Inferring the effective fraction of the population already infected with Covid-19 by comparing rates in different regions of the same country.

Robert Thorne

May 22nd 2020

University College London

I use a simple model for the spread of Covid-19 in a large population.

I compare the relative decay of the number of deaths per day between different regions in Italy, Spain and England, each applying in principle the same social distancing procedures across the whole country.

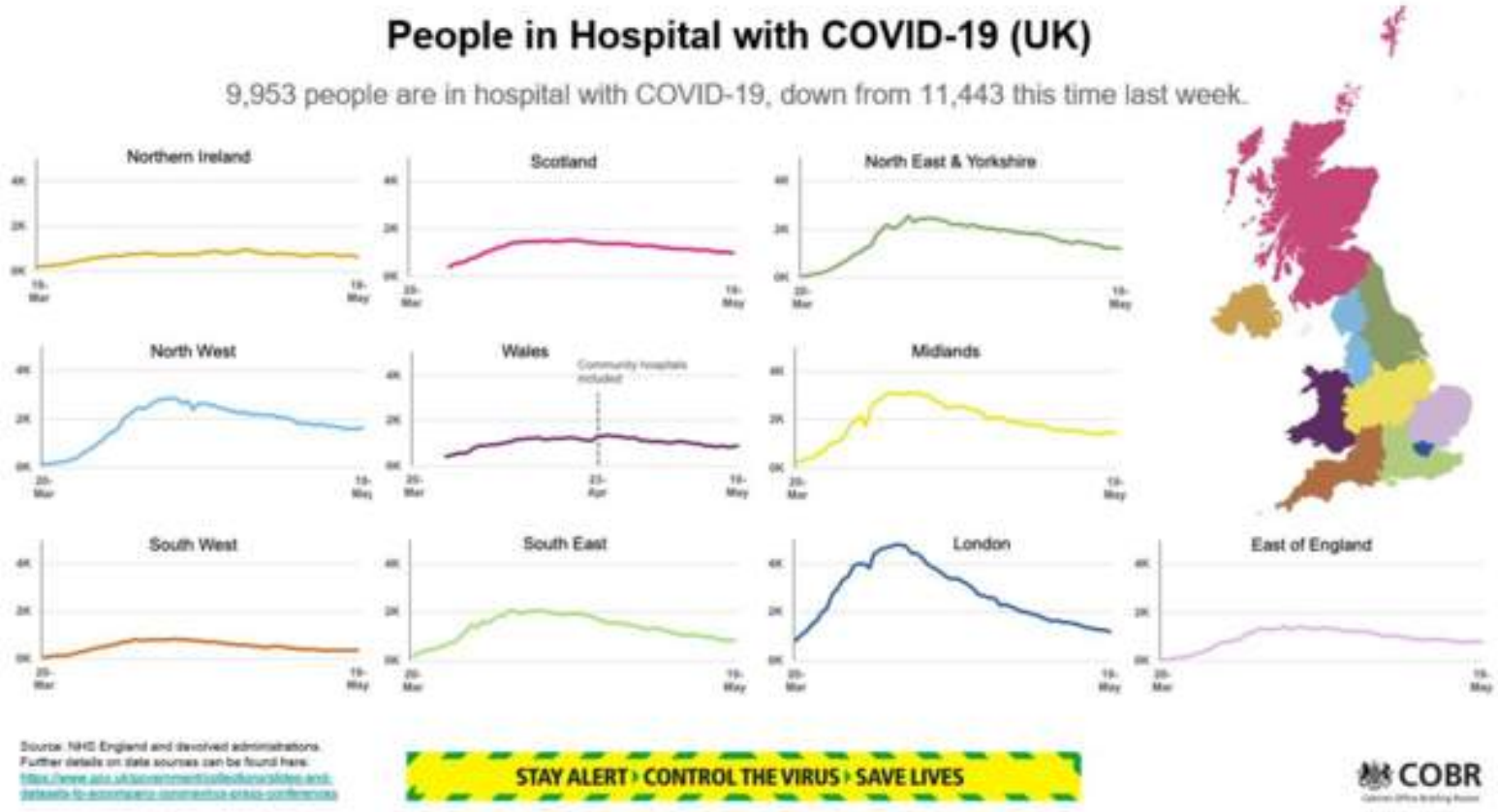
I obtain an estimate of the total fraction of the population which has already become infected. In the most heavily affected regions, Lombardy, Madrid and London, this fraction is higher than expected, i.e. ≈ 0.3.

This can be converted to a determination of the infection fatality rate ifr, which appears to be $ifr \approx 0.0025 - 0.005$, somewhat lower than usually assumed ~ 0.01 (arXiv:2005.00495 - posted 30th April).

Alternatively, this can also be interpreted as a effectively larger fraction of the population than simple counting would suggest if there is a variation in susceptibility to infection.

The implications are very similar for either interpretation or for a combination of effects.

The Motivation.



It is clear that the initially most heavily infected regions in a variety of countries, including the UK (or for practical reporting reasons England) have been having much faster falling rates of covid infection ever since the peak was reached, and well after “lockdown”, when everything should be the same throughout the countries.

The Model

I consider a basic deterministic model for the spread of Covid-19, which is arguably appropriate when a significant fraction of a population has become infected, and the system can be treated as continuous rather than discrete.

This has an analogy in Physics where one can discuss interactions between nearby atoms, spins etc. in the “mean-field” approximation.

By comparing the relative decay of the number of deaths per day between different regions, each applying equivalent social distancing procedures, one can obtain information about the total fraction of the population which has become infected.

The evolution of the system is written using the equations and conventions in e.g. doi.org/10.1101/2020.03.24.20042291.

$$\frac{dy}{dt} = \beta y(1-z) - \sigma y, \qquad \frac{dz}{dt} = \beta y(1-z).$$

y is the fraction of the population who are currently infectious

z is the fraction who are no longer susceptible

$1/\sigma$ is the infectious period in days

β is the transmission coefficient, $\beta = R\sigma$, R is the reproduction number.

At t_0 a significant fraction of the population have become infected.

Treatment of y and z as continuous variables is appropriate.

Also when R is approximately 1, or less – a very slow rate of growth for y, or even a slow decay.

$$\left(\frac{dz}{dt}\right)_t = \beta(1-z)y_0 \exp((R(1-z)-1)\sigma t).$$

At t the rate of deaths is the infection fatality rate ifr times $(dz/dt)_{t-20}$, asssuming a delay of 20 days between infection and death.

This is not constant, and there should be a convolution of (dz/dt) with a function with mean $t \approx 20$ and a width.

I use the simplest model here (less important for almost constant rate than rapid growth or decay). I define $\tau = t - 20$.

Normally assumed that z is sufficiently small that it is having negligible effect.

Whether there is growth or decay is governed entirely by whether R is greater or less than 1.

ifr is also usually taken to be a fixed value close to 0.01 https://doi.org/10.1016/S1473-3099(20)30243-7.

By comparing the relative rate of deaths of regions within the same country one can infer the value of z, and (possibly) the value of ifr.

Two regions subject to exactly the same social distancing procedures, so assume that R has a common value $R \approx 1$.

The fall in death rates is mainly due to $(1 - z)$.

Taking the ratio of $(dz/dt)_\tau$ for the two regions

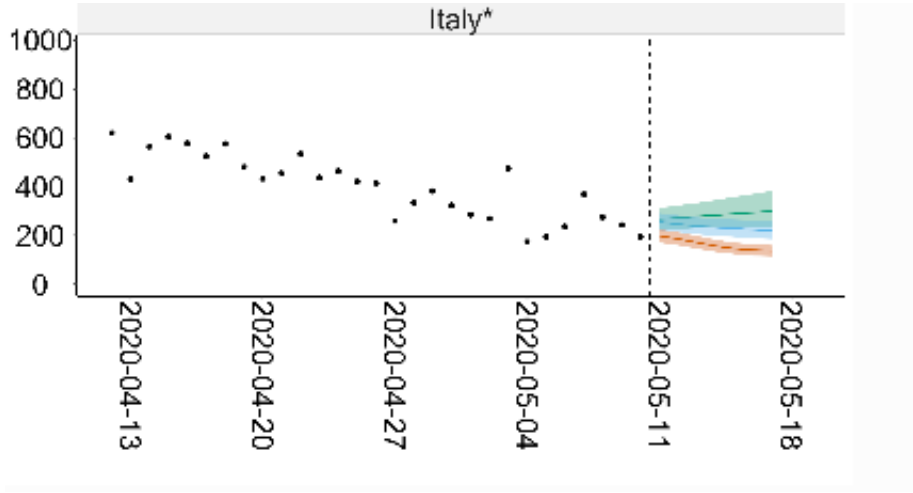
$$\left(\frac{dz_1}{dt}\right)_\tau \Big/ \left(\frac{dz_2}{dt}\right)_\tau = R_{12} \quad \propto \quad \frac{\exp((R(1-z_1)-1)\sigma\tau)}{\exp((R(1-z_2)-1)\sigma\tau)}$$

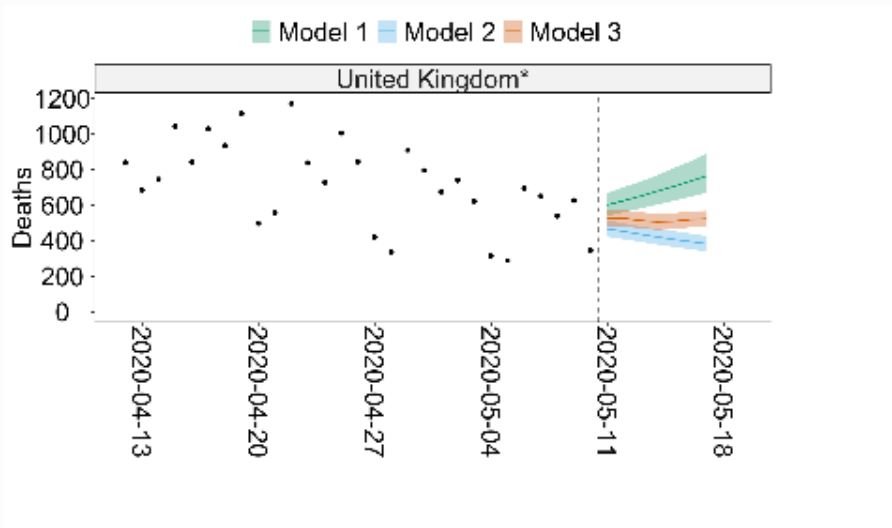
$$\rightarrow R_{12} \quad \propto \quad \exp(-R(z_1 - z_2)\sigma\tau).$$

If at time t_0 then if $z_1 > z_2$ then R_{12} will fall with time.

If the rate of decay of the number of deaths in two regions is clearly different, then assuming R is very similar in each, the only explanation is the effect of the differing values of z.

Treatment approximate, but consider degree of uncertainty in most sophisticated models.

