Correlations in US COVID-19 mortality age profiles: epidemic start dates, geography and the PCF hypothesis

T J Newman, SOLARAVUS

correspondence to: tjnewman@solaravus.com

Abstract

Using data from the United States Centers for Disease Control and the United States Census Bureau we measure normalised mortality age profiles (NMAPs) of COVID-19 for the 25 worst-affected States. There are clear trends in NMAPS with the start date of the epidemic for each State and its geographical location, with deaths increasingly concentrated in the older age groups for States with later epidemics. Deviations from this trend are correlated with the health index of each State. These findings are predicted outcomes of the recently proposed pre-conditioning field (PCF) hypothesis. A more detailed analysis is performed for the north-eastern States to analyse the effect of the initial epidemic in New York City (NYC) on neighbouring States. Strong geographical trends in NMAPs from west/south-west to east of NYC are highlighted. The NMAPs for the States of New York and Massachusetts show a remarkably precise overlap if the NY data is multiplied by a well-established immune system ageing function, confirming the inter-regional predictions of the PCF hypothesis. We briefly discuss other possible underlying causes for the measured trends, and suggest more sophisticated data analyses to test the PCF hypothesis.

Contents: *page number*

SOLARAVUS TECHNICAL REPORT

Report #: STR 004
Date: August 8th 2020
Title: Correlations in US COVID-19 mortality age profiles: epidemic start dates, geography and the PCF hypothesis
Author: T J Newman

This is a technical report from the company SOLARAVUS and will remain freely available to all interested parties via the CERN Open Science archive Zenodo and the SOLARAVUS website.

Note, given the universal interest in the current COVID-19 pandemic, this report is written with both scientists and the general public in mind as potential readers. An attempt has been made to minimise technical jargon, and to provide more comprehensive explanations of certain technical methods than would be usual in scientific articles appearing in journals.

If a citation to this report is required, please use:

Newman TJ 2020 *Correlations in US COVID-19 mortality age profiles: epidemic start dates, geography and the PCF hypothesis* Solaravus Technical Report 004
doi: 10.5281/zenodo.3976802

1. Introduction and rationale

The COVID-19 pandemic originated in Wuhan, Hubei province, China in late 2019 (Zhu *et al*. 2020), and has spread across the globe. COVID-19 is the disease caused by the SARS-CoV-2 virus (Bar-On *et al*. 2020 and references therein). To date (early August 2020) COVID-19 has infected over 18 million people with a death toll in excess of 700,000 (*worldometers.info*). The United States has been particularly badly hit, with over 5 million people infected, and a death toll in excess of 160,000 (*worldometers.info*). With 48 of the States geographically contiguous, and with centralised databases available for both COVID-19 fatalities and underlying demographics, it is possible to perform in-depth data analysis of spatio-temporal aspects of the disease spread in the US, as demonstrated in recent studies (e.g. Loomba *et al*. 2020, Mollalo *et al*. 2020, Zhang and Schwartz 2020). In this report we concentrate on the mortality age profiles due to COVID-19 for the 25 worst-affected States, and study their variation with epidemic start date and geographical location.

The rationale for studying the age profile of mortality with respect to start dates of epidemic status is to test the pre-conditioning field (PCF) hypothesis recently advanced by this author (Newman 2020). The PCF hypothesis posits that infected populations immunologically pre-condition neighbouring and more distant susceptible populations. A candidate mechanism for the PCF is viral detritus (ViDe), exhaled by infected individuals, carried by air currents, and subsequently inhaled by susceptible individuals. ViDe is hypothesised to trigger a low-level immunological response, thus providing a degree of protection on later exposure to live SARS-CoV-2 virus. Mathematical modelling has shown that the PCF hypothesis is able to explain a number of features of the COVID-19 pandemic, such as highly disparate mortality rates from one region to another, very slow decays in epidemic curves, and generally low levels of seroprevalence (as discussed in detail in Newman 2020).

Studying normalised mortality age profiles (NMAPs) for different States enables the PCF hypothesis to be robustly tested. If the hypothesis is correct then those States experiencing the epidemic weeks later than others would be expected to benefit from long-ranged pre-conditioning originating from those earlier States. Individuals will benefit if their immune systems are sensitive and selective enough to recognise the PCF (e.g. low concentrations of SARS-CoV-2 viral fragments in the ViDe instantiation). Given the steep decline of immune function with advanced age (Montecino-Rodriguez *et al*. 2013, Palmer *et al*. 2018) it is therefore expected that the most elderly sub-populations will benefit least from the PCF, and therefore, *relative to their fellow citizens in younger age classes of the same State*, they will fare worse in terms of serious outcomes, this relative worsening increasing in States whose epidemic started later.

Indeed, we find clear and statistically significant correlations in NMAPs depending on the start date of the epidemic for each State. The correlations clearly show that the elderly sub-populations in a given state tend to be more severely affected, relative to younger sub-populations, the later their State experienced the epidemic. In addition, a detailed analysis of the north-eastern States reveals highly correlated NMAPs relative to geography. The NMAP of Massachusetts differs from the NMAP for New York precisely by a well-known function describing the decline of the immune system. The precision of this relationship is remarkable in and of itself, and is strong evidence for the PCF hypothesis.

The outline of the paper is as follows. In section 2 we define our terms and the datasets we work with, and in section 3 we report our results. In section 4 we summarise the findings and discuss them in the context of the PCF hypothesis. We also examine other possible causes for the trends and discuss future work to test and exploit the PCF hypothesis.

2. Methods and definitions

In this technical section we describe the methods and definitions used in this study, with brief descriptions of the rationale for each.

2.1 Epidemic start date

We use two different definitions for the start date of "epidemic status" in a given State. Definition 1: the first date at which the State reports over 1000 cases of COVID-19; definition 2: the first date at which the State reports over 50 deaths from COVID-19. These start dates range over a three-to-four week period as shown in Table 1 (spanning mid-March to mid-April). Defining precise dates of epidemic outbreaks is not straightforward (Texier *et al*. 2016). However, given the exponential increase in cases and/or deaths at the start of the epidemic in a given region, the start date will be weakly (logarithmically) sensitive to the precise number of cases or deaths that define start dates. Population size and density vary considerably from one State to another, and so the values of 1000 cases and 50 deaths have been chosen as large enough to be relatively insensitive to statistical fluctuations and small enough to be relatively insensitive to population size/density of the State in question.

2.2 Normalised mortality age profile (NMAP)

Comparing mortality statistics between States requires care, given the large number of variables that differ from State to State, such as population size, population density, responses to the pandemic, and underlying demographic and genetic distributions. Indeed, technical notes from the US Centers for Disease Control (CDC) warn against comparing mortality data from State to State, in part because of differing protocols and time scales for defining and officially registering COVID-19 deaths. This is sensible advice and rules out naïve comparisons of absolute numbers of deaths from one State to another. That said, the contiguous United States does provide a rich dataset of the spread of the epidemic and our view is that comparisons of one State to another are possible if defined carefully. Our approach therefore is to define a relative measure of deaths which compares one sub-population in a given State to the remaining population in that same State. This relative measure will eliminate to a large degree many of the uncontrolled variables listed above. We define such a relative measure, which we denote the "normalised mortality age profile" (NMAP). The NMAP is defined across a set of age classes, determined largely by the structure of CDC and US Census Bureau datasets, which are decadal. Given the relatively small number of deaths for those under the age of 45, we do not differentiate age classes below this age, and therefore collect data for the following six age classes: <45, 45-54, 55-64, 65-74, 75-84, ≥85. For each age class in a given State we define the mortality rate as the cumulative number of deaths in that class as of a certain date divided by the total number of individuals in that class. These mortality rates will define a profile of deaths across age classes for that State, which for brevity we call the mortality age profile (MAP). It is the *shape* of the MAP that is of interest to us, not the absolute values of mortality rates which will, of course, depend on the intensity of the epidemic in that state, which in turn will be highly dependent on variables such as population density and local lockdown regulations. We therefore normalise the MAP, such that its integral (i.e. the sum of the values of the profile) is equal to unity. We call this the normalised mortality age profile (NMAP). In short, the shape of the NMAP describes the relative burden of COVID-19 deaths across age classes within a given State and can be compared from one State to another. A mathematical definition of the NMAP is given in Appendix A.1. In addition, we measure NMAPs separately for males and females in a given State, given the significant disparity of mortality rates with gender. A similar measure to the NMAP, used in epidemiology and actuarial science, is the standardised mortality rate (Everitt and Skrondal 2010, US DHHS 2012) measuring mortality relative to standardised population tables or expected values.

