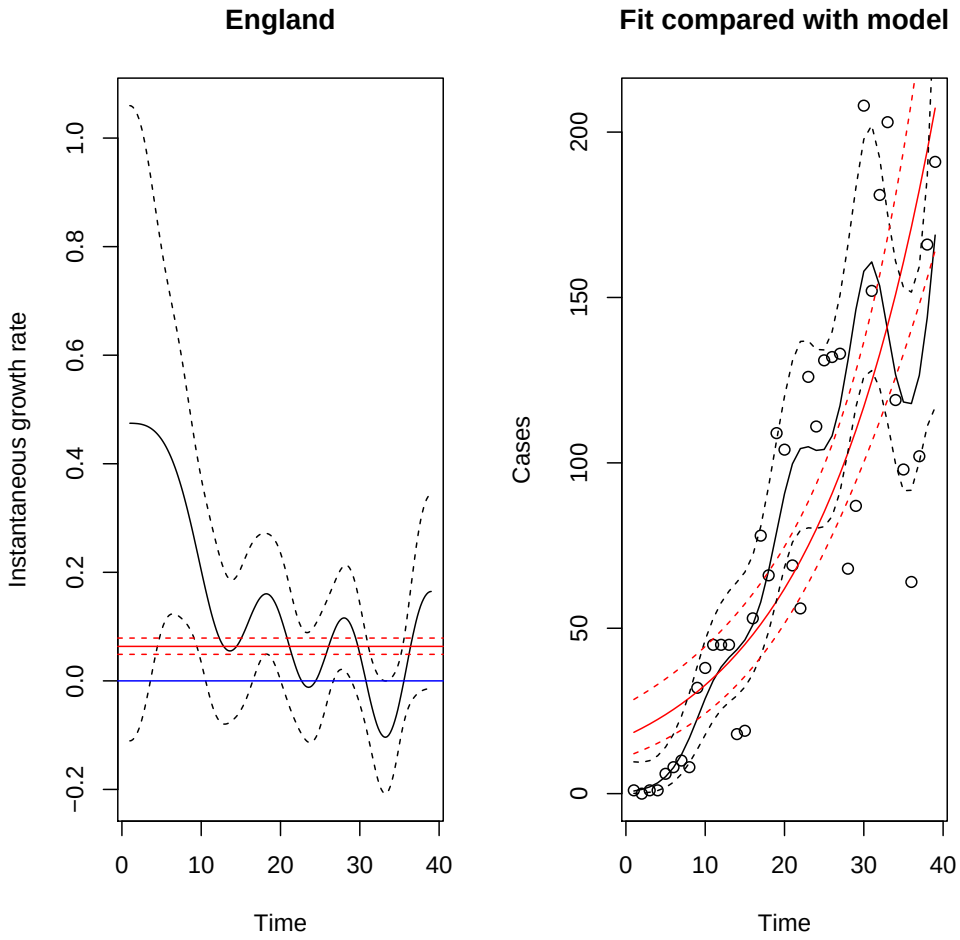

Figure 2: Left panel: Instantaneous growth rate in number of reported care homes in England from GAM (black) and GLM (red); Right panel: reported number of care home outbreaks (dots), GAM best fit curve (black line) and CI (dashed lines), and comparison to GLM (red)

$\Lambda = \lambda$

If the force of infection is constant then we may derive an ordinary differential equation (ODE) of the form

$$\dot{I} = \lambda N - (\gamma + \lambda)I$$

which has solution

$$I = I_0 e^{-(\gamma+\lambda)t} + \frac{\lambda N}{\gamma + \lambda}(1 - e^{-(\gamma+\lambda)t}) = A - (A - I_0)e^{-\mu t}$$

which will tend to a constant $A = \lambda N/(\gamma + \lambda)$. However, the new reported affected care homes $I_{\text{new}} = \lambda(N - I)$ and whilst the data is currently apparently constant the early reported show signs of growth differnt to hat predicted here, suggesting this is not a credible model (assuming that early growth is epidemiological rather than surveillance based).

$\Lambda = \beta I$

The force of infection Λ is not simply a constant. We can assume that the force of infection is proportional to the number of nearby care homes currently suffering an outbreak ($\lambda = \beta I$). This is an approximation, clearly a physical home cannot infect another, but staff working between homes, staff becoming infected from community or infected visitors (prior to lockdown) may pass infection.