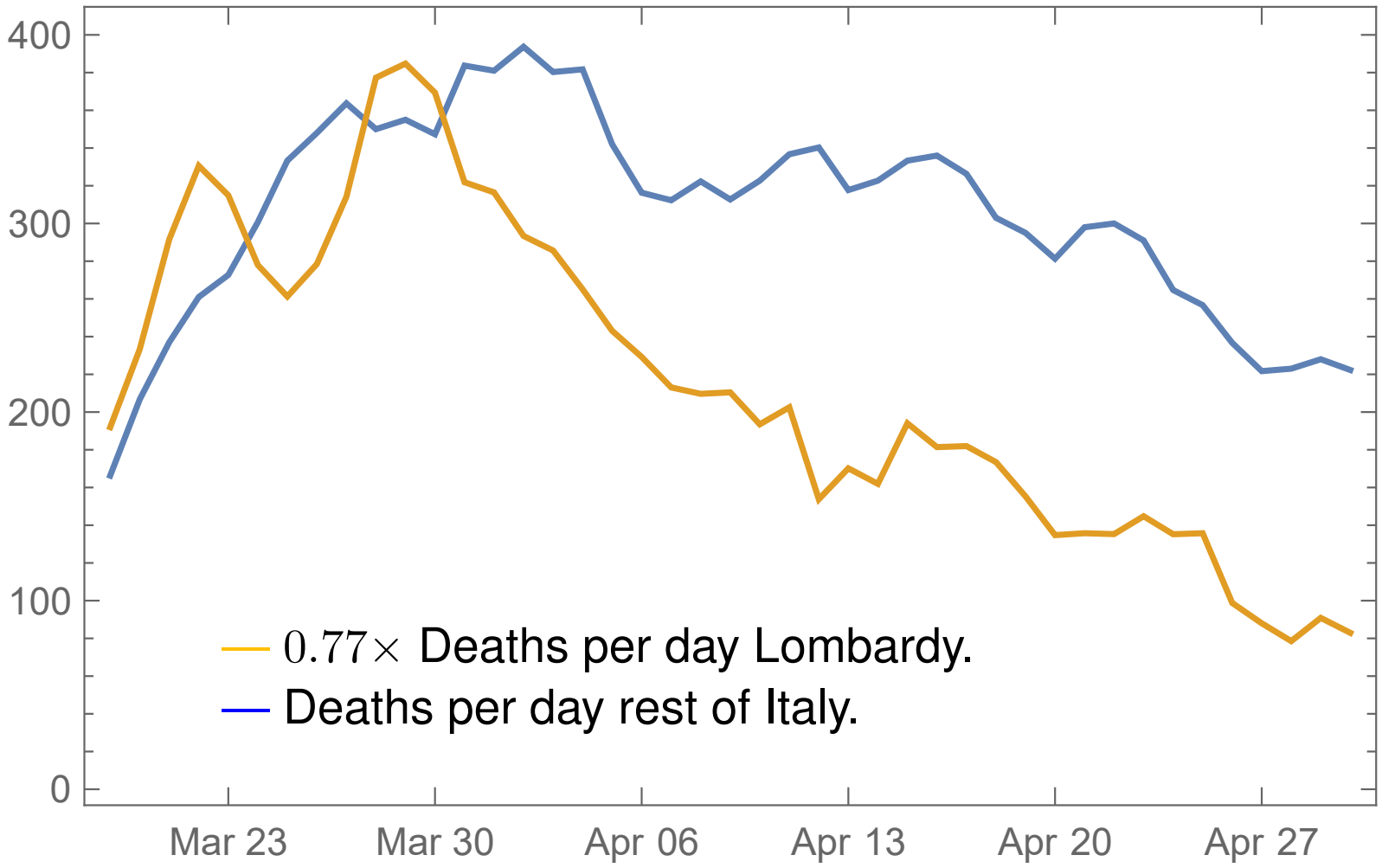




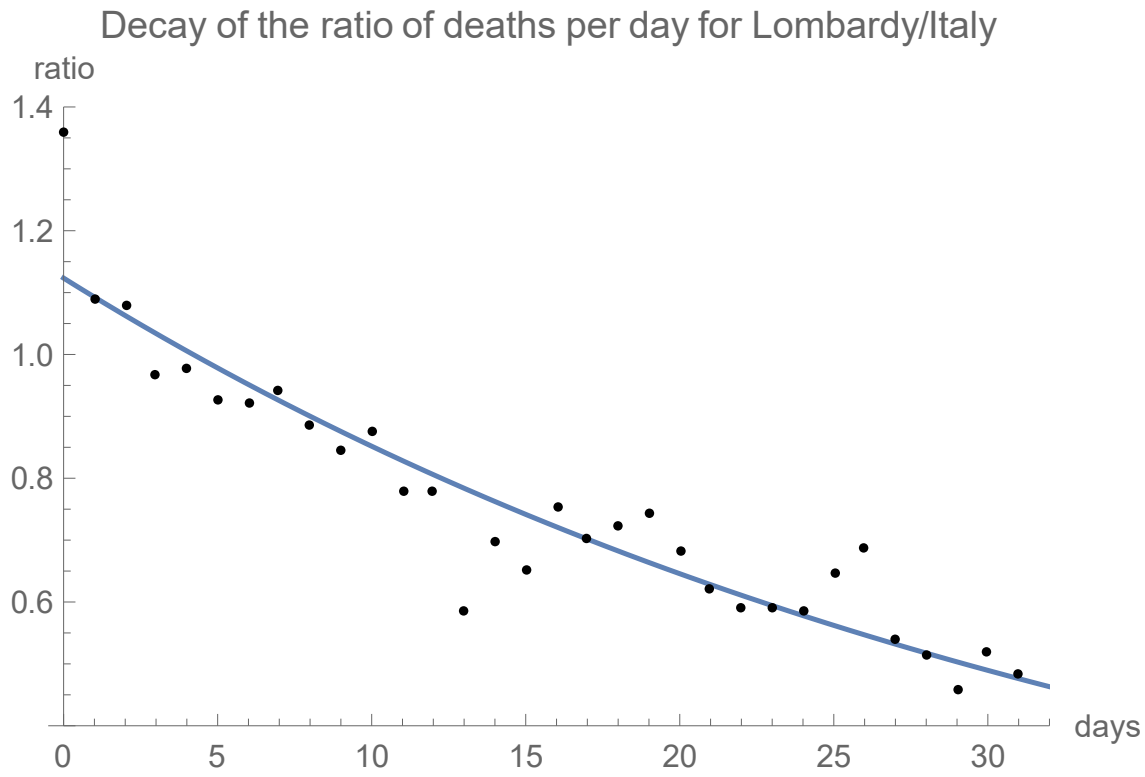
https://mrc-ide.github.io/covid19-short-term-forecasts/index.html.

Lombardy and Italy

First major outbreak in Europe and are where full social distancing was first applied and a peak was first reached on 27 March.



The 3-day average from 28-30 March, the ratio R_{12}, (1 denotes Lombardy and 2 the rest of Italy), was $R_{12} \sim 1.3$. For dates near 28-30 April it is $R_{12} \sim 0.5$ and falling (data taken from https://github.com/pcm-dpc/COVID-19/tree/master/schede-riepilogative/regioni).



If the decay in both Lombardy and the rest of Italy is due only to lockdown reducing the effective R this is difficult to understand.

Taking 32 values of R_{12} and fitting a form $a\exp(-\lambda\tau)$ find $\lambda = 0.028$.

Fixing $R = 1$ and $1/\sigma = 7$ days $\rightarrow z_1 - z_2 = 0.196$

The largest uncertainty is from varying the first day of data included $\rightarrow$ uncertainty of about 15%.

Assuming ifr is common throughout Italy $ifr \equiv ifr_I$, and using the accumulated deaths in the middle of the period

$$z_1 \approx \frac{11100}{(ifr_I 10^7)},\ z_2 \approx \frac{9900}{(ifr5 \times 10^7)}\,\frac{z_1 - z_2}{ifr_I} = 0.196 \rightarrow ifr_I = 0.0046 \pm 0.0006.$$

Hence, inferred ifr is rather lower than the common assumption, by a factor of about 2.5.

z_1 for Lombardy approximately 20 days before last data used, i.e. on 10th April, was $(13,800/0.0046) \times 10^7 = 0.30 \pm 0.05$. For the remainder of Italy $z_2 = 0.06 \pm 0.01$.

Now confirm that $R \approx 1$ is a good assumption.

Using $R = 1$ the total decay for the remainder of Italy is a fall of 0.83 over 32 days, entirely due to $(1 - z)$.

In fact $\sim 0.6 \rightarrow R \approx 0.95$.

This R gives a slightly larger $z_1 - z_2$, by a factor of $1/0.95$.

The absolute fall in Lombardy largely due to $(1 - z)$ but partially R.