2.3 Datasets

To construct the NMAPs we require centralised moderated datasets at the federal level, to remove as much State-to-State administrative and reporting variability as possible. We therefore use demographic data from the US Census Bureau, which was updated in 2018; in particular we collect the number of individuals of each gender and in each of the age classes defined above. In this study we do not focus on other important demographic variables such as socio-economic status or ethnic background. Data on COVID-19 mortality is taken from the CDC, and we downloaded on July 12th 2020 the complete dataset which accounted for all confirmed COVID-19 mortalities State-by-State as of July 4th 2020. We use the cumulative number of deaths up to this date for all States, regardless of the start date of the epidemic for that State. This data will be for deaths over an approximately three-month period for all States, and as such will be relatively independent of transient effects, as the peak of the (first wave of) the epidemic will have occurred for every State prior to this data collection date. Note, subsequent second waves, particularly in the western States occurred after this date. The historical data for epidemic start-dates for each State were taken from *worldometers.info*, a website that charts cases and mortalities for each State over the entire course of the epidemic, and sources its data from State and federal databases. As discussed above, the epidemic start date will be relatively insensitive to differences in reporting State-to-State, given the exponential increase in the epidemic in each State.

2.4 States and other regions considered

Of the 50 States in the US we decided to analyse data for the 25 worst-affected States in terms of absolute number of COVID-19 deaths as of July 4th 2020 according to the CDC dataset. We denote this set of States as US25. States outside of this group had, as of July 4th 2020, absolute mortality counts well below 1000 deaths, which once distributed across age classes are increasingly non-robust for analysis due to small-number statistics. We also perform an analysis of the north-east of the US centred on NYC, and for this we calculate NMAPs for additional States outside of the US25 along with the District of Columbia. Finally, in discussing trends in NMAPs on the west coast of the US we include mention of the Canadian Province of British Columbia; thus NMAPs for an additional US State (Oregon) and the three largest Canadian Provinces are calculated (though not accounting for gender).

2.5 Health scores

Our analysis is informed by measures of health for each State. We use the 2018 annual health scores for each State produced by the United Health Foundation (their methodology uses 35 factors integrated within a broad WHO definition of "health"). The scores have values given by the standard deviation across all States, varying between approximately –1 and +1. A negative score indicates a health score below the national average and *vice versa*. We group States into categories: low, medium and high, defined by scores < –0.35, between –0.35 and +0.35, and > +0.35. UHF health scores for US25 are given in Table 1.

2.6 Immune system decline

The human thymus declines with age in both size and T-cell output (Gui *et al*. 2012, Sottini *et al*. 2014, Palmer *et al*. 2018). The decline follows an approximately exponential decay:

$$\text{immune decline} \sim \exp[-\alpha \times \text{age class}]$$

We will refer to this as the immune decline function (IDF). It has a single parameter α, the rate of immune decline, with the approximate value α =0.044 /year. The COVID-19 data is reported in decadal age classes; on that scale α = 0.44 /decade. We also define the inverse of the IDF, which we term the immune barrier function (IBF) (cf. Appendix A.2).

3. Results

In this section we present our analysis of US COVID-19 data. The results of the analysis will be of general interest, independent of the author's motivation for measuring NMAPs to test the PCF hypothesis. To aid the reader, we have broken the Results section into subsections, of which 3.2 and 3.4 provide context and motivation from the perspective of the PCF hypothesis. The remaining subsections present the results of data analysis. Nine key findings emerge from this analysis and are presented as numbered underlined statements. These findings are summarised in subsection 4.1. Detailed discussion of how the results support or refute the PCF hypothesis is provided in section 4.

3.1 NMAPs: quality of data and properties

We first illustrate the normalisation procedure that takes mortality age profiles (MAPs), which differ markedly in their magnitude from State to State, to unit integral NMAPs, thereby allowing more robust comparison between States. In Figure 1a we show the MAPs (upper panel) and NMAPs (lower panel) for NY, IN and OH. The data used to generate these profiles (and all the NMAPs for US25) are provided in Table 2. We note that the MAPs for all three States rise rapidly with increasing age class, consistent with the fact that COVID-19 mortality is strongly dependent on age. The MAP for NY is much higher than that for IN which in turn is higher than OH, reflecting the significantly different proportional mortality of COVID-19 in these three States. By normalising each MAP, such that each profile has unit area (i.e. an integral of unity), we arrive at the NMAPs shown in the lower panel. There are two striking observations to make. First, the NMAPs for IN and OH are almost identical. In fact, to show the difference between the two NMAPs we use larger open circles for the IN data to allow the OH data to be seen. Second, the NMAP for NY is flatter than the IN/OH NMAPs. This indicates that, relatively speaking, COVID-19 mortality was more spread across age classes in NY than it was in IN and OH. This will be a recurring theme in the analysis.

F1: NMAPs are well approximated by exponential functions

Given that the values across the age classes of an NMAP typically range over two to three orders of magnitude, it is necessary to plot these curves on a log-linear scale in order to see the structure of the profile across the entire set of age classes. Consequently, in Figure 1b (upper panel) we show log-linear plots of the NMAPs for NY and OH. It is remarkable that the points of the NMAP for each State lie on almost perfectly straight lines. The interpretation of this behaviour is straightforward, namely that the NMAPs are extremely well approximated by exponential functions. A recent study of age dependence of COVID-19 mortality curves has reported this exponential dependence, which is indeed well known in a similar form for all-cause mortality across age groups (Promislow 2020).

F2: NMAPS can be described by a single parameter: the NMAP index (NI)

As the NMAPs are constructed to have an integral of unity, the exponential behaviour implies that the entire NMAP can be captured by a single parameter, which we call here the NMAP index (NI). The NI is a rate, having dimensions of /decade. Explicitly, the dependence of the NMAP on age class can be written as an exponential function as follows:

$$\mathrm{NMAP} = C \exp[\mathrm{NI} \times \text{age class}] \quad . \tag{1}$$

The prefactor *C* is a constant related to NI ensuring unit area. On taking natural logarithms of both sides, this will reduce to the equation of a straight line on a log-linear plot, namely:

$$\log(\mathrm{NMAP}) = \mathrm{NI} \times \text{age class} + \log(C) \tag{2}$$

Therefore, the NI can be directly inferred as the slope of a straight line fit to the log-linear plot of the NMAP. Such straight-line fits are shown for NY and OH in Figure 1b (upper panel). Their slopes, which are the NIs, are given in Table 3, along with the NIs for all States in US25. The exponential nature of the NMAP is shown explicitly in Figure 1b (lower panel) where we plot the data points for the NMAP of OH along with the exponential curve derived from the NI for OH. The curve passes through the data points very accurately, and we see how extraordinarily well the simple exponential function characterises the OH NMAP.

This fact provides us with a useful tool to quantitatively compare the NMAPs between different States. So long as the NMAP for a given State is well approximated by an exponential function we can use the NI as a single parameter to describe the NMAP, allowing a straightforward quantitative comparison between NMAPs, and also allowing the measurement of potential NMAP correlations from one State to another.

Returning to Figure 1b (upper panel) we see that the flatter NMAP for NY compared to OH, when shown on a log-linear curve, corresponds to a straight line for NY with a lower slope than that for OH. Thus, the NI for NY will be lower than that for OH. Indeed their respective values are 0.827 (1.5%) and 1.077 (0.8%) (cf. Table 3). Given the NMAPs for OH and IN are so similar we expect their NIs to be very close, and indeed this is the case. The NI for IN is 1.088 (1.7%). We have given in each case the percentage error in the NI, which arises from fitting a straight line to the log-linear plot of the NMAP. This is a guide to how well the NMAP is approximated by an exponential function. As can be seen from Table 3, the vast majority of the States in US25 have very low percentage errors in their NIs. However, there are a few exceptions. In Figure 1c we show the NMAPs for the two States with the largest NI errors, AZ (upper panel) and AL (lower panel), along with the best fits of an exponential function. It is clear that the data is not well described by a simple exponential function, namely that the NMAPs cannot be characterised by a single parameter. The NIs for AZ and AL have percentage errors of 10.9% and 6.1%. We therefore introduce a cut-off of 5.0% as the maximum tolerable error in the NI for a given State's NMAP to be accepted as well-represented by an exponential function. This cut-off removes only AZ and AL from US25 in studying the NMAPs for total populations (i.e. including both male and female sub-populations). We have also calculated NIs for the male and female sub-populations in each State (Table 3). Four States fail the 5% cut-off within the male data, and six States fail the 5% cut-off within the female data. Thus, with a 5% error threshold we retain 84% of the possible data in using the NIs for subsequent analysis (63 of the possible 75 NMAPS: 19, 21 and 23 States of US25, respectively, for the male, female, and total population analyses).

3.2 Expectations of trends in NMAP indices (NIs) from the PCF hypothesis

A high-level schematic of the PCF hypothesis is shown in Figure 2. For a detailed motivation and explanation of the PCF hypothesis we recommend that the reader consult the original paper (Newman 2020). In brief, it is hypothesised that infected individuals are the source of an immunological pre-conditioning influence or field (denoted by PCF) which can spread far more widely than the short-range infection dynamics of the live virus. One possible instantiation of the PCF is viral detritus (ViDe); namely, that along with live virus, infected individuals emit (through exhalation and coughing) viral fragments and immune system detritus into the air. ViDe will travel long distances in air currents and will have biological potency if inhaled by distant susceptible individuals, so long as the molecular structure (e.g. folded protein conformations) of ViDe components, such as SARS-CoV-2 spike proteins, is retained under atmospheric conditions. Theoretical analysis (Newman 2020) has shown that the PCF hypothesis provides a parsimonious explanation for many features of the COVID-19 disease dynamics.