The model constructed above can be written in deterministic ODE form and is the usual SIS form, this means we can write down an explicit solution for I of the form:

$$I = \frac{(R_{CH} - 1)N}{R_{CH}\left(1 + \exp(-\gamma(R_{CH} - 1)(t - t_0))\right)}$$

where $R_{CH} = \beta/\gamma$ and $t_0 = (\ln(R_{CH}(N - I_0) - N) - \ln(R_{CH}I_0))/(\gamma(R_{CH} - 1))$ controls behaviour of equation initially (assuming the number of infected care homes at time zero is I_0) such that when $t = t_0$ we have I half the value it will attain eventually. We note that I can be written in the form

$$I = \frac{pN}{1 + e^{\mu_1(t_0 - t)}}$$

where $\mu_1 = \gamma(R_{CH} - 1)$, $p = \left(1 - \frac{1}{R_{CH}}\right)$.

When $R_{CH} \gg 1$ this would suggest simple logistic regression is suitable or when $t_0 \gg t$ we can assume exponential growth (log-linear regression).

However the data presented shows new care homes reported each day with no information about cessation of outbreaks. Thus in SIS framework this data would be those entering the I state namely

$$I_{\text{new}} = \beta I(N - I)/N = \frac{q + \mu_1 e^{\mu_1(t_0 - t)}}{(1 + e^{\mu_1(t_0 - t)})^2} N$$

Region	μ_1	qN	t_0	R_{CH}	$1/\gamma$	Attack ratio
EM	0.26	11	9.2	22	136	0.97
EE	0.23	14	10.6	36	115	0.96
London	0.18	11	5.0	23	120	0.96
NE	0.27	6	12.2	34	121	0.97
NW	0.17	28	14.7	11.7	62	0.91
SE	0.39	16	6.7	73	184	0.98
SW	0.15	14	16.8	22	139	0.95
WM	0.19	15	14.8	21	107	0.95
YH	0.09	13	19.1	10.3	103	0.90
England	0.17	121	10.7	22	122	0.95

Table 1: Care home model outputs

where μ_1 and t_0 are defined as above and $q = p\gamma = \mu_1/R_{CH}$ and we can extract the final attack ratio $p = q/\gamma = 1 - 1/R_{CH}$.

We can fit this to the data provided, and the output for each region is shown in Table 1 and for the worst affected region (North West) and England as a whole in Figure 3.

The 'recovery' rate is about 4 months (120 days) on the whole – this seems long as an an outbreak may be expected to last 1 months in a care home of about 50 residents, and that some of these outbreak will be caused by other infections). The R_{CH} estimates of about 20-30 are very high, but this is not a traditional disease model the units being care homes not individuals. Given the 4 month recovery rate this suggests 5-6 care homes are affected each month for each current home reporting outbreaks. Whilst larger than expected these 'disease' parameters are outputs from the model for the parameter fits based on data. Uncertainty quantification will not show a move to different area of parameter space.

Attack rates at time of report of outbreak

We can extract from the data the number of staff and residents affected at the time of report and the number of total residents and staff and so calculate the attack rate are time of report (Figure 4). This shows fairly established outbreaks at time of report with similar disease burden in staff as residents.