Using $R = 0.95$ we obtain

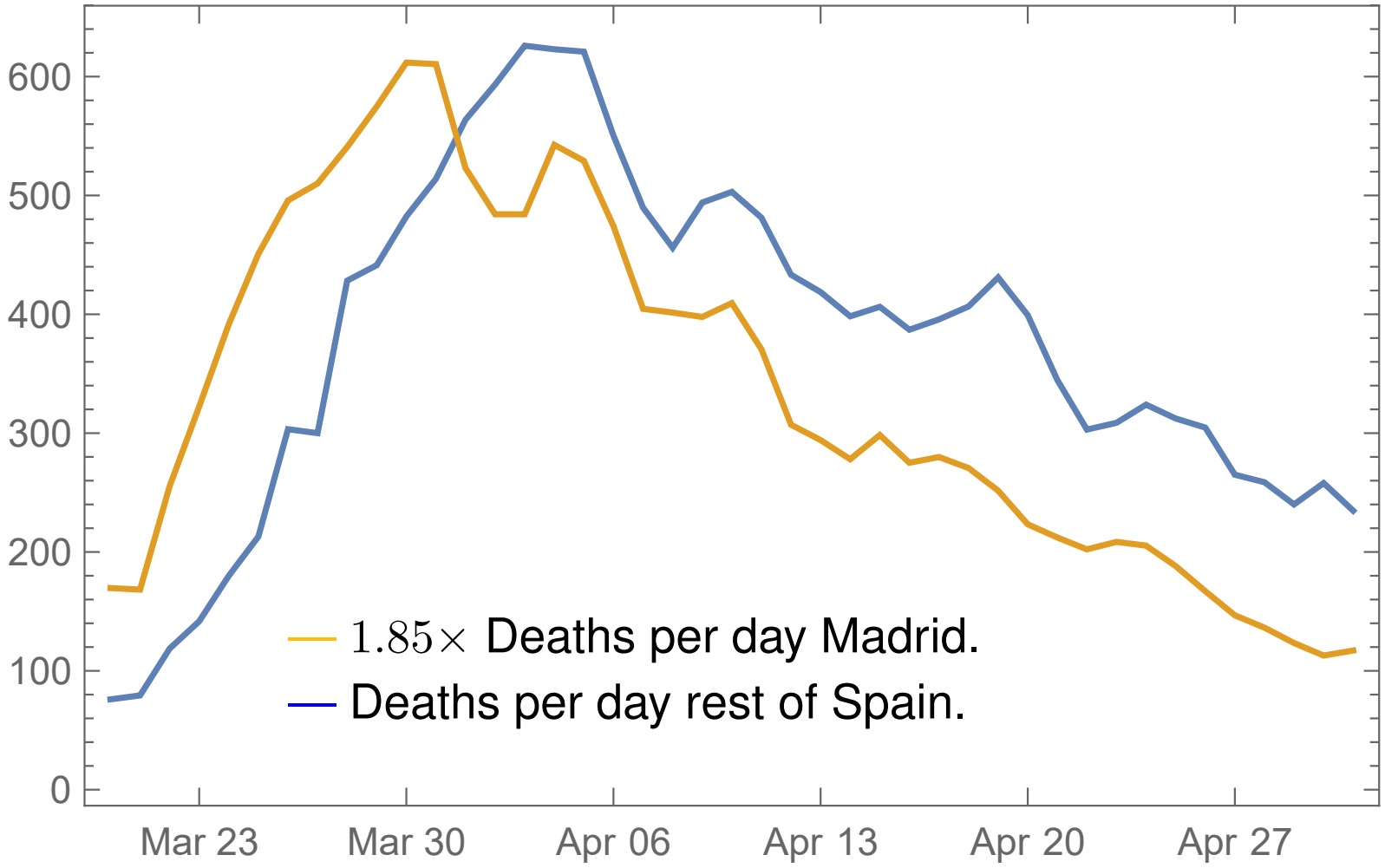
$$ifr_I = 0.0044 \pm 0.0006,$$

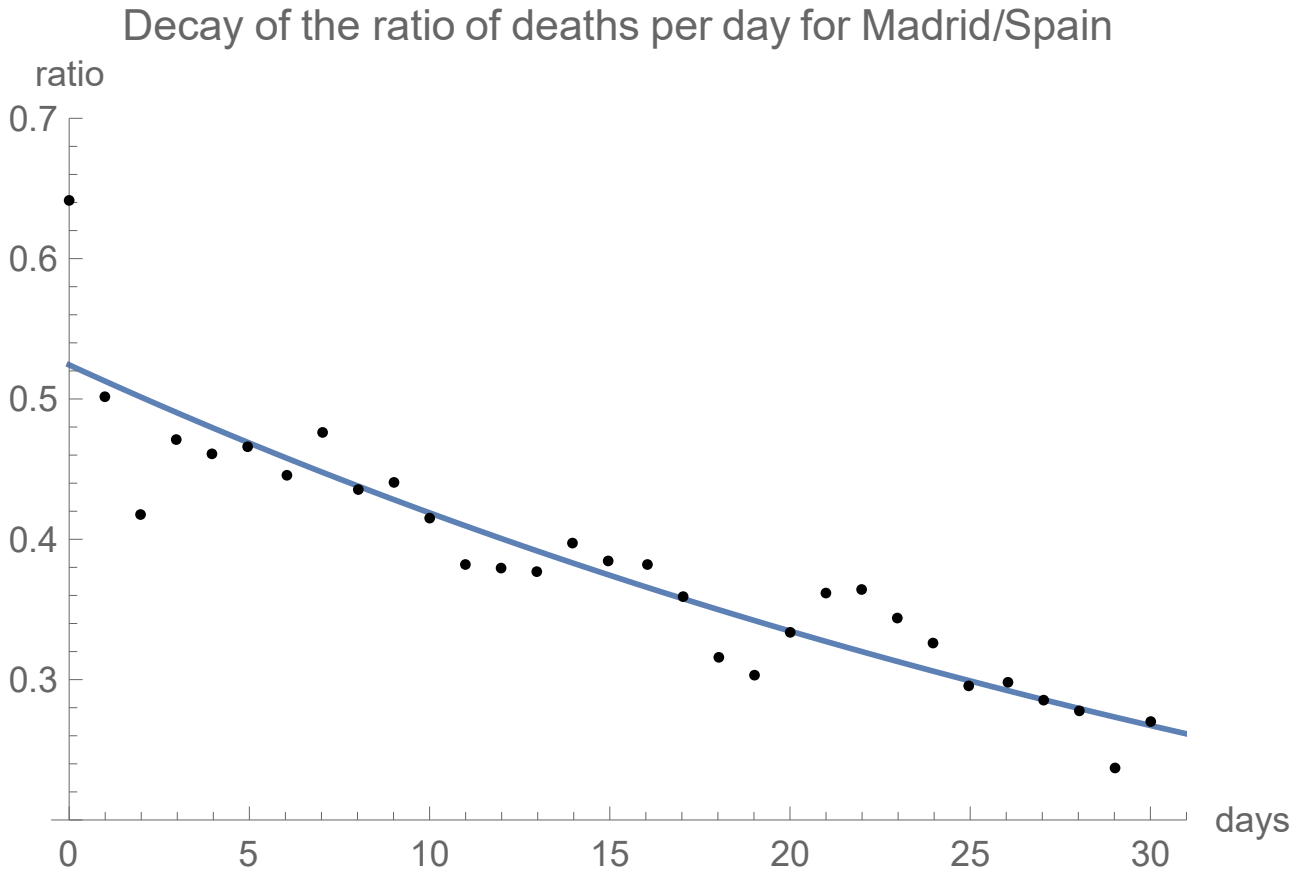
For Lombardy $z_1 = 0.32 \pm 0.05$ on 10th April.

For the rest of Italy $z_2 = 0.065 \pm 0.01$.

Madrid and Spain

The peak was reached only a couple of days later than in Italy.





R_{12} for the 3-day average for 29-31 March was $R_{12} \sim 0.6$, but near 28-30 April it is $R_{12} \sim 0.25$.

Fit the last 31 values (data taken from `https://covid19.isciii.es/`.)

$\rightarrow \lambda = 0.0225$ and $z_1 - z_2 = 0.158 \pm 15\%$.

Using the values for acummulated deaths and populations in the middle of the period

$$z_1 \approx \frac{6550}{(ifr_S 6.7 \times 10^6)}, \quad z_2 \approx \frac{11500}{(ifr_S 4 \times 10^7)} \rightarrow \frac{z_1 - z_2}{ifr_S} = 0.158$$
$$\rightarrow ifr_S = 0.0043 \pm 0.0006.$$

Infer that for Madrid $z_1 = 0.28 \pm 0.05$ on 10th April and for the rest of Spain $z_2 = 0.09 \pm 0.02$.

$R \approx 1$ is a good assumption, but gives a fall for the remainder of about 0.7 where it is more like 0.5.

A value of $R = 0.9$ works a little better, and means ifr and z_1, z_2 are raised by 10%

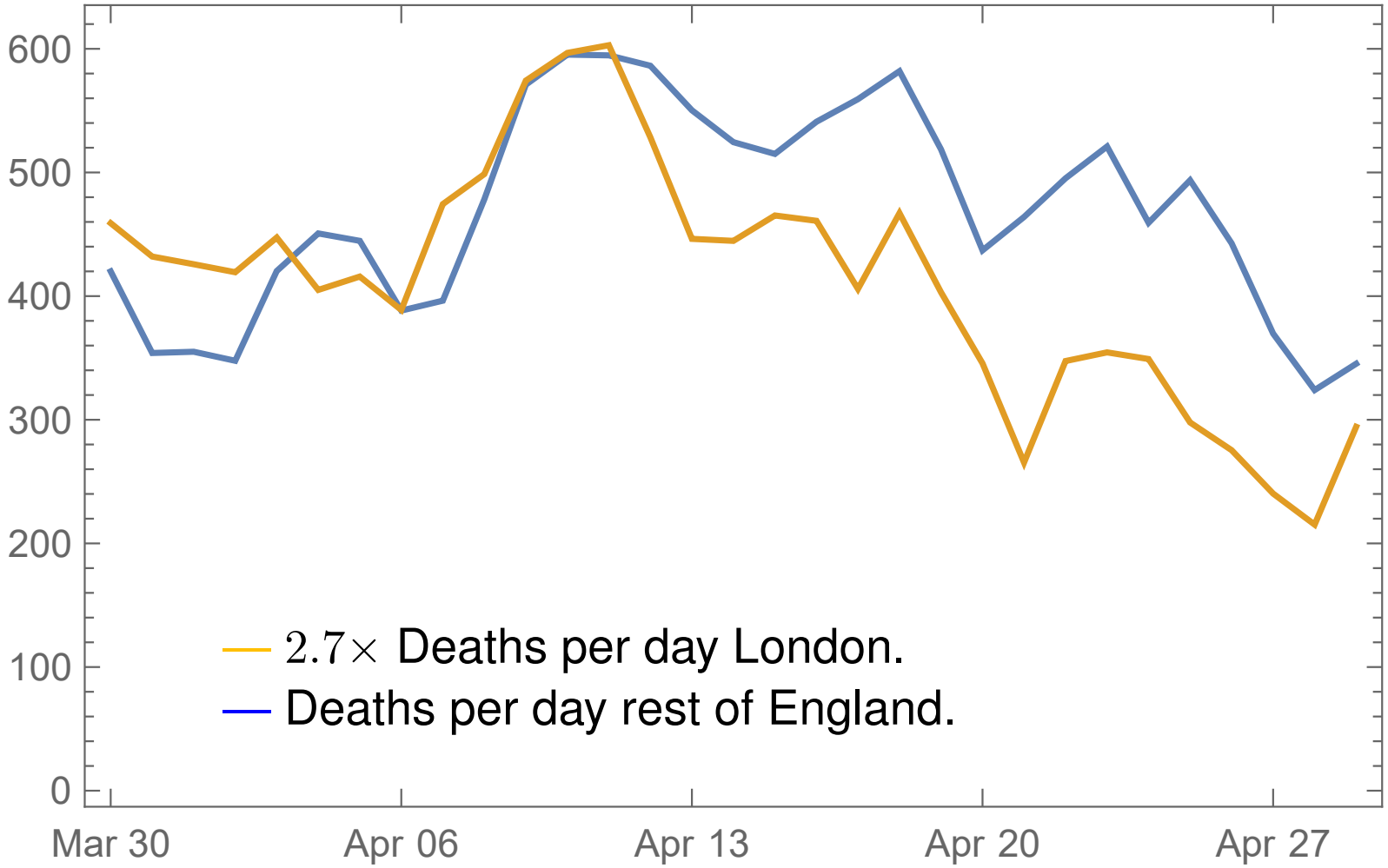
$$ifr_S = 0.0039 \pm 0.0005.$$

For Madrid $z_1 = 0.31 \pm 0.05$ on 10th April.

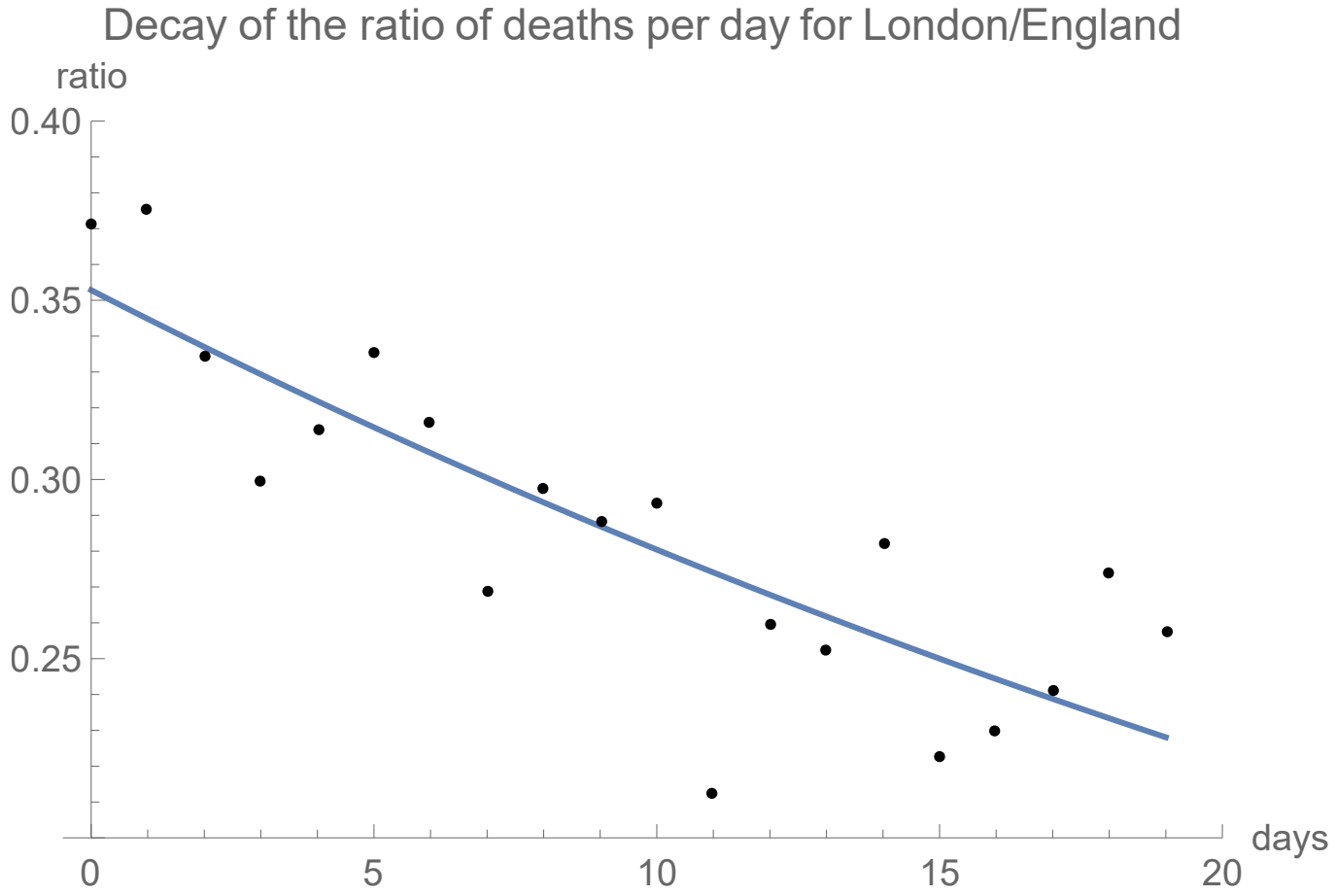
For the rest of Spain $z_2 = 0.10 \pm 0.02$.