Referring to Figure 2, an epicentre of COVID-19 will have dynamics dominated by short-range contact infection. As the infected population grows, the PCF in turn increases in amplitude and will spread to neighbouring regions which are yet to experience high levels of infection. Within the PCF hypothesis, individuals in these regions, on exposure to the PCF, are immunologically pre-conditioned. The degree of pre-conditioning for an individual will vary depending on the sensitivity and selectivity of that individual's immune system. It is expected that pre-conditioning will significantly alter the overall impact of the epidemic on the neighbouring region as compared to that in the original epicentre. It is not straightforward to measure this difference in terms of absolute numbers of cases and deaths, because these depend strongly on other factors, such as the flux of infected individuals into the region, population density, underlying demographic distributions, and timing and enforcement of control measures such as lockdowns. The normalisation of the NMAP removes much of the effect of region-to-region demographic and population density differences, and differences in absolute numbers of cases and deaths. As such, through a study of NMAPs, we can seek to measure differences in neighbouring regions which might arise from the PCF. In this study of US25 we generally consider a "region" to be a State.

As mentioned above, within the PCF hypothesis pre-conditioning requires both sensitivity and selectivity of an individual's immune system in order to provoke an immunological response, providing some degree of protection on later exposure to live virus. The vigour of the immune system declines steeply with age (Montecino-Rodriguez *et al*. 2013, Palmer *et al*. 2018), and so we expect that, likewise, the efficacy of pre-conditioning falls off steeply with age. Thus, the *relative impact* of COVID-19 on the elderly compared to younger age classes will be *greater* in a pre-conditioned population than in a non-pre-conditioned population. As such, we predict that the NMAPs for regions with pre-conditioning will be shifted towards older age classes. Given the NMAPs of nearly all States have an exponential form, this predicted shift towards older age classes will be reflected by a higher NI (i.e. a straight line NMAP which is steeper when plotted on a log-linear curve).

Therefore, a prediction of the PCF hypothesis is that NIs for States neighbouring an originative epicentre will be higher than that of the epicentre itself. On a larger geographical scale, States which have remained free from COVID-19 infections for a longer time period will have experienced a more prolonged period of pre-conditioning, and so the PCF hypothesis predicts that the later a State is first exposed to the epidemic (through infected individuals eventually entering that State) the higher will be its NI compared to States which experienced the epidemic earlier.

3.3 Variation of NMAP indices with the start date of epidemics

F3: NIs are positively correlated with epidemic start dates

To test this second large-scale prediction of the PCF hypothesis we plot the NIs for US25 against the date when each State reached epidemic status. We use two different measures of such a date, as explained in subsection 2.1. Definition 1 is the date of 1000 or more cases in a particular State, and definition 2 is the date of 50 or more deaths in a particular State (cf. Table 1). NIs for each State, along with relative errors, are given in Table 3. Plots of NI versus start date are shown in Figure 3a for total populations (combined male and female sub-populations) using definition 1 (upper panel) and definition 2 (lower panel). Straight line fits are also shown. The correlation coefficients and associated *p*-values are given on the plots. We see that for both definitions of start-date there is a moderately strong positive correlation between NI values and start date of epidemic, and the correlation is statistically significant. We note that the two originative States, NY and WA, have disparate NIs. We will look at these two States in more detail in subsections 3.5 and 3.6 respectively.

F4: NIs for male sub-populations are lower than those for female sub-populations
Figures 3b and 3c show the corresponding plots for male and female sub-populations respectively. In three of the four plots there is a statistically significant positive correlation between NI values and start-date of epidemic (note the *p*-value for the male plot using definition 2 of the start-date is 0.0572 and so although the data shows a positive correlation, this lies outside the commonly used 5% definition of significance). Note that the NIs for males are lower than those for females in all the States within US25 (except for AL, but for this State all NIs fail the 5% error threshold).

F5: Deviations in the NI/start-date correlation are strongly dependent on State health scores
In all of the plots in Figure 3 we note that there are some States which deviate quite strongly from the line of best fit. Some insight into these deviations is possible by colouring the points according to the UHF health scores of the corresponding States: green, blue and purple corresponding to high, medium and low health scores respectively (see subsection 2.5). Those States which deviate strongly in the positive direction, most notably MA, CT and CO, have high health scores. Those States which deviate strongly in the negative direction, most notably MS and TX, have, respectively, low and low-medium health scores.

This is in accordance with the predictions of the PCF hypothesis. The population in a State with a lower health score will tend to have relatively higher rates of illnesses such as diabetes and cancer in individuals of younger age classes, which in turn, on average, reflects greater challenge with immune health. As such, pre-conditioning will be less effective in the younger age classes of a State with a low health score than in a State with a high health score. In such a State, the impact of COVID-19 will thereby be more spread out across age classes, and the NI consequently lower than expected based purely on start date.

By contrast, in States with high health scores, the PCF hypothesis predicts full response of pre-conditioning, which implies, *within the PCF hypothesis*, that a majority of individuals in most age classes will have a degree of buffering from COVID-19. Only the very elderly, whose immune systems are in a late stage of decline, will fail to benefit from the PCF. As such, the NMAP will be steeply tilted across the age classes yielding a higher value of the NI.

F6: The total variation in NIs is comparable to the rate of immune system decline
The NI has dimensions of /decade (i.e. per decade), and, as such, is a rate constant, measuring the rate of impact of COVID-19 as a function of age class. From Figures 3a-c it is striking that in all three sets of plots (combined, male, female) the total variation of NIs is between 0.40 to 0.45, very close to the decadal rate of decline of the immune system, α =0.44 /decade. This suggests that the difference in NIs across the States is directly related, mechanistically, to immune system function. In the following subsection we explain how this result can be naturally explained within the framework of the PCF hypothesis.

3.4 The PCF hypothesis in terms of inter-regional disease dynamics

In Figure 4 we provide a schematic figure explaining the inter-regional effects of the PCF within the classic susceptible-infected-recovered (SIR) compartment model of epidemiology. (*Intra*-regional PCF dynamics are explored in detail in Newman 2020.) In Figure 4, Region 1 experiences the epidemic well before Region 2. The PCF will increase in magnitude as the number of infected individuals in Region 1 grows, and will spread into neighbouring Region 2. In this simplified model, the PCF is assumed to act on susceptible individuals in Region 2 before the epidemic has subsequently taken hold in that region. In practice, the influence of pre-conditioning may be concurrent with a significant influx of infected individuals from Region 1 (purple arrow), and the subsequent scale of cases and mortality in Region 2 will be

strongly sensitive to the relative intensities of these two influences (influx of infected individuals and the spreading PCF) from Region 1. Assuming PCF to be the dominant influence in the short-term, all susceptible individuals in Region 2 will experience the PCF. However, the degree to which the PCF provides immunological buffering for a given individual in Region 2 will depend strongly on the sensitivity and selectivity of that individual's immune system. In particular, pre-conditioning will be significantly less effective for the elderly. Hence, as discussed in subsection 3.2, we expect the NI to be higher in Region 2 compared to Region 1. It is not possible to make predictions about case numbers and mortality rates in Region 2 compared to Region 1 because of the many confounding factors discussed earlier, such as population density, socio-economic and ethnic distributions, regional responses such as lockdowns, influx of infected individuals etc. If all of these factors were equal in both regions then the PCF hypothesis would predict significantly smaller numbers of cases and deaths in Region 2 compared to Region 1.

The key point of Figure 4 is to make clear at which point immune system health affects pre-conditioning response in Region 2 (see red arrow Figure 4). Assuming all individuals have experienced PCF the susceptible population becomes a pre-conditioned population. As infected individuals from Region 1 interact with this population, local contact infections may occur. For those individuals who have strongly benefitted from pre-conditioning, infection with live virus will cause very mild or even no apparent symptoms, and they will essentially not experience the infection – we label them in Figure 4 as "immune". For those pre-conditioned individuals in older age classes, pre-conditioning is not expected to provide immunity *per se*, but will present an additional immune hurdle for the infection to overcome if it is to cause severe infection. It is well documented (Palmer *et al*. 2018 and references therein) that the incidence of illness from many infectious diseases follows an exponential increase with age, precisely of the form $\exp(\alpha \times \text{age class})$, namely the IBF (inverse of the IDF). This has been argued to arise from stochastic fluctuations in the pathogen load in the early course of the infection, with illness occurring if an immune escape threshold is overcome, that threshold declining with age according to the IDF (Palmer *et al*. 2018).