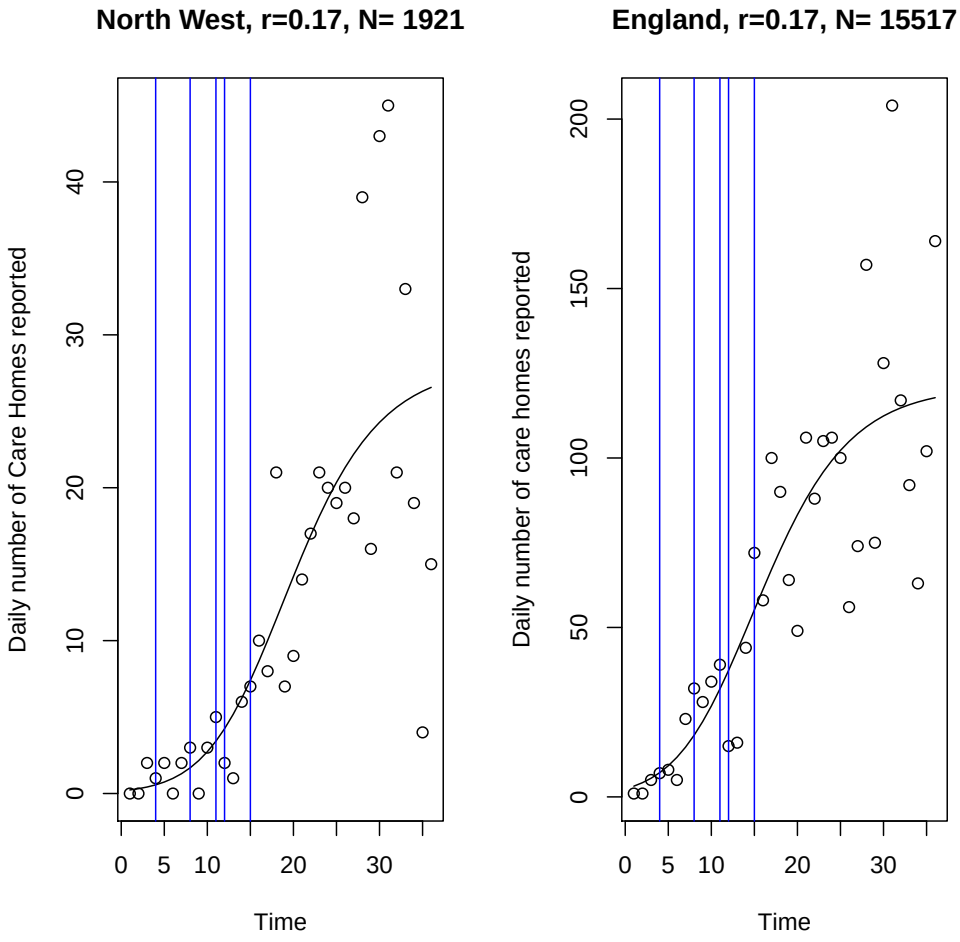

Figure 3: Number of care homes reporting COVID19 outbreaks of at least one case each day: left, North West and right, England, Time zero is 07/04/2020.

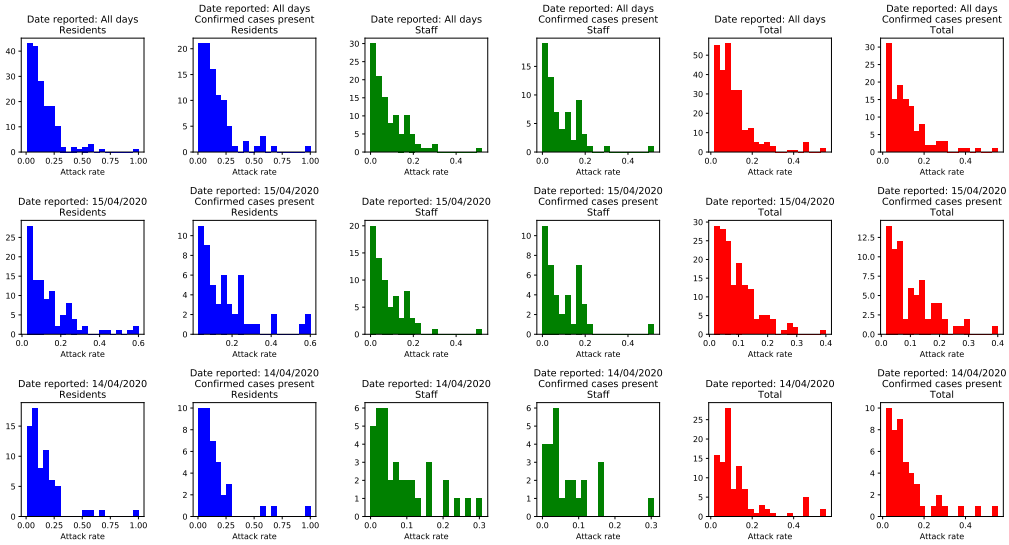

Figure 4: Attack rates in residents and staff at the time of outbreak report for those reported on 14th April and 15th April.