London and England

Behind Italy and Spain, but England now has slow rate of decline. London rather distinct from the rest of England.



Over 21 days from 9th April the ratio of deaths reported per day (data from https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-daily-deaths/) fallen from $R_{12} \sim 0.37$ for 7-9 April to $R_{12} \sim 0.23$ up to 28-30 April.



Fit to $a\exp(-\lambda t)$, finding $\lambda = 0.023$ and hence, $z_1 - z_2 = 0.161 \pm 20\%$.

Making a common assumption on ifr_E

$$z_1 \approx 3650/(ifr_E \times 8.9 \times 10^6), \quad z_2 \approx 9700/(ifr_E \times 48 \times 10^7).$$

This results in $ifr_E \approx 0.0013$ much less than Italy and Spain. However, the comparison is not so straightforward.

Most deaths from Covid-19 are amongst the population older than 65.

In Lombardy the fraction of the population over 65 is 22%, similar to Italy as a whole, and in Madrid 20%, marginally higher than the national percentage.

In London it is just 12%, as opposed to 18% for England $\rightarrow$ ifr for London lower than the rest of England (and Italy and Spain).

Ratio of the fraction of the population in London over 65 years old to that for the reminder of England is $12\%/19\% = 1/1.6 \rightarrow ifr_{E-L} = 1.6 ifr_L$.

Using this

$$\frac{z_1 - z_2}{1.6 ifr_L} = 0.161 \rightarrow ifr_L = 0.0018 \pm 0.0003 \rightarrow ifr_{E-L} = 0.0028 \pm 0.0004.$$

On 10th April $z_1 = 0.31 \pm 0.07$

For the rest of England $z_2 = 0.11 \pm 0.02$.

The absolute fall for the remainder of England is 0.65 over the past 20 days, reasonably consistent with $R = 1$ (within about 5%)

Times before the peaks

Include data for times before the peak (defined as the day on which I begin the fit to the ratio). This includes dates only a week or so after the lockdown in each country, when the number deaths was far smaller.

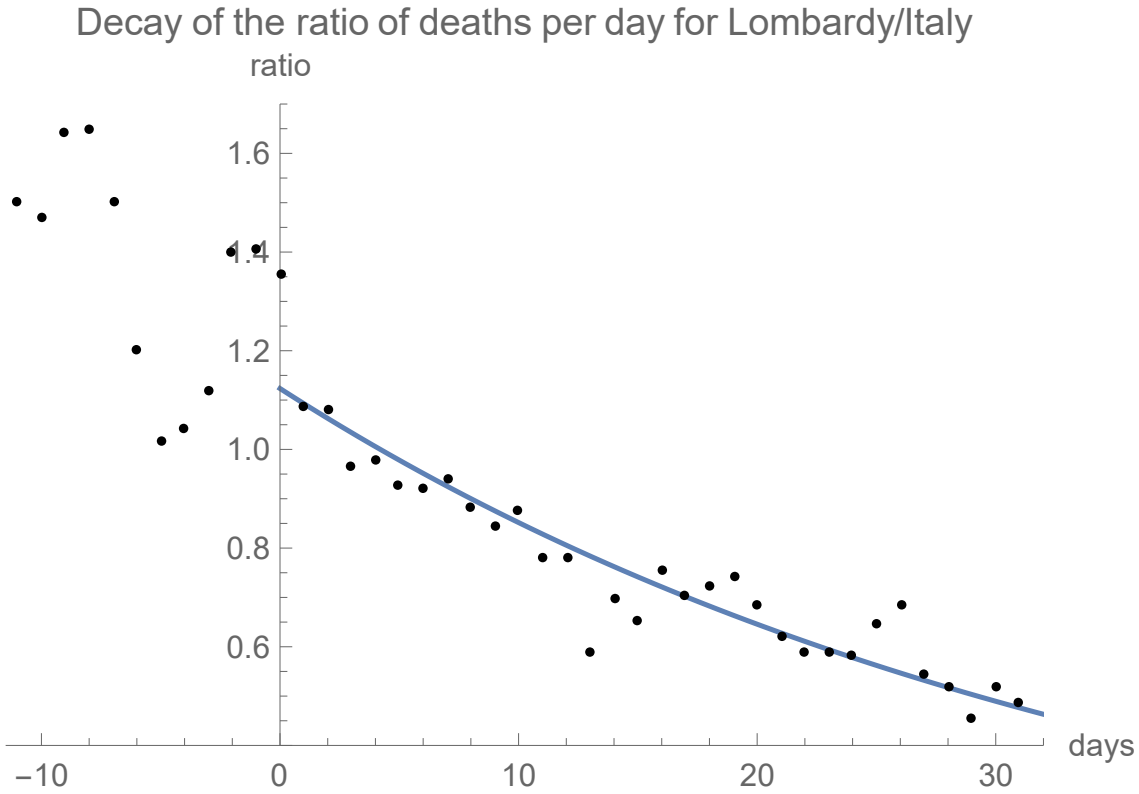
In each case in the week or more before the peak the ratio falls.

Hence, even though before full social distancing one might suspect more densely populated regions with the highest rates of infection to have the largest values of R, the opposite is true.

The region with the highest number of deaths per population assumes this role very early, but even before lockdown is applied the remainders then tend to catch up.

The most affected regions have a smaller effective R even at earlier times when the absolute value of $R >> 1$.

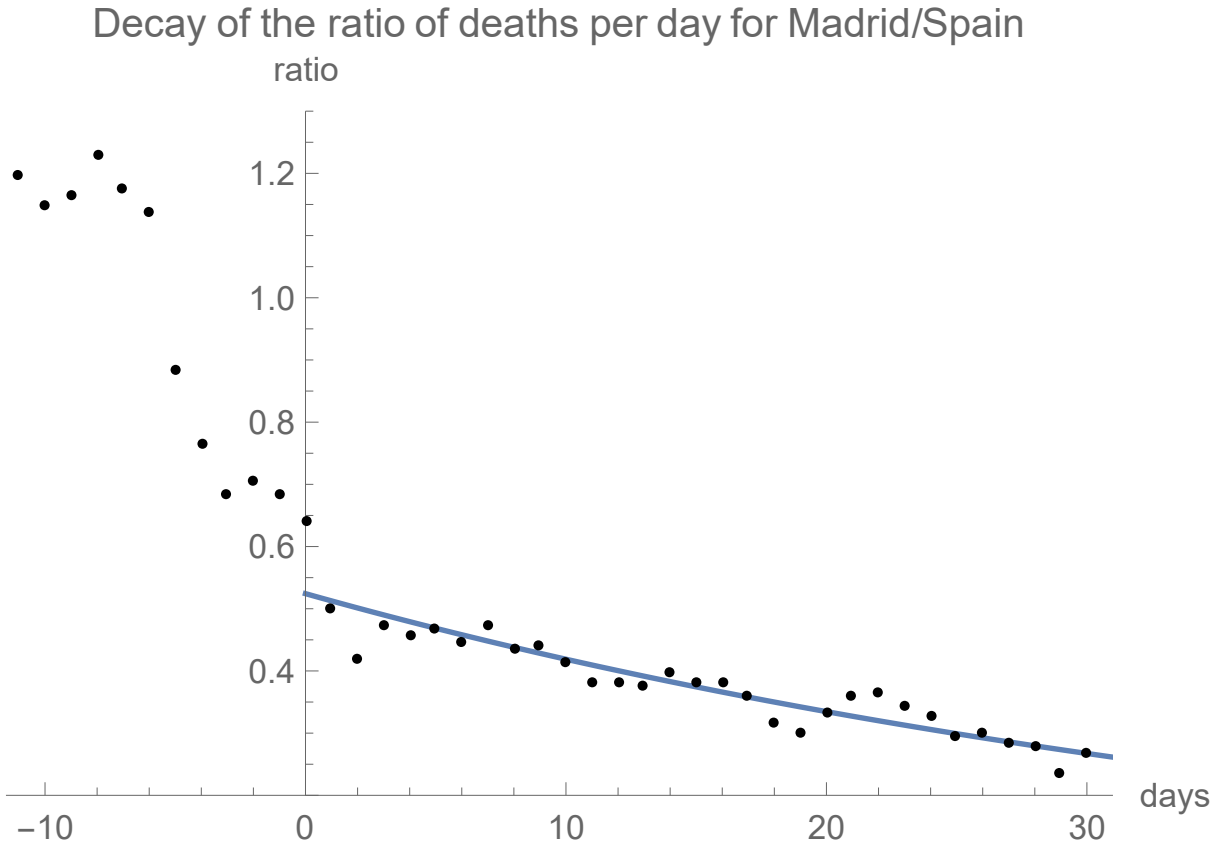
Either R is smaller for Lombardy, Madrid and London than the remainders for all but the very earliest times, or even before the peak the $(1 - z)$ factor is playing a significant role.



Assuming that before the peaks, R is the same in all regions to a good approximation, then the ratio $R_{12} \propto \exp(-R(z_1 - z_2)\sigma\tau)$.

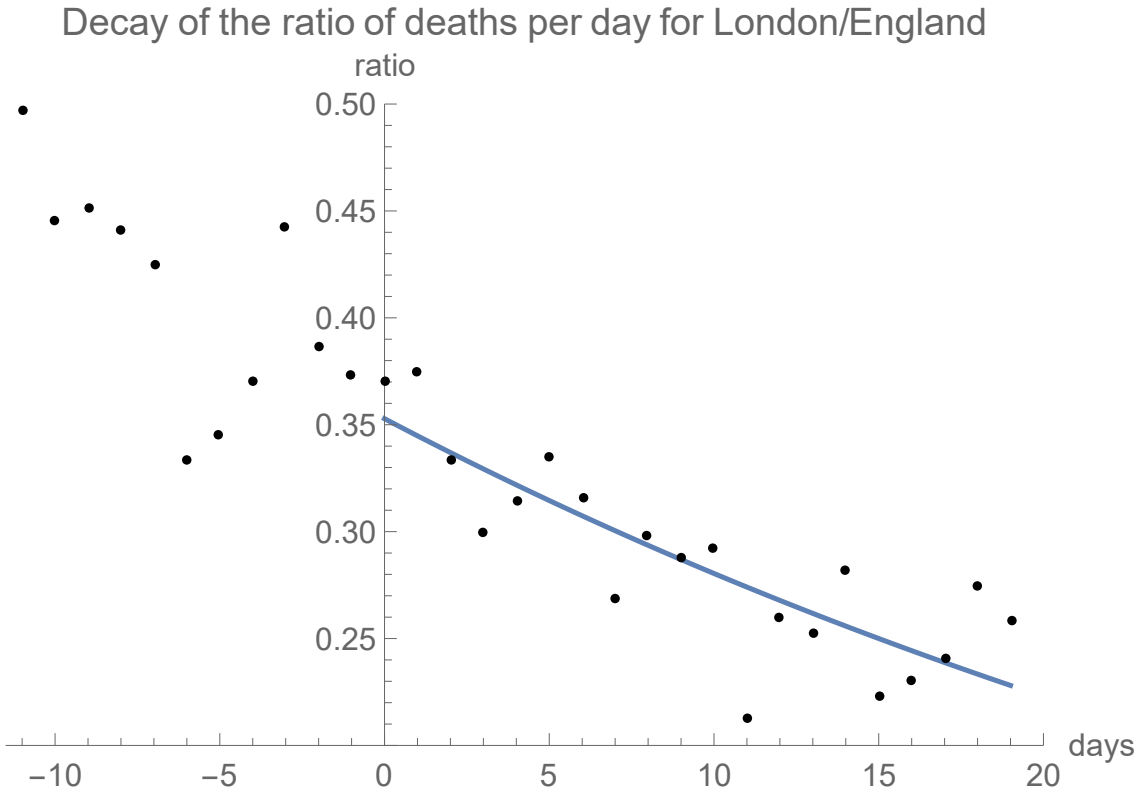
z_1 and particularly z_2 very small, e.g. using the value of ifr_I then 5 days before the peak z_1 for Lombardy will be $\approx 4200/((0.0044 \times 10^7) \approx 0.1$.