Thus, according to this simplified model of a population which has experienced the PCF before significant influx of infected individuals, the NMAP of Region 2 will be equal to NMAP(Region 1) × IBF (the IBF factor arising from the immune barrier provided by pre-conditioning as explained above). Explicitly, one has

$$\text{NMAP (Region 2)} = \text{NMAP (Region 1)} \times C' \exp[\alpha \times \text{age class}] \quad , \tag{3}$$

where the factor *C'* is a constant ensuring that the NMAP for Region 2 has unit area. Informally, one might say that the NMAP of Region 2 is an "immune boosted" form of the NMAP for Region 1. Given that the NMAPs follow exponential functions themselves, cf. Equation (1), on taking natural logarithms of both sides we have the prediction:

$$\text{NI (Region 2)} = \text{NI (Region 1)} + \alpha \tag{4}$$

In reality this simple relationship will be potentially distorted by various factors. For example, the relationship will fail if the underlying health scores of the two regions are significantly different. The relationship will also fail if there is a very high rate of influx of infected individuals from Region 1 into Region 2 (primarily commuters) prior to a significant period of pre-conditioning. Generally, we would expect pre-conditioning to provide benefit to age classes with strong immune health, and predict that $\text{NI (Region 2)} > \text{NI (Region 1)}$, which is consistent with the data analysis described in the previous subsection. To test Equation (4) requires a study of neighbouring States, so we turn to an in-depth analysis of the US region surrounding the major COVID-19 epicentre of the NYC metropolitan area.

3.5 In-depth analysis of the north-eastern States

NY State and neighbouring States such as NJ, MA and CT have experienced some of the highest COVID-19 mortality rates of any region in the world to date (Richardson *et al*. 2020, *worldometers.info*). NY and WA (on the west coast of the US) experienced epidemic status prior to other States in the US. Within 3 weeks, all US25 States were experiencing the epidemic (according to definition 1, as described in section 2). Given that the State of NY contains one of the largest metropolitan areas in the US, it is helpful to differentiate the State level epidemic from that of NYC. This is possible as the CDC dataset collects data for NYC and NY State separately (cf. Table 4a). For the purpose of analysing the north-eastern States surrounding NYC we also look at mortality data for north-eastern States outside of US25, namely DE and NH, along with the District of Columbia (DC) (cf. Table 4a). Robust statistical analysis is not possible for additional States in the region (ME, VT and WV) due to their relatively low numbers of COVID-19 deaths.

F7: NIs are directionally clustered about NYC increasing from west to east

Figure 5 provides a simplified map of the north-eastern US. Each State is labelled with its abbreviation along with its NMAP index (NI). The NIs for three additional regions outside of US25 (cf. Table 5) are statistically less robust due to their relatively small number of deaths: the NIs of DC and DE have errors above the 5% threshold, whilst NH has missing data in half of the age classes, due to the privacy protocol of the CDC (applied when there are less than 10 deaths per age class). The less robust NIs are shown in Figure 5 in parentheses.

It is noteworthy that the NI for NYC (0.78) is significantly lower than that of NY State when NYC is excluded (0.95), consonant with the predictions of the PCF hypothesis; namely, that the NI will be lowest in an originative epidemic epicentre. The NI for NY State as a whole is 0.83, close to the NYC value, showing the dominance of the NYC metropolitan area in the NMAP for the State as a whole. The geographical clustering of NI values is striking. For those States in US25 east of NYC (i.e. the New England States) the NIs have values 1.18, 1.18 and 1.25 (for CT, RI and MA respectively). For those States west and south-west of NYC, the NIs have values 0.92, 1.00, 1.06, 1.06 (for NJ, MD, PA and VA respectively). In light of the discussion of inter-regional effects of PCF, and considering NYC as Region 1 and New England as Region 2, we expect NI (New England) = NI (NYC) + $\alpha = 0.78 + 0.44 = 1.22$, (cf. Equation 4). The NIs for the New England States are indeed clustered around this value. It is noteworthy that the NIs for the States to the south and west lie almost exactly midway between the NYC and New England values, possibly due to partial pre-conditioning.

As discussed in subsection 3.4, the PCF prediction given in Equation (4) will be valid for Region 2 assuming that region is able to respond fully to the PCF. Indeed, those States in New England are in the category of high UHF health scores. Consistent with this, the States to the west and south of NYC, with the exception of NJ, are in the category of medium UHF health scores, and thus, within the PCF hypothesis, might not be expected to respond fully to pre-conditioning. Further discussion of these geographical correlations is provided in section 4.

F8: NI (MA) = NI (NY) + rate of immune system decline

To illustrate the impact of Equation (4), it is instructive to compare the NMAPs for NYC and MA. (We choose MA as it is the most populous and distant New England State of MA, CT and RI, thereby having more robust statistics and presumably proportionately less influx of infected individuals through commuter traffic). In Figure 6a we show the NMAPS for NYC and MA on both linear and log-linear scales. The straightness of the lines illustrates how well these NMAPs are approximated by the exponential function (indeed, the percentage errors

in the NIs for NYC and MA are well below the 5% threshold, being 3.5% and 1.1% respectively). Following the discussion of subsection 3.4, we can ask how closely will the "immune boosted" NMAP for NYC compare to the NMAP for MA. In Figure 6b we show the NYC NMAP data boosted by the IBF, alongside the MA NMAP. The degree of overlap is very high. It is necessary to use larger open circles as symbols for the NYC immune boosted data in order for some of the MA data to be visible.

More consistent with the inter-regional conceptualisation of Figure 4 is to compare equivalent regions, i.e. one State to another. Thus, we compare the "immune boosted" NY State with MA. The NI for the NY State is 0.83, close to the value for NYC itself, and this value will be dominated by COVID-19 deaths in the wider NYC metropolitan area. We might expect an even higher degree of overlap than the comparison of immune-boosted NYC to MA, since the NY State and MA NMAPs are both based on State-level demographics (e.g. both containing urban, suburban and rural communities). The comparison of NY State and MA NMAPs without and with the immune boost to NY State are shown in Figures 6c and 6d respectively. The precision of overlap is remarkable. These findings provide very strong support for the inter-regional predictions of the PCF hypothesis.

3.6 A preliminary analysis of the west coast States

The analysis of the north-eastern US provides an insight into the geographical correlations between NIs. It will be interesting to perform similar analyses across the US (e.g. across the Mid-West and across the southern States), but this is beyond the scope of the current Report. However, given the important role of WA as a co-originative State, along with NY, we provide a preliminary view of the west coast States.

F9: NIs increase from south to north along the west coast of US and S. Canada

We first note that the NI for WA is significantly higher than for NY, with the value 1.05. Interestingly, this is similar to the NI values of the States to the south and south-west of NYC, which within the PCF hypothesis would be considered as partially pre-conditioned. It is therefore interesting to ask whether it is possible that WA experienced weak pre-conditioning prior to its epidemic status. If there was an early pre-conditioning influence in that region, we might expect to see a trend in NI values along the west coast. In addition to the NI for CA already to hand we therefore collect data for Oregon and for the region north of WA, namely the Province of British Columbia (BC) in Canada (cf. Tables 4.1 and 4.2). The calculated values of NIs for these four regions are shown in Figure 7 (cf. Table 5). Note that the value for OR is statistically less robust (due to relatively low COVID-19 mortalities) and is given in parentheses.

There is a clear and uniform trend of increasing NI values from south to north, ranging from 0.89 in CA to 1.25 in BC. To get a sense of the robustness of the NI value for BC we looked at the two largest Provinces in Canada, namely Ontario (ON) and Quebec (QC), and calculated their NIs (cf. Table 4b and Table 5). We find that the BC value is similar to those of ON and QC (1.14 and 1.28 respectively). All three of the States on the west coast have positive UHF health scores (that for OR is 0.295, at the upper end of the medium range, cf. subsection 2.5). An interpretation of these NI values within the PCF hypothesis is briefly explored in section 4.

4. Summary and discussion

4.1 Summary of data analysis

In this Report we have presented a detailed analysis of the age profiles of COVID-19 mortality across the US. We have used CDC data for the period 02/01/20 to 07/04/20, and in order to achieve robust statistical analysis we have concentrated on the 25 States with the highest numbers of COVID-19 mortalities, this group denoted by US25.

We have defined a normalised mortality age profile (NMAP), cf. subsection 2.2 and Appendix A.1, which through normalisation is less dependent on variables that differ intrinsically from State to State, such as demographic distributions, population density, and manner of enforcement of lockdown procedures. As such, NMAPs can be calculated for each State in US25 and meaningfully compared. The NMAP describes how the burden of COVID-19 mortality is spread across age classes within a given State. A flatter NMAP indicates that mortality is more widely spread in age groups outside of the elderly, whilst a steeper NMAP indicates that mortality is very highly concentrated in the elderly population of that State.

The decision to study NMAPs was originally made in order to test the PCF hypothesis (Newman 2020), and we will discuss how the analysis of NMAPs has informed this hypothesis in subsection 4.2. The NMAP results are, however, interesting in their own right, and so we first summarise the set of nine findings resulting from the analysis. (These numbered findings will be referred to in bold font in the discussion below.)

F#	Summary
F1	NMAPs are generally well approximated by exponential functions
F2	NMAPs can be described by a single parameter: the NMAP index (NI)
F3	NIs are positively correlated with epidemic start dates
F4	NIs for male sub-populations are lower than those for female sub-populations
F5	Deviations in the NI/start-date correlation are strongly dependent on State health scores
F6	The total variation in NIs is comparable to the rate of immune system decline
F7	NIs are directionally clustered about NYC increasing from west to east
F8	NI (MA) = NI (NY) + rate of immune system decline
F9	NIs increase from south to north along the west coast of US and S. Canada

F1 is a critical property of NMAPs (cf. Figure 1). It begs the question: why are COVID-19 mortality curves so well approximated by the exponential function? For our purposes it allows a simple quantitative comparison between the NMAPs of different States, since the normalised exponential function is described by a single parameter, which we call the NMAP index (NI). This is the content of **F2**. To determine the NI for a given State, we plot the NMAP on a log-linear scale, yielding a set of points approximately lying along a straight line. The NI is the slope of a best-line fit through that set of points. We find that 23 of the 25 NMAPs for US25 are well-approximated by exponential functions, having errors in the fitting of NIs ≤ 5% (cf. Table 3). Likewise, when analysing NMAPs for male and female sub-populations, 40 of the 50 NMAPs satisfy the 5% error threshold. Thus, 84% (63/75) of the original data can be subsequently analysed and interpreted using NIs. The exceptions, those NMAPs which are not describable by a single parameter, presumably have exceptional underlying causes for this, possibly relating to unusual COVID-19 mortality rates in younger age classes.

We defined for each State in US25 two different measures of epidemic start date (based on critical numbers of either cases or deaths, thereby providing additional robustness to the analysis). We plotted NIs for US25 against epidemic start dates for total populations in each State (Figure 3a) and male and female sub-populations in each State (Figures 3b and 3c respectively). As summarised in **F3**, in all six cases we found that there was a positive correlation which for five of the six cases was statistically significant at the 5% level (the plot for male sub-populations, definition 2 for start dates, has a *p*-value slightly greater than 0.05). This is one of the major results of the Report. There is a clear trend in the burden of mortality across age groups in a given State depending on the date at which that State first experienced the epidemic. States experiencing the epidemic later in time have higher NIs, which implies that in those States the elderly are carrying more of the burden of COVID-19 mortality relative to younger sub-populations in that same State. One can rephrase this equally accurately as: in States experiencing the epidemic later, the younger age classes are carrying less of the burden of COVID-19 mortality relative to the more elderly sub-populations. This is a direct prediction of the PCF hypothesis (cf. subsection 4.2) but may be due to, for example, sociological rather than biological causes (cf. subsection 4.3).

F4 states that in all cases where the NIs satisfy the 5% threshold the NI for females is greater than that for males in a given State (cf. Figures 3b, 3c and Table 3), and this is regardless of the start date of the epidemic in a given State. This finding is similar to (though not equivalent to) the fact that male mortality outweighs female mortality from COVID-19 (cf. Table 2, Dowd *et al*. 2020). This bias is consistent with the higher burden of many infectious diseases for men as compared to women (Guerra-Silveira and Abad-French 2013). Crucially, though, since the NMAPs and associated NIs are normalised measures, **F4** means that in females the relative burden of COVID-19 mortality is *non-linearly* shifted to higher age classes than in males (potentially related to gender differences in rate of immune decline).

In all six plots shown in Figure 3 there are States which deviate significantly from the line of best fit. The points representing the States are coloured green (high health score), blue (medium health score) and purple (low health score) using the 2018 UHF dataset. We see that the outliers well above the line tend to be green, whilst the outliers well below the line tend to be purple (or blue with low health scores within the wide definition of "medium"), as summarised by **F5**. Thus, States with lower health scores tend to have NMAPs that are relatively less sensitive to the epidemic start date, whilst States with higher health scores tend to have NMAPs that are relatively more sensitive to the epidemic start date. This is a prediction of the PCF hypothesis as discussed below. Finally, as summarised by **F6**, we see from Table 3 and Figure 3 that, for male, female and combined populations the total span of NIs is between 0.4 and 0.45. This is precisely in the range of the rate of immune system decline α =0.44 /decade. It indicates that States with later epidemics have, relative to earlier States, a shift in mortality to higher age classes which is commensurate with the ageing impact of immune decline. This is a noteworthy finding in its own right. It provides very strong support for the PCF hypothesis as discussed below.

Shifting from temporal to spatial correlations, Figure 5 shows the range of NIs for the States (and DC) in the north-east of the US. Some additional NIs have been calculated for States outside of US25. Because of their lower relative COVID-19 mortality these NIs are less statistically robust and are shown in parentheses in Figure 5 for purposes of completeness and transparency, but are not used in subsequent analysis. As expressed in **F7**, there is a striking clustering of NI values about the epicentre of NYC (NI $\simeq$ 0.78). The States to the west and south-west of NYC have NI values clustered around 1.00 ($\simeq$ 0.92, 1.00, 1.06, 1.06 for NJ, MD, PA and VA respectively), whilst the States to the east of NYC have NI values clustered

around 1.20 ($\simeq$ 1.18, 1.18, 1.25 for CT, RI and MA respectively). This clustering may be due to intrinsically different common factors between these two States, such as health scores, which tend to be medium in the first group and high in the second. It might also indicate a geographical or environmental factor influencing the burden of COVID-19 mortality across age classes. We will discuss this below in terms of the PCF hypotheses.

F8 summarises a remarkably precise finding; a prediction of the PCF hypothesis, and an intriguing fact in its own right. The NIs for the New England States differ from that of NYC by an amount almost exactly equal to the rate of immune system decline α =0.44 /decade. In particular, Figures 6a and 6b show first the NMAPs for NYC and MA, and then replotted with the NYC NMAP boosted by the immune barrier function (IBF) (an exponential function with α =0.44 /decade). The overlap between the MA and "immune boosted" NYC NMAPs is very high. A corresponding analysis, consistent with the inter-regional dynamics of Figure 4, compares MA to NY State, and is given in Figures 6c and 6d. The overlap between the NMAPs of MA and "immune boosted" NY State is precise to a remarkable degree.

The analysis of spatial correlations of NIs in the north-eastern region of the US shows spatial correlations in the form of clustering, and indicates that similar analyses are worth pursuing for other regions of the US. Given the finite scope of this Report, we do not pursue these analyses here, beyond a brief survey of the States surrounding the other originative State, namely WA on the north-western coast. States directly inland of WA cannot be analysed due to low relative numbers of COVID-19 mortalities, but an analysis is possible south to north along the western coast of US/S Canada taking in CA, OR, WA and BC. As expressed in **F9**, we find a south to north trend of increasing NIs. With only four data-points, statistical power is low, so this finding should be viewed as indicative only, and is discussed briefly below.

4.2 Findings in the context of the PCF hypothesis

The original motivation for developing a robust measure for each State that could be then compared across States, was to test a prediction of the PCF hypothesis; namely, consistent trends of the impact of COVID-19 with respect to the start dates of the epidemic in each State. Thus, NMAPs were developed, and the fact that they follow exponential curves and are thereby describable by a single parameter, the NI, is a serendipitous empirical finding greatly simplifying subsequent data analysis. **F1** and **F2** are consonant with the exponential nature of age mortality curves, and recent analysis for COVID mortality (Promislow 2020).

We refer the reader to the recent SOLARAVUS Report (Newman 2020) for the motivation and development of the PCF hypothesis. The core idea is that infected individuals are the source of an influence that immunologically pre-conditions susceptible individuals over larger distances than short-range contact infection from live virus (cf. Figure 2). This influence is termed the pre-conditioning field (PCF). The use of the word "field" is not common in the life and biomedical sciences, but is the standard term in the natural sciences for an influence that is distributed over space. One possible instantiation of the PCF, that is discussed at great length in the previous Report, is viral detritus (ViDe). That is, the non-live viral fragments and associated molecular debris emitted by infected individuals through exhalation and coughing, including potentially bioactive molecules from the immune response. ViDe is hypothesised to retain biological potency after being carried potentially quite long distances by air currents, and could then provoke a weak but non-trivial immune response in distant susceptible individuals. Global estimates in the previous Report showed that with 1 million infected individuals, the concentration of ViDe was sufficient that each individual on Earth would inhale of order one thousand molecular ViDe fragments per day. It might be helpful for the reader to have ViDe as their mental model of the PCF framework,

but it is important to stress that the general predictions of the framework are not dependent at this stage on any particular instantiation of the PCF, such as ViDe.

The PCF hypothesis provides a clear prediction for the shift in NMAPs that would occur in a pre-conditioned population relative to a non-pre-conditioned population, and this is summarised in Figure 4. A description of this Figure explaining the role of the PCF is given in subsection 3.4. The key point is that a pre-conditioned individual presents the virus with an additional immune hurdle to overcome. The height of the hurdle, so to speak, depends on immune health, i.e. the ability of that individual's immune system to respond to the PCF. The dominant determinant of immune health is age (Montecino-Rodriguez *et al*. 2013, Palmer *et al*. 2018). Therefore, we expect that the effectiveness of pre-conditioning decreases with increasing age. Consequently the elderly in a pre-conditioned population experience a "double hit" relative to younger age classes – increased risk of mortality with age from COVID-19 and a decreased benefit from pre-conditioning. This second hit steepens the NMAP, and so we expect from the PCF hypothesis that NMAPs will tend to be steeper the longer a State has benefitted from pre-conditioning prior to epidemic status.