$R \sim 2 \rightarrow R_{12} \approx \exp(-0.03\tau)$, leading to a fall of about 0.7 in the 10 days before the peak. This is roughly what is observed for Lombardy/Italy.



For Madrid/Spain the relative fall is ~ 0.5, implying a higher value of R, consistent with the larger absolute rate of increase before the peak for Spain.

It is clear that there is a distinct kink in the ratio near the peak, consistent with a sudden change in R.



For London, there is a clear decrease even before the peak.

Qualitatively similar to Lombardy/Italy, a fall of about 0.7 in the 10 days before the peak.

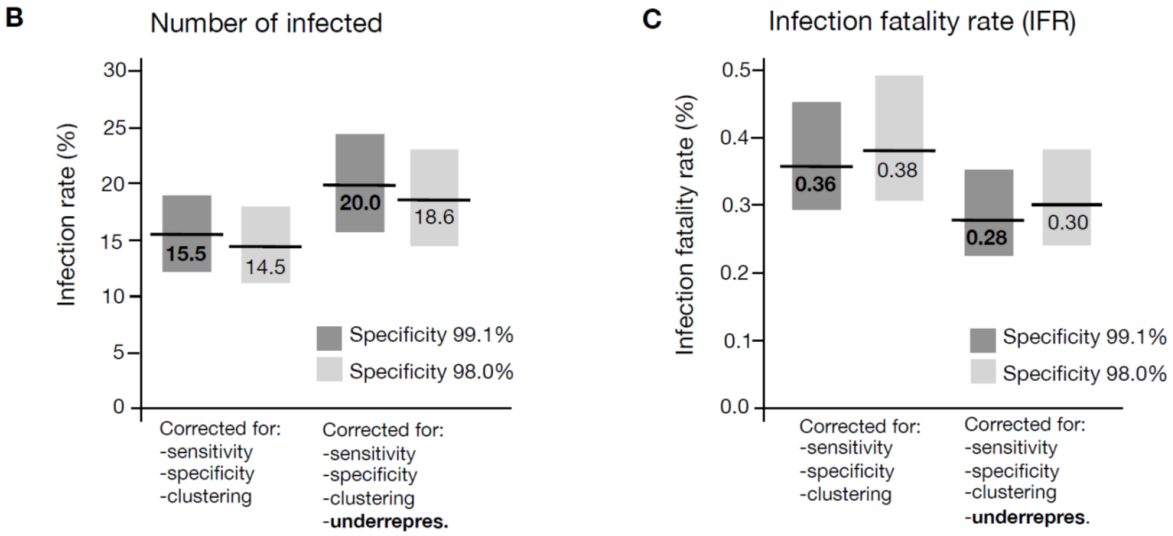
This is consistent with the fact that the absolute rate of increase in the approach to the peak is similar to Italy.

Evidence from seroprevalence tests

Early small-scale results seemed to fit in well with this picture.

Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event

Hendrik Streeck[1], Bianca Schulte[1], Beate M. Kümmerer[1], Enrico Richter[1], Tobias Höller[5], Christine Fuhrmann[5], Eva Bartok[4], Ramona Dolscheid[4], Moritz Berger[3], Lukas Wessendorf[1], Monika Eschbach-Bludau[1], Angelika Kellings[5], Astrid Schwaiger[6], Martin Coenen[5], Per Hoffmann[7], Birgit Stoffel-Wagner[4], Markus M. Nöthen[7], Anna-Maria Eis-Hübinger[1], Martin Exner[2], Ricarda Maria Schmithausen[2], Matthias Schmid[3] and Gunther Hartmann[4]



COVID-19 Antibody Seroprevalence in Santa Clara County, California

Eran Bendavid[1], Bianca Mulaney[2], Neeraj Sood[3], Soleil Shah[2], Emilia Ling[2], Rebecca Bromley-Dulfano[2], Cara Lai[2], Zoe Weissberg[2], Rodrigo Saavedra-Walker[4], Jim Tedrow[5], Dona Tversky[6], Andrew Bogan[7], Thomas Kupiec[8], Daniel Eichner[9], Ribhav Gupta[10], John P.A. Ioannidis[1,10], Jay Bhattacharya[1]

We can use our prevalence estimates to approximate the infection fatality rate from COVID-19 in Santa Clara County. Through April 22, 2020, 94 people died from COVID-19 in the County. If our estimates of 54,000 infections represent the cumulative total on April 1, and we assume a 3 week lag from time of infection to death, up to April 22[24], then 94 deaths out of 54,000 infections correspond to an infection fatality rate of 0.17% in Santa Clara County. If antibodies take longer than 3 days to appear, or if the average duration from case identification to death is less than 3 weeks, then the prevalence rate at the time of the survey was higher and the infection fatality rate would be lower. On the other hand, if deaths from COVID-19 are under reported or the health system is overwhelmed than the fatality rate estimates would increase. Our prevalence and fatality rate estimates can be used to update existing models, given the large upwards revision of under-ascertainment.

CORONAVIRUS | APR. 23, 2020

Antibody Test Estimates 1.7 Million NYC Residents Have Had Coronavirus

By Adam K. Raymond

THL: Coronavirus may have infected dozens of times more than confirmed in Finland

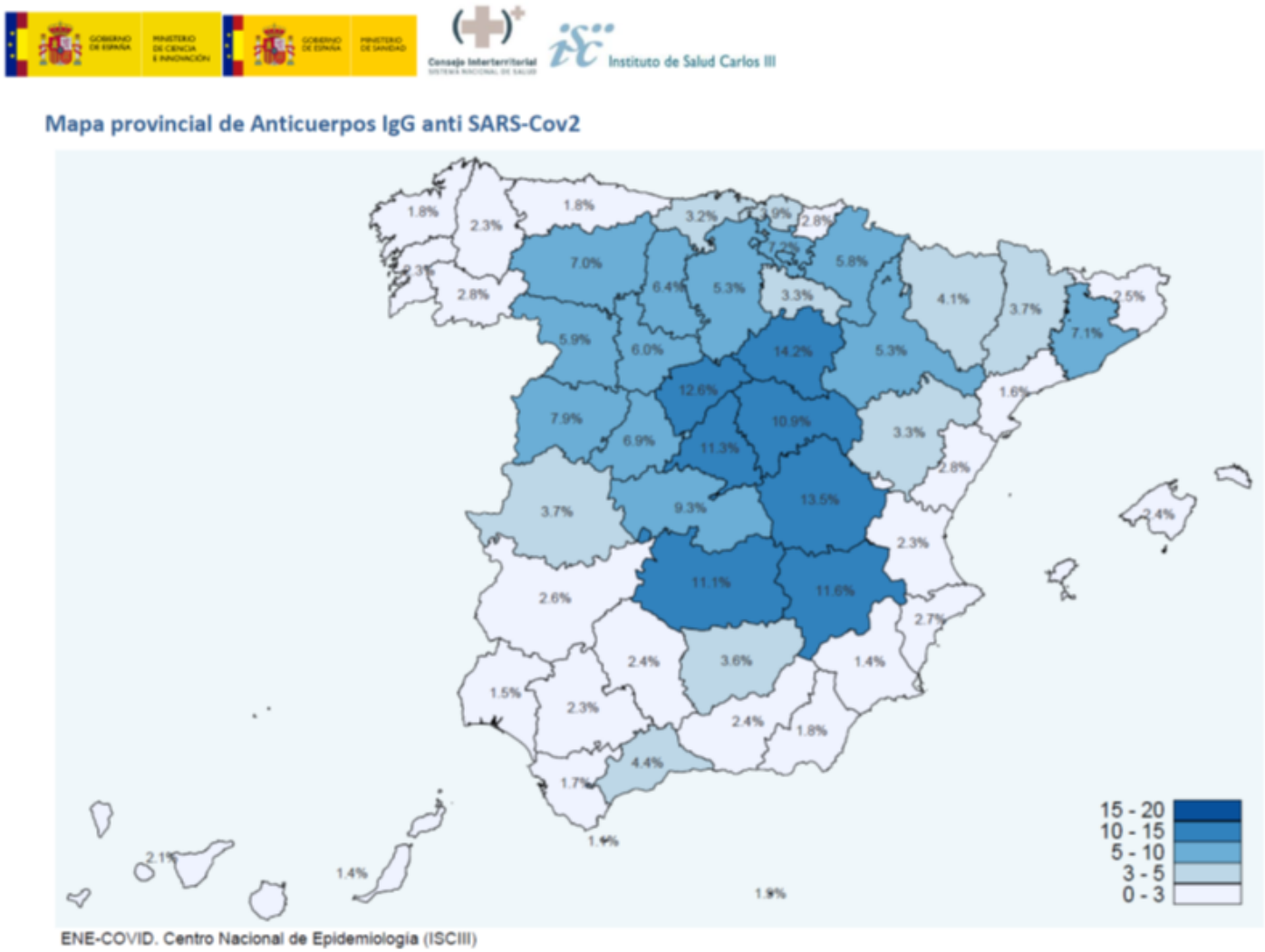
FINLAND / 17 APRIL 2020

THE ACTUAL NUMBER of people infected with the new coronavirus may be dozens of times higher than the number of laboratory-confirmed infections, reports the Finnish Institute for Health and Welfare (THL).

Its estimate is based on the findings of a newly completed antibody study.

Plus increased rates in reporting influenza like symptons also indicative.

Later studies in some of these pulled back. Also large scale study in Spain reported national average of about 5%.



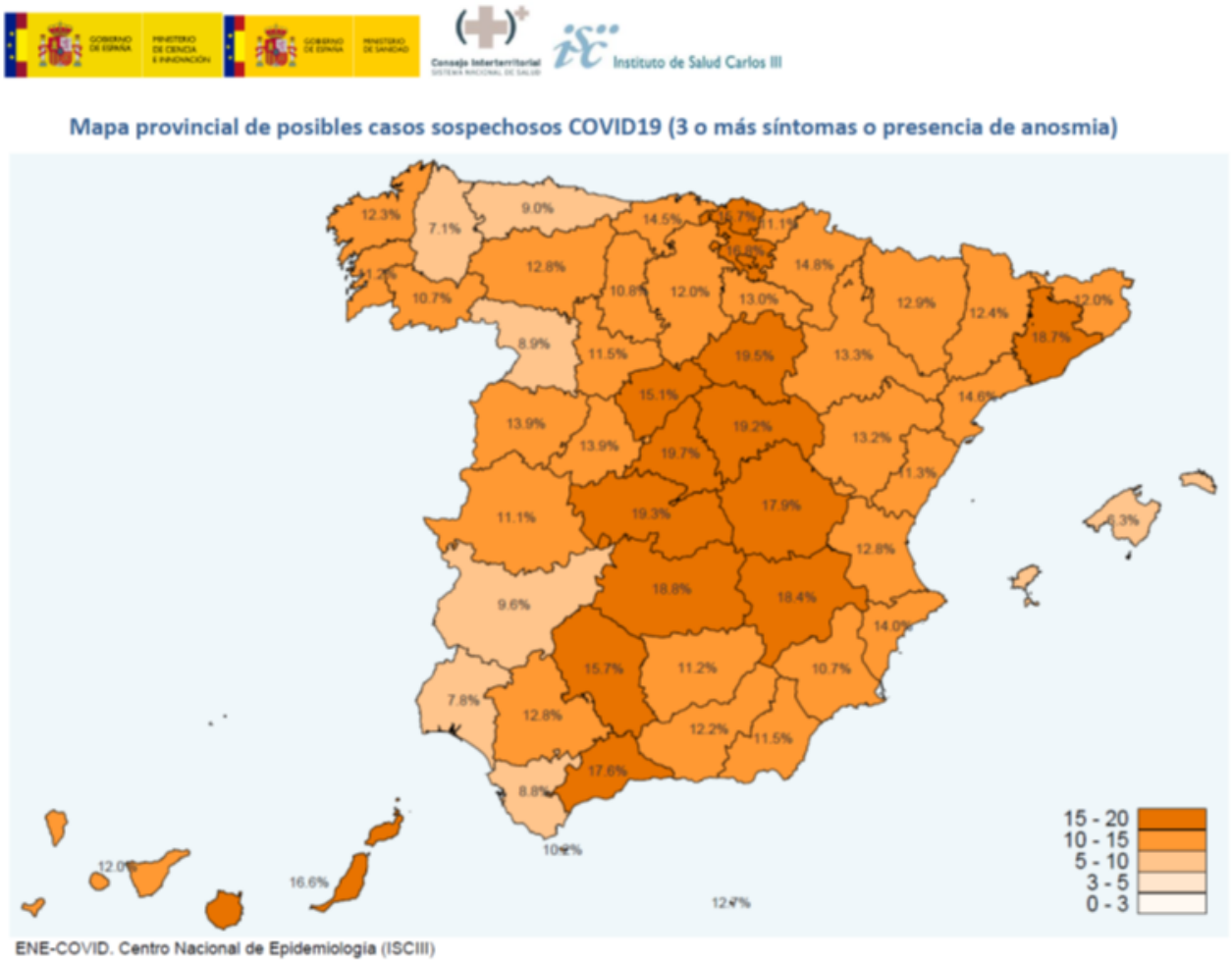
https://www.mscbs.gob.es/gabinetePrensa/notaPrensa/pdf/13.