This prediction is borne out very clearly as described in **F3**. From the PCF perspective, we would expect to see deviations from the average trend based on the general underlying immune health of the population, as stated in **F5**. States with high immune health would benefit most from pre-conditioning, and contrariwise, States with low immune health would benefit least. In a State with poor general immune health, we would expect only modest impact from pre-conditioning, and in such a State there is barely a "double hit", so to speak, on the elderly population. As such, such States would tend to have lower than expected NIs, more similar to NIs of non-pre-conditioned States.

The decline in immune health with age is driven by the inexorable decline of the thymus starting at early age, which in turn leads to a decline in T-cell diversity with age. This decline is well known to be approximately exponential, with the thymus approximately halving in size every 16 years. The rate of decline, which we denote by α, has the approximate value α =0.44 /decade (Sottini *et al*. 2014). This exponential immune decline has been studied in some detail (Palmer *et al*. 2018) and theoretical arguments predict consequent exponential increases in disease incidence with age which are in good agreement with data, for several infectious diseases (bacterial and viral) and a wide range of cancers. More recently, the role of the decaying thymus has been studied in relation to the age dependence of hospitalisations from COVID-19 (Donnelly and Palmer 2020) and discussed as potentially pivotal to the wider immune response to SARS-CoV-2 (Mourreau *et al*. 2020). We assume that the age dependence of the response to PCF follows this same exponential dependence, and we term this the immune barrier function (IBF) (cf. subsection 2.6 and Appendix A.2).

According to Figure 4, the additional IBF, which SARS-CoV-2 has to overcome in a pre-conditioned population, leads to a quantitative prediction on the steepening of the NMAP. In "ideal" pre-conditioning circumstances, i.e. a population with a high health score and a sufficiently strong PCF, we would expect nearly all COVID-19 cases to have been filtered by the IBF, and as such mortality rates across age class would occur according to the *product* of the "original" NMAP (which is a function of the disease dynamics in a non-pre-conditioned individual) and the IBF. Given that both NMAPs and the IBF are exponential functions of age class, this leads to the very simple prediction, namely Equation 4 in subsection 3.4 – an immune boost to the NI by an amount α =0.44 /decade. Thus, **F6** is a direct prediction of the PCF as embodied in Figure 4, and it is striking that those States which approach the greatest boost in their NIs are CT, MA and MN, the States with the three highest health scores.

One might argue that NIs are sure to be greater with increasing health score, based on an argument that the elderly will have weaker immune systems in all populations, but that younger age classes will have stronger immune systems in healthier States. The second part of this statement is no doubt true, and is used above in arguing why PCF predicts deviations from the NI/age class trend in terms of health scores, but this argument alone cannot explain the trend in Figure 3. Note that States with early epidemics, such as NY, NJ and CA, also have high health scores but low NIs. Within the PCF hypothesis, this is simply because those States have had no (or slight) benefit from pre-conditioning prior to epidemic onset.

Another test of the PCF hypothesis is possible from Figure 3, by looking at the largest variation of NIs for the male and female sub-population data (Figures 3b and 3c respectively). The thymus is known to decline faster for males than females (Gui *et al*. 2012) This is thought to underlie some of the differences in disease susceptibility between the genders, and has been argued to account for the widespread differences in cancer incidence between men and women (Palmer *et al*. 2018). Precise values of the α parameter are not available for separate genders, but generally we have $\alpha_{\text{male}} > \alpha_{\text{female}}$. As such, the maximal immune boost from pre-conditioning should be greater in males than females. Thus, the largest variations in NI for the male sub-populations should be greater than for the female sub-populations. This is indeed the case, as seen from Table 3, with the largest variation in the male data being 0.452 compared to that in the female data, which is 0.422. These values depend on outliers are so are not on the same footing of statistical robustness as the nine main findings summarised in subsection 4.1.

A critical aspect of the PCF hypothesis, which is missing from the schematic in Figure 4, is the role of geography. If the PCF is a spreading influence from an infected population, then regions experiencing a higher concentration of PCF will benefit more than regions with a lower concentration. The concentration of PCF may not be solely related to distance from the infected population, because of environmental factors. For example, if ViDe is considered as the instantiation of the PCF, then it will be carried by air currents, and concentration will be mainly determined by wind direction rather than distance from the epicentre of infection.

Thus, on taking the role of space and geography into account, the PCF hypothesis does not predict a perfect correlation between NI and epidemic start date. Critical too are the roles of geographical distance and PCF concentration in relation to the disease epicentre(s). Such a spatio-temporal study on the national scale of the US is possible, but beyond the scope of this Report. As a first step, though, we can analyse a region of the US instead of the entire country, and this was the motivation to study the north-eastern States surrounding the major US COVID-19 epicentre of NYC. What is striking from **F7** and Figure 5 is the clustering of NI values around NYC. Those States to the south-west and west of NYC have NIs clustered around 1.00, whilst those States to the east of NYC (i.e. New England) have NIs clustered around 1.20. A deeper analysis is required to tease apart the underlying reasons for this clustering. The metropolitan area of NYC extends into some neighbouring States, and the commuter traffic between these States in the weeks before lockdown will have been very high. The scale of mobility of infected individuals in the region will play a major role in the net efficacy of pre-conditioning in the weeks following the NYC epidemic. Nevertheless, one can explore some consequences of the PCF hypothesis with NMAPs alone.

For example, given the high health score of MA and its relatively larger distance from NYC (somewhat mitigating the flux of infected individuals from NYC into the State), and assuming

that it experienced a high level of PCF, one can seek a test of Equation 4 in comparing the NIs of NYC and MA, and also NY State with MA (given the greater NYC metropolitan area dominates the NY State data). This is the content of **F8**. The fact that the "immune boosted" NMAP for NY State is almost perfectly coincident with the NMAP of MA is a remarkable fact. It provides a tremendous degree of support for the PCF hypothesis. Regardless of PCF, it is a fact that deserves to be understood. That it is a prediction of the PCF hypothesis is important in building the case for the PCF idea, but deeper testing of this idea is of course still required. It is often said that "extraordinary claims require extraordinary evidence", and in that regard **F8** provides some extraordinary evidence supporting the PCF hypothesis.

The clustering of NI values about NYC, when viewed through the lens of the ViDe instantiation of the PCF, provides another test. Given that NI values are higher to the east of NYC, this might imply stronger levels of pre-conditioning, and for ViDe this requires dominant air currents from west to east in that region over the early period of the epidemic. A detailed study of such a question requires the expertise of an atmospheric scientist. However, a preliminary analysis can be performed by looking at wind direction data for major cities in the region around NYC (obtained from *weatheronline.co.uk* for the period March to April 2020). Calculating the net east-west component of wind directions (namely a vector sum of the compass directions weighted by the percentage of wind from that direction over the period), we find that for each of the major cities in NJ, NY, CT, RI and MA (namely Newark, NYC, Hartford, Providence and Boston) the net component of wind along the west-east axis was predominantly westerly (i.e. from the west), by a factor ranging from 25% to over 100%. Thus, in the period March-April 2020, airborne ViDe emitted in the NYC epicentre would on average, along the west-east axis, be taken east, spreading over New England. As mentioned above, a quantitative analysis requires a far more detailed analysis, since air currents will have different directions depending on altitude, and also because the PCF would have new sources arising as the epidemic spreads into neighbouring States. We appeal to interdisciplinary teams of atmospheric scientists, biophysicists, epidemiologists and immunologists to tackle this interesting and challenging project, should they be convinced that the PCF hypothesis is plausible enough to deserve more in-depth study.

The very brief analysis of the west coast States (and British Columbia in southern Canada) is required even in this first Report on NIs as WA State is significant in experiencing the epidemic very early (either contemporaneous with or slightly earlier than NYC depending on definition of start dates). The NI for WA is moderately high, at 1.05, and raises an obvious challenge to the PCF hypothesis. The NI for WA implies a degree of pre-conditioning, yet it is an originative State in the US in terms of COVID-19. **F9** is a first step towards answering this question. We see that WA's NI is part of a larger trend of increasing NIs from south to north along the west coast. In the spirit of the PCF hypothesis this would imply that the west coast States had already experienced some form of pre-conditioning, increasing from south to north. This was in fact discussed in the first Report (Newman 2020), where it was pointed out that regions experiencing severe epidemic, such as the US eastern seaboard and Brazil, are on the opposite side of the globe from the original epicentre of Wuhan, China. Thinking in terms of ViDe, and considering the fact that Pacific air currents in the band 35^{o}–65^{o} latitude are strongly westerly (AMS 2020), the west coast States would have received higher levels of PCF than the east of the US. Experts in atmospheric science can provide more detailed analysis, but it is noteworthy that the band of Westerlies starts just north of Los Angeles (ca. 34^{o}), and that its centre (where presumably the flux is greatest) lies at the same latitude as Vancouver (ca. 50^{o}), thus increasing south to north as the NIs would indicate in Figure 7. We note that this continental scale of PCF dynamics also helps to explain the remarkably low levels of COVID-19 mortality in Japan in the earlier phase of the pandemic

(*worldometers.info*). It might appear counter-intuitive, at first glance, to seriously consider biological relevance of continental scale dispersal of ViDe, and we therefore refer the reader to recent estimates from an order-of-magnitude analysis (Appendix 1, Newman 2020).

On a final technical note in this subsection, we stress that the inter-regional conceptualisation of the PCF hypothesis in Figure 4 is more detailed than the simplest compartment model of PCF described in the first Report (Newman 2020, Figure 1), which is essentially an *intra*-regional conceptualisation, modelling the effect of a PCF generated within a region on the subsequent disease dynamics within the same region. In addition, that simplest of models explicitly disregards age classes, and labels individuals either as susceptible or pre-conditioned. By introducing age classes in Figure 4 we are introducing a slightly more complicated model for the sake of realism, in which now all individuals are pre-conditioned, but by varying degrees depending on age class. Infected individuals will arise from this heterogeneous pre-conditioned compartment, filtered by the IBF.

4.3 Alternative explanations for key findings

We offer a brief discussion of alternative explanations for the more striking findings from this analysis. It is possible that similar arguments have already been made on related data in the exponentially growing COVID-19 literature, and we apologise here if key arguments already published or in preprint form are unwittingly duplicated in this subsection without reference. Given this author is advancing the PCF hypothesis, and finding it to be generally supported by the data analysis in this Report, the onus is on others in the community to support and/or refute it with new insights and analyses, should they be so inclined.

Turning first to **F3**–**F6** which arise mainly from Figure 3, the clear fact that needs to be explained is what caused NMAPS to shift towards older ages in States that experienced the epidemic later. This finding shows that the NMAP of the epidemic in a given State is very sensitive to the absolute time at which that State first experienced the epidemic, relative to the course of the larger scale epidemic throughout the US. It is very different from the recent simulation result predicting that in a given region the burden of mortality shifts to older age groups during the course of the epidemic in that region (Davies *et al*. 2020). It is also important to stress that this finding is not concerned with absolute numbers of cases or mortalities within a given State, but about the cumulative relative impact on different age classes within the same State. An explanation for this dependence on start date that comes to mind readily is sociological; States that experienced the epidemic later had more time to prepare. For example, MN had three weeks to prepare, from the epidemic status of NY and WA in mid March, until its own epidemic status in early April. One would presumably then argue that more time to prepare allows a State in particular to optimise its protection of the most vulnerable members of the population, particularly the elderly, through lockdowns and other preventative measures. This argument then predicts relatively less exposure of the elderly to infection, and hence lower levels of mortality in the highest age classes – precisely the opposite trend to what is in fact observed. It may indeed be the case that this negative correlation is "hiding" in the data but being overwhelmed by the positive correlation, that we hypothesise is being caused by PCF. However, this natural sociological explanation (assuming equal rates of compliance with lockdown measures across age groups) does indeed predict a correlation, but a negative one, and completely at odds with the data. A negative correlation, at odds with the data, would also arise from a similar argument that medical treatment of the elderly improved over the course of the US epidemic, thereby relatively benefitting elderly patients in States with later start dates compared to elderly patients in States with earlier start dates.

Perhaps “start date” is a red herring? Detailed US studies are emerging, examining the dependence of US COVID-19 mortality on a multifactorial range of demographic and environmental factors (e.g. Loomba *et al*. 2020, Mollalo *et al*. 2020, Zhang and Schwartz 2020). Perhaps there are intrinsic properties of States that explain the differences in NMAPs between States, and which coincidentally give rise to the correlations shown in Figure 3. In other words, are the correlations, despite their statistical significance (which are in some cases quite strong), just coincidences after all? To answer this question requires such intrinsic factors to be identified, and then to show that the statistical correlation of NIs with the variation of the intrinsic factor is stronger than the correlations shown here in terms of start date. Such an explanation would also need to account for **F5** and **F6**. The first of these is not surprising, but the second is, and it might start to stretch the credulity of the objective observer if this too were to be labelled as coincidence. As such, it does appear that the decline of the immune system is playing a key role in determining the NMAPs, and so an intrinsic factor underlying NMAP variation would likely be closely related to immune health.

Considering the geographical clustering of NIs seen in **F7**–**F9**, this provides a good testing ground for identification and analysis of possible intrinsic properties of States that might explain trends in NIs. The intrinsic factor would need to have a common value across those States south-west and west of NYC, and a different common value across New England. And, these common values would need to be different in both cases to the value of that intrinsic factor in NYC. This seems difficult to achieve with only one factor, and, more likely, weighted combinations of factors may be necessary to attempt to describe the clustering of NIs.

4.4 PCF: further tests, relevance and application

The analysis of NMAPs across the contiguous US has revealed some fascinating properties of the impact of COVID-19 across age classes, both as a function of start date of epidemic and geography. It would therefore be of interest to analyse similar datasets for other large regions of the world, e.g. continental Europe and S. America. Key will be moderated datasets across the entire region to minimise intrinsic differences in reporting COVID-19 mortalities, thus allowing a relatively unbiased comparison of NMAPs from one country to the next.

A related test is to study the time evolution of the NMAP within one country. There will be potential variations due to changing social conditions, e.g. introduction of or ending of lockdowns, but over a period of relatively unchanging social behaviours it might be possible to test the PCF hypothesis. The PCF prediction is that intra-regional pre-conditioning will increase over the period that infection numbers increase, and as such the NMAP will steepen over time, until some form of saturation is reached. This shift will not be observable if a significant PCF already existed in the region, having originated from elsewhere prior to high levels of infection in the region, so it will be important to study this effect in a region that was isolated from PCF influences prior to epidemic. Of course, this requirement is difficult to meet if the precise mechanism of PCF is unknown (should it even exist), but, for example, if ViDe is taken to be the instantiation of the PCF, one can select a region which was relatively isolated from global air currents from prior areas of high infection. Brazil might serve as a good candidate for this study, or, indeed, NYC itself.

The PCF hypothesis is unconventional and the natural reaction of many life scientists and biomedical experts will be to doubt it, particularly when thinking through the mechanistic aspects of, for example, the ViDe instantiation of the PCF. This study was conducted for precisely that reason – to help support or refute the PCF hypothesis by studying hard data. Looking back over the study, we see that the findings in this Report generally support the PCF hypothesis, in some cases very strongly and with surprising precision. Thus, despite the need to continue to treat the PCF hypothesis with scientific conservatism, and to devise

stronger and better tests to support or refute it, the hypothesis has evidential support and we are justified in thinking how to take the PCF hypothesis forward, assuming it is correct.

There is an imperative here for action and innovation. We are living in unprecedented times and the pandemic has yet to abate. What then are consequences of the PCF hypothesis?

Humans innovate, and the PCF concept appears to offer many opportunities for innovation to better protect communities from COVID-19 and subsequent infectious disease epidemics. Early pre-conditioning, by definition within the hypothesis, provides buffering against infection. It is increasingly ineffective with age (or more generally in those with immuno-compromised status), and so enhancement of the PCF (or enhanced preventative measures) for these individuals would be the primary technological response required. This, of course, requires some knowledge of the PCF itself, but a good candidate appears to be ViDe, and this can be experimentally tested relatively quickly. If ViDe is indeed supported by such experiments, then the hypothesised PCF might be enhanced through administration of carefully optimised and tested ViDe directly to individuals.

Equally importantly, to the mind of the author, is the wider long-term consequence – that PCF arises as an emergent population-level immunological response. In the first Report (Newman 2020) this was discussed in terms of evolution and sociobiology. Given many species (from bacteria to amoebae to insects to fish, birds and mammals) have evolved complex biological responses at the population level, then why would humans have foregone such valuable evolutionary innovations? If our species has dispensed with such population-level responses, then our immune systems are truly independent, one human from another. Consequently each individual's immune system has to "craft its own key to unlock a new pathogen"; this appears to be an enormous waste of effort, compared to sharing immunological keys and solutions. This, ultimately, is the biological and evolutionary underpinning of the PCF idea: that the fitness advantages of sharing immunological information outweigh the fitness disadvantages (the primary one of which is, essentially, wholly selfish: that one individual is helping unrelated individuals to survive infectious disease). The PCF, for example through the ViDe instantiation, is an early warning system. Infected individuals unwittingly share information with distant others about a new airborne pathogen, in the form of viral and immune system fragments (possibly even airborne antibodies), allowing on inhalation some form of immunological response that buffers those individuals to the effects of subsequent infection.

A fundamental concept that emerges from the PCF hypothesis is that individuals are far more connected in terms of their health than is currently thought. One might call this "community immunity". There are strong positive feedback mechanisms within PCF – the immunological strength of a community is far more than the sum of its parts.