Prevalencia de anticuerpos IgG anti SARS-Cov2 según antecedentes relacionados con COVID19

	Totales			Hombres			Mujeres		
	Nº	%	IC 95%	Nº	%	IC 95%	Nº	%	IC 95%
Antecedentes de PCR positiva									
No	60640	**4,7**	4,4 - 5,0	29150	**4,7**	4,3 - 5,1	31490	**4,8**	4,4 - 5,1
Sí	247	**83,0**	76,2 - 88,2	102	**87,6**	77,8 - 93,4	145	**79,4**	69,8 - 86,6
Síntomas relacionados con COVID19*									
Asintomáticos	40202	**2,5**	2,3 - 2,8	20366	**2,5**	2,2 - 2,8	19836	**2,5**	2,2 - 2,8
Paucisintomáticos (1-2 síntomas**)	12362	**4,6**	4,1 - 5,1	5619	**4,8**	4,1 - 5,5	6743	**4,4**	3,7 - 5,2
3-5 síntomas**	5431	**8,2**	7,1 - 9,4	2186	**9,9**	8,3 - 11,8	3245	**6,9**	5,8 - 8,2
>5 síntomas**	1035	**14,7**	11,9 - 18,0	366	**21,6**	16,6 - 27,7	669	**10,7**	7,9 - 14,3
Anosmia	1867	**43,3**	39,9 - 46,8	718	**44,9**	40,2 - 49,8	1149	**42,2**	38,1 - 46,4
3 o más síntomas* en las últimas 2 semanas									
No	58504	**4,7**	4,4 - 5,0	28434	**4,7**	4,3 - 5,1	30070	**4,6**	4,3 - 5,0
Sí	2393	**14,1**	12,0 16,6	821	**14,0**	11,1 17,7	1572	**14,2**	11,7 17,1

Not entirely reliable. Misses up to 17% of definite positive cases (delay in antibodies being formed?).

Picks up less than half with symptoms of loss of taste/smell.



Additional “possible” missed cases with loss of sense of taste/smell or more than 3 standard symptoms.

Yesterday, suggestions London $\sim 17\%$, rest of England $\sim 5\%$.

Evidence now pointing against low ifr and large fraction of population already infected?

Well certainly some strong evidence partially against the former. Not so clear for the latter.

Additional possible feature to consider – variable susceptibility.

Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold

Authors: M. Gabriela M. Gomes[1,2,3]*, Rodrigo M. Corder[4], Jessica G. King[5], Kate E. Langwig[6], Caetano Souto-Maior[7], Jorge Carneiro[8], Guilherme Gonçalves[9], Carlos Penha-Gonçalves[8], Marcelo U. Ferreira[4], Ricardo Aguas[10].

Abstract: As severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spreads, the susceptible subpopulation is depleted causing the incidence of new cases to decline. Variation in individual susceptibility or exposure to infection exacerbates this effect. Individuals that are more susceptible or more exposed tend to be infected earlier, depleting the susceptible subpopulation of those who are at higher risk of infection. This selective depletion of susceptibles intensifies the deceleration in incidence. Eventually, susceptible numbers become low enough to prevent epidemic growth or, in other words, the herd immunity threshold (HIT) is reached. Although estimates vary, simple calculations suggest that herd immunity to SARS-CoV-2 requires 60-70% of the population to be immune. By fitting epidemiological models that allow for heterogeneity to SARS-CoV-2 outbreaks across the globe, we show that variation in susceptibility or exposure to infection reduces these estimates. Accurate measurements of heterogeneity are therefore of paramount importance in controlling the COVID-19 pandemic.

doi.org/10.1101/2020.04.27.20081893

Variable Susceptibility (Not in arXiv:2005.00495)

The discussion so far assumes that all members of a population are equally susceptible to infection.

Relax this assumption. Instead of the fraction who have been infected, z, use the fraction still susceptible to infection, S.

Assume that S is a function of the susceptibilty to infection x, i.e $S(x)$.

The probability distribution function (pdf) of the population before infection starts is $q(x)$.

The total fraction susceptible is 1, so by definition $\int_0^\infty q(x)\,dx = 1$.

We also define

$$\bar{x}(t=0) = \int xq(x)\,dx = 1.$$

However, $q(x)$ has some non-zero variance V.

At some later time we define

$$\int S(x)\,dx \equiv S, \qquad \int y(x)\,dx \equiv y.$$

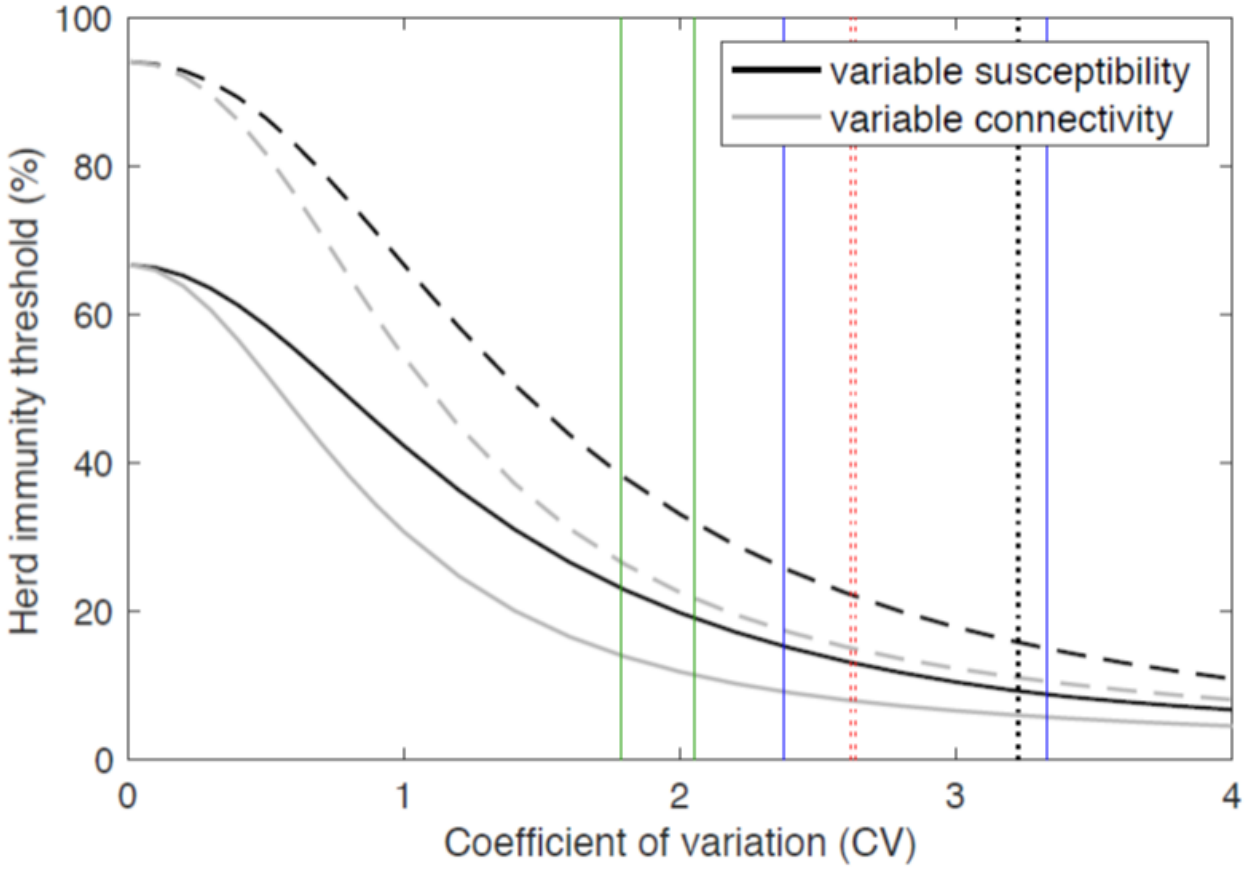
Using $S(x)$ rather than z we can rewrite the evolution equations as

$$\frac{dy(x)}{dt} = \beta y x S(x) - \sigma y(x), \qquad \frac{dS(x)}{dt} = -\beta y x S(x).$$

The increase in the newly infectious fraction and decrease in the susceptible fraction are proportional to the susceptible fraction weighted by the susceptibility.

Assume this is uncorrelated to the force of transmission from the currently infected fraction y – the virus can be spread equally efficiently by all infected people.

$$\frac{dy}{dt} = \big(R \int x S(x)\,dx - 1\big)\sigma y \equiv (R\bar{x} - 1)\sigma y.$$



Gomes *et al* study studies effect on "herd immunity" threshold by solving (slightly different) equations numerically.

Some estimates, e.g. malaria, tubercolosis, SARS-Cov-1.

Does not so far consider possible data constraint on variance V.

Compared to our previous equations we have the replacement

$(1 - z) \rightarrow \int xS(x)\,dx$ – we have an *effective* fraction infected $z_{eff} = 1 - \int xS(x)\,dx$.

The rate of infection starts to decrease when $R \int xS(x)\,dx < 1$.

From the equation for the rate of change of $S(x)$, at any time the decay in $S(x)$ is proportional to $xS(x)$ (assuming the variation of y due to time varying $\bar{x}$ is relatively slow)

The fraction remaining with susceptibility x can be written as $\exp(-\delta(t)x)$, for some $\delta(t)$.

Hence we can always write $S(x)$ at some time during virus spread as

$$S(x,t) = q(x)\exp(-\delta(t)x).$$

Any variation in susceptibilty leads to $S(x)$ decreasing more quickly for larger x so $\int xS(x)\,dx$ decreases more quickly than $\int S(x)\,dx$.

Force of infection $R \int xS(x)dx$ falls faster than $R(1 - z)y \equiv RSy$.

Example - let intial $q(x) = S(x, t = 0)$ be a gamma function pdf. This is defined in terms α and β such that

$$q(x) = N(\alpha, \beta) x^{\alpha - 1} \exp(-\beta x).$$

$N(\alpha, \beta)$ is the normalization and if $\int q(x)\, dx = 1$ then

$$N(\alpha, \beta) = \beta^{\alpha} / \Gamma(\alpha).$$

We also have specific expressions for the mean and variance,

$$\bar{x} = \alpha / \beta, \qquad V = \alpha / \beta^2.$$

Since we define $\bar{x} = 1$, $\alpha = \beta \rightarrow V = 1/\alpha$ and

$$q(x) = \frac{\alpha^{\alpha}}{\Gamma(\alpha)} x^{\alpha - 1} \exp(-\alpha x).$$

Consider $S(x,t)$ at some time during the spread of the virus.

$$S(x,t) = q(x)\exp(-\delta(t)x) \equiv \frac{\alpha^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp(-(\alpha+\delta)x).$$

$S(x,t)$ is now a different gamma function pdf with $\beta = \alpha + \delta$ rather than $\beta = \alpha$.

The normalization should now be $\frac{(\alpha+\delta)^\alpha}{\Gamma(\alpha)}$. This reindentification means that we can calculate $S = \int S(x,t)\,dx$ and $\bar{x} = \int xS(x,t)\,dx$ very easily.

$$\begin{aligned} S &= \frac{(\alpha)^\alpha}{(\alpha+\delta)^\alpha} = (1+\delta/\alpha)^{-\alpha} \\ \bar{x} &= \alpha/(\alpha+\delta)\int S(x,t)\,dx = (1+\delta/\alpha)^{-1}(1+\delta/\alpha)^{-\alpha}. \end{aligned}$$

The first term is the change in the mean value of x if the normalization were correct, while the second is due to the change in the fraction of the population still susceptible.

The fall in the fraction susceptible for susceptibility of 1 is $\exp(-\delta)$.

Let us consider δ much less than 1 and α.

$$(1+\delta/\alpha)^{-1} = 1 - \frac{\delta}{\alpha} + \mathcal{O}(\delta^2), \qquad (1+\delta/\alpha)^{-\alpha} = 1 - \delta + +\mathcal{O}(\delta^2).$$

So to first order in δ

$$\int xS(x,t)\,dx = 1 - \delta(1+1/\alpha) \equiv 1 - \delta(1+V) \quad \rightarrow z_{eff} = \delta(1+V).$$

The effective fraction of the population infected is changed from $z = \delta$, due just to the decrease in $\int S(x,t)\,dx$, to $z_{eff} = \delta(1+V)$, due also to the change in the average susceptibility.

The fraction of the population that needs to become infected in order to obtain so-called "herd immunity" is decreased by a factor of $1+V$ (up to corrections of order δ^2)

If we need $z_{eff} = 0.6$ and the variance of the susceptibility is $V = 2$ then we need only that $z = \delta = 0.2$.

Variation in susceptibility offers an alternative explanation to why the rates of infection and deaths are falling more quickly in Lombardy, Madrid and London in comparison to the rest of Italy, Spain and England respectively.

It is the effective fraction that is obtained from my analysis. The larger than expected fraction could be due to a genuinely larger infected fraction than assumed, i.e. lower ifr, or due to a significant variance in the susceptibility, or some element of both.

z_{eff} about 3 times larger than expected could be due to $V \approx 2$.

In practical terms, in either case the evidence from the relative fall in the most affected regions suggests that these regions already have a large "effective" infection rate which suppresses the reproduction rate R.

The remainder of each country will also already have reduced effective R.

Whether this is due to a lower ifr than normally assumed or a variable susceptibilty with variance ~ 2, the main result is the same either way.

Discussion and further "evidence"

The number of deaths reported is subject to upwards corrections due to some potential omissions (I have used hospital deaths in England). Corrections could raise the ifr by a factor of maybe 50%.

The extracted values of $z_{1,eff}, z_{2,eff}$ are insensitive to this potential shortfall (assuming it is consistent in a given country).

The observed difference in the decay rates between regions could be due to those regions with highest initial number of cases observing the rules of social distancing better (see arXiv:2004.07827 for the case of municipalities within Lombardy).

In each case, Lombardy, Madrid and London need to have a value of R about 0.2 lower than the rest of the country, despite being more densely populated – R might instead be greater due to greater proximity of population, more use of public transport etc.

If R is greater in Lombardy, Madrid and London the behaviour of the ratios requires raising the values of $z_{1,eff}$ and $z_{2,eff}$ further.

The model I apply is extremely simple and there are numerous sources of uncertainty, which are difficult to quantify.

Since the values of $1/z_{eff}$ and ifr are linearly proportional to σ, a change in this (which could be $\sim 20\%$) translates directly into their values.

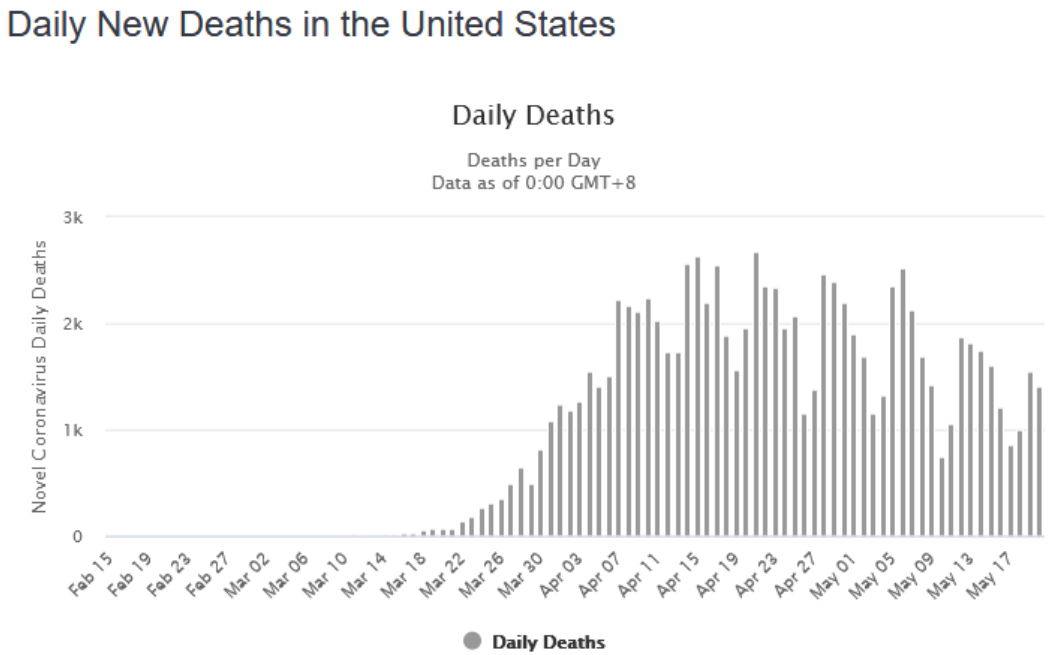
The assumption of constant z_{eff} fractions is also a simplification, though the difference $z_{1,eff} - z_{2,eff}$ is less sensitive.

Asumptions about a constant time from infection to death can be improved upon as can time variation of R which may effectively be larger at the beginning (opposite to $(z_{1,eff} - z_{2,eff})$ dependence).

Checked by varying the start date by 2 days either way, and does indeed lead to the largest uncertainty.

The type of ifr and/or variable susceptibility inferred here would suggest that New York (in particular), New Jersey and Belgium should now be experiencing significant effects from an extra $(1 - z_{eff})$ factor multiplying their R value. Deaths per population very high $\mathcal{O}(1/1000)$.

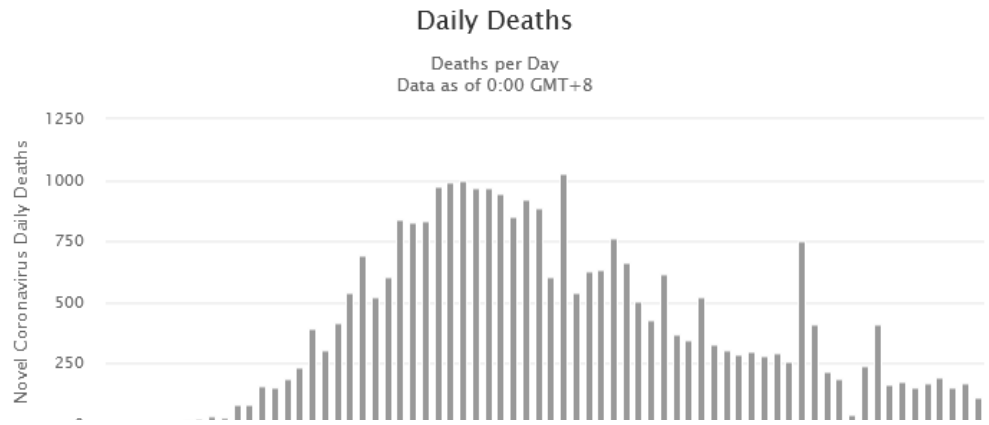
It seems likely that this is indeed the case. New York and New Jersey have a consistently declining number of deaths, while in general the remainder of the USA displays a relatively flat rate. (Plots from https://www.worldometers.info/coronavirus.)



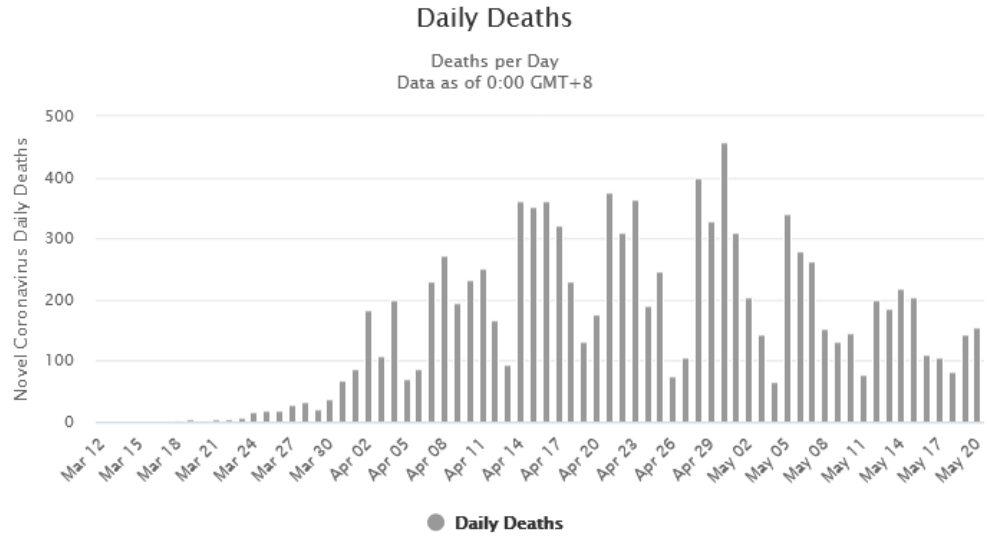
Peak reached in New York and New Jersey when they had very similar accumulated deaths, $\sim (0.8/1000)$.

Data on social mobility extremely similar in the two, including start dates.

Daily New Deaths in New York



Daily New Deaths in New Jersey



For Belgium it is notable that despite a lockdown only 4 days after Spain (when Belgium had a tiny number of deaths), and well before England, the peak was only reached 10-14 days after Spain.

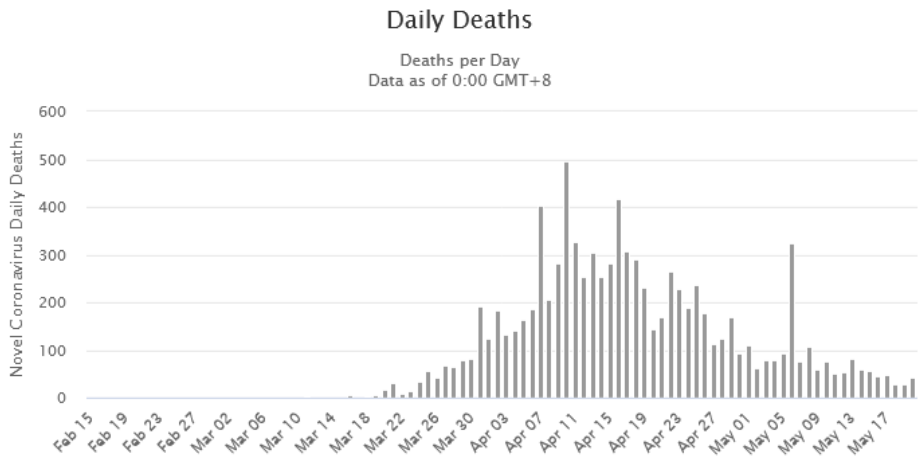
The timing fits far better with the point where z_{eff} would be significant

Assuming an ifr or variation in susceptibility the same as England, peak when $z_{eff} \sim 0.15$.

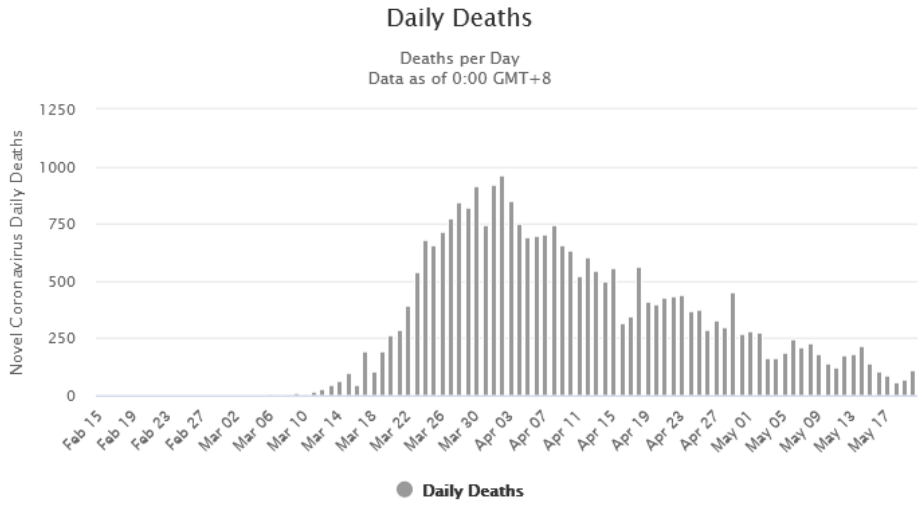
A week earlier, when we expect the peak to occur, $z_{eff} \sim 0.08$.

Now a rapid decline for Belgium, consistent with current $z_{eff} > 0.25$.

Daily New Deaths in Belgium



Daily New Deaths in Spain



(Plots from https://www.worldometers.info/coronavirus.)

Results also suggest that in the Stockholm region of Sweden the rate should be at the peak or beyond despite no lockdown in Sweden - deaths per population $\mathcal{O}(0.8/1000)$.

Appears to be true whereas Västra Götaland County, with lower number so far is now catching up.

Yesterday report Stockholm had 7.3% with antibodies three weeks ago.

Country	% of total population infected (mean [95% credible interval])	
Austria	0.76%	[0.59%-0.97%]
Belgium	8.73%	[6.73%-11.27%]
Denmark	1.06%	[0.82%-1.39%]
France	3.48%	[2.72%-4.46%]
Germany	0.89%	[0.69%-1.14%]
Greece	0.13%	[0.10%-0.18%]
Italy	4.65%	[3.75%-5.81%]
Netherlands	3.44%	[2.72%-4.36%]
Norway	0.46%	[0.35%-0.60%]
Portugal	1.11%	[0.88%-1.43%]
Spain	5.59%	[4.44%-7.07%]
Sweden	4.06%	[3.04%-5.46%]
Switzerland	1.93%	[1.52%-2.43%]
United Kingdom	5.38%	[4.22%-6.87%]

Posterior model estimates of percentage of total population infected over the course of the pandemic. Estimates as of 2020-05-07.

In general, the inferred ifr and/or variation in susceptibilty suggests that the effective number of people infected is rather higher, by a factor of about 3, than estimates in e.g. https://mrc-ide.github.io/covid19estimates/#/total-infected.

Finally, if the is correct, it does also rely on the fact that those people who have become infected are no longer susceptible to further infection, at the very least for some short period of time.

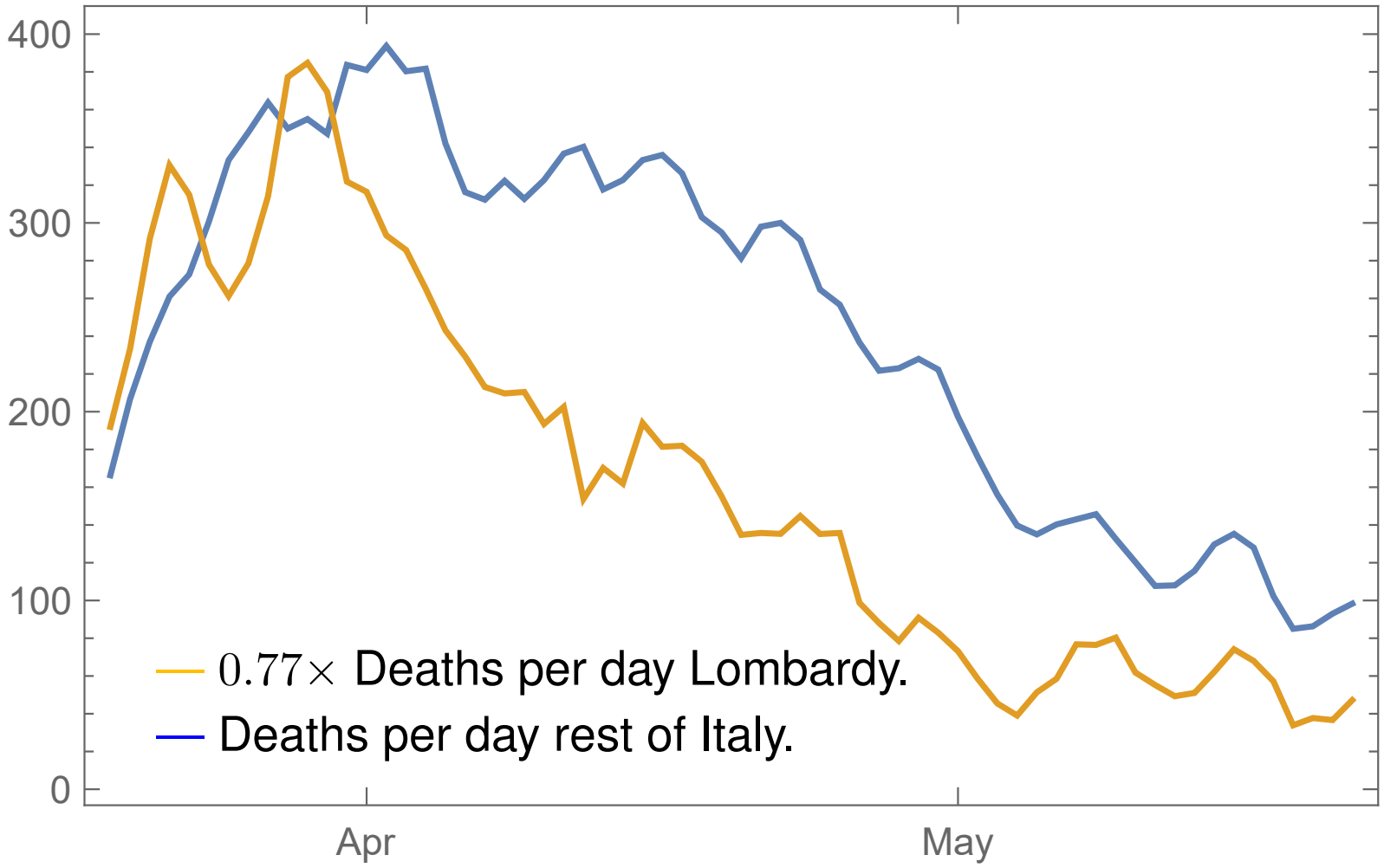
If so, then for most practical purposes it is much the same if a large effective number of the population is infected is due to a lower ifr or variation in suceptibility, or both. The implications for at what stage the infection no longer transmits freely in the population is determined by z_{eff}.

The lower ifr may seem more attractive, but actually it implies many asymptomatic or very mild cases.

The latter suggests fewer cases, so easier for e.g. “track and trace” to work, especially as fewer and fewer average transmissions will occur.

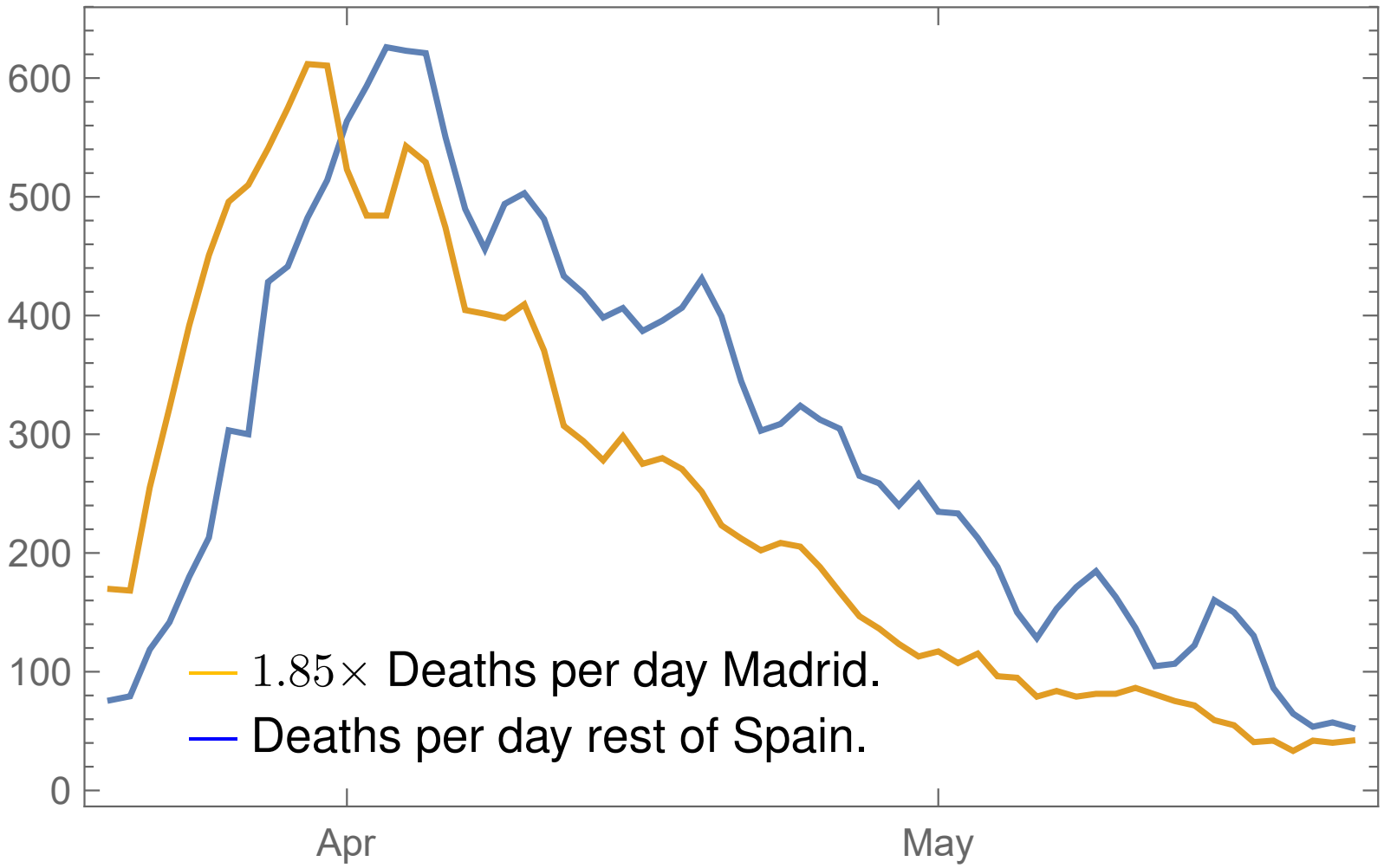
Lombardy and Italy

Up to date numbers.



Madrid and Spain

Up to date numbers.



London and England

Up to date numbers.