Since the first Report (Newman 2020) was released the author has received quite a number of news stories from colleagues chiming with the PCF hypothesis: Japan's low levels of COVID-19 mortality, the intense debate around bioaerosols and live virus transport, the apparent intrinsic immunity in some individuals who have not been infected with SARS-CoV-2 etc. The author hopes that this second Report, testing the PCF hypothesis with hard data, provides a spur to the wider community to further examine and test the PCF idea, and to take the idea forward as a platform for innovation.

If the PCF hypothesis is correct, it will change the way we think about infectious disease, emphasising emergent and non-linear community protection, and thereby catalysing entirely new approaches to combat COVID-19 and other infectious diseases.

Appendix A: mathematical details

A.1 defining the NMAP

We define a set of $M+1$ age classes, indexed by $m = 0, \cdots, M$ corresponding to age τ in the range: $0 \leq \tau < \tau_1, \tau_1 \leq \tau < \tau_2, \cdots, \tau_{m-1} \leq \tau < \tau_m, \tau_m \leq \tau$.

For each State we define the number of individuals in class m by n_m, such that

$$\sum_{m=0}^{M} n_m = N \, ,$$

where N is the total number of individuals in that State.

Likewise, for each State we define the number of individual COVID-19 deaths in class m by k_m, such that

$$\sum_{m=0}^{M} k_m = K \, ,$$

where K is the total number of COVID-19 deaths in that State.

The COVID-19 mortality rate for a given State within a particular age class is given by

$$\mu_m = {}^{k_m}/_{n_m} \, .$$

These values across age groups constitute what we refer to in the main text as the mortality age profile (MAP). We denote the sum of these values by S, namely

$$S = \sum_{m=0}^{M} \mu_m \ .$$

Finally, we define the normalised mortality age profile (NMAP) for a given State by the set of values

$$\mathrm{NMAP}_m = {}^{\mu_m}/_{S} \ ,$$

from which it is seen that the NMAP for each State is normalised to unity.

A.2 defining the immune decline and barrier functions

In the main text we refer to the exponential decay of the thymus, and define the immune decline function (IDF) and its inverse, the immune barrier function (IBF). Representing age by the symbol τ, these functions have the explicit mathematical expressions:

$$\mathrm{IDF} = \exp(-\alpha\tau) \ ,$$

$$\mathrm{IBF} = \frac{1}{\mathrm{IDF}} = \exp(\alpha\tau) \ .$$

For age τ measured in units of decades, the parameter α has the approximate value 0.44 /decade (Sottini *et al*. 2014, Palmer *et al*. 2018).

References

AMS 2020 *Glossary of Meterology* American Meterological Society (*glossary.ametsoc.org*)

Bar-On YM, Flamholz A, Phillips R, Milo R 2020 *SARS-CoV-2 (COVID-19) by the numbers* eLife **9** 357309

Davies NG *et al*. 2020 *Age-dependent effects in the transmission and control of COVID-19 epidemics* Nat. Med. doi:10.1038/s41591-020-0962-9

Donnelly R, Palmer S 2020 *Risk of COVID-19 hospitalisation rises exponentially with age, inversely proportional to T-cell production* (preprint)

Dowd JB *et al*. 2020 *Demographic science aids in understanding the spread and fatality rates of COVID-19* PNAS **117** 9696-9698

Everitt BS, Skrondal A 2010 *The Cambridge dictionary of statistics* (C.U.P., Cambridge)

Guerra-Silveira F, Abad-Franch F 2013 *Sex bias in infectious disease epidemiology: patterns and processes* PLOS ONE **8** e62390

Gui J, Mustachio LM, Su D-M, Craig RW 2012 *Thymus size and age-related thymic involution: early programming, sexual dimorphism, progenitors and stroma* Aging & Dis. **3** 280-290

Loomba RS *et al*. 2020 *Disparities in case frequency and mortality of coronavirus disease 2019 (COVID-19) among various states in the United States* (preprint) doi:10.1101/2020.07.28.20163931

Mollalo A, Vahedib B, Riveraa KM 2020 *GIS-based spatial modeling of COVID-19 incidence rate in the continental United States* Sci. Total Env. **728** 138884

Montecino-Rodriguez E, Berent-Maoz B, Dorshkind, K 2013 *Causes, consequences, and reversal of immune system aging* J. Clin. Invest. **123** 958–965

Newman TJ 2020 *Epidemic suppression via an emergent pre-conditioning field* Solaravus Tech. Rep. 003 doi: 10.5281/zenodo.3937508

Palmer S, Albergante L, Blackburn CC, Newman TJ 2018 *Thymic involution and rising disease incidence with age* PNAS **115** 1883-1888

Promislow DEL 2020 *A geroscience perspective on COVID-19 mortality* J. Ger. doi:10.1093/gerona/glaa094

Richardson S *et al*. 2020 *Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area* JAMA **323** 2052–2059

Rousseau M-A *et al*. 2020 *Understanding the thymus with applications to COVID-19 pathophysiology and susceptibility with potential therapeutics* (ResearchGate preprint)

Sottini A *et al.* 2014 *Simultaneous quantification of T-cell receptor excision circles (TRECs) and K-deleting recombination excision circles (KRECs) by real-time PCR* JoVE J. Vis. Exp. e52184

Texier G *et al*. 2016 *Outbreak definition by change point analysis: a tool for public health decision?* BMC Med. Inform. & Dec. Mak. doi:10.1186/s12911-016-0271-x

US DHHS 2012 *Principles of epidemiology in public health practice, 3rd ed.* (CDC, US DHHS)

Zhang CH, Schwartz GG 2020 *Spatial disparities in coronavirus incidence and mortality in the United States: an ecological analysis as of May 2020* J Rural Health **36** 433-445

Zhu N *et al*. 2020 *A novel coronavirus from patients with pneumonia in China*, 2019 New Eng. J. Med. **382** 727-733

Data Sources

www.cdc.gov
[CDC "Provisional COVID-19 death counts by sex, age and state", dataset up to and including July 4th 2020]
data.census.gov/cedsci
[US Census Bureau 2018 data for underlying demographics (age and sex) of each state]
www.worldometers.info/coronavirus
[Internationally sourced COVID-19 datasets (incidence and mortality), source for US25 epidemic start dates]
www.americashealthrankings.org
[United Health Foundation 2018 data on health across the United States]
www.bccdc.ca/health-info/diseases-conditions/covid-19/data
[COVID data for British Columbia: BC_Surveillance_Summary_July_16_2020.pdf]
files.ontario.ca/moh-covid-19-report-en-2020-07-20.pdf
[Ontario Agency for Health Protection and Promotion; January 15, 2020 to July 19, 2020]
www.quebec.ca/en/health/health-issues/a-z/2019-coronavirus
[COVID data for Quebec: situation-coronavirus-in-quebec/#c63035]
www12.statcan.gc.ca
[Demographic data for Canadian Provinces from the 2016 census]
www.weatheronline.co.uk
[Climate record data: net wind direction March-April 2020 for north-eastern US]

Acknowledgements

The author is grateful to Ruairi Donnelly and Sam Palmer for sharing their analysis on immune system dependence of COVID-19 hospitalisations prior to completion of their paper. The author is grateful to Janis Antonovics, Robert Austin, Myles Byrne, Karl Friston, James Le Fanu, John Golden, Josh LaBaer, David Liao, Stuart Lyndsey, Irwin McLean, Ron Milo, Denis Noble, James O'Dwyer, Rupert Sheldrake, Karol Sikora, and Bernhard Strauss for feedback on the PCF hypothesis. The author sincerely thanks Eva Lundell for assistance with data gathering and checking, and for a careful reading of the manuscript, Luca Albergante for advice on statistical analysis, and Rupert Sheldrake and Greg Vogel for support in bringing the PCF hypothesis to a wider audience.

Short biography

Thea Newman has thirty years' experience in scientific research and has held academic positions at the University of Virginia (2000-2002), Arizona State University (2002-2010) and the University of Dundee (2011-2017). Her training is originally in theoretical physics. Since 2000 she has pursued theoretical biophysics research across a range of biological areas: ecology, developmental biology, molecular biology and most recently cancer. Since 2018 she has directed SOLARAVUS, working as an independent educator and researcher.

Thea Newman is not an expert *per se* in immunology, epidemiology or atmospheric science. This technical report is offered to the public and the scientific community as both an analysis of spatio-temporal correlations in US COVID-19 data and as a statistical test of the PCF hypothesis. In sum, it is hoped to provide utility to all those tasked with better understanding and managing the COVID-19 pandemic and subsequent epidemics.

List of acronyms

We use the following acronyms as a short-hand throughout the article. In scientific papers it is conventional to use the unabbreviated form only once, thereafter only using the acronym. We do not slavishly follow that practice here as this paper is written for both the scientific community and the general public. Thus, we occasionally use unabbreviated forms if we feel it improves clarity (e.g. in figure captions, section headings and discursive text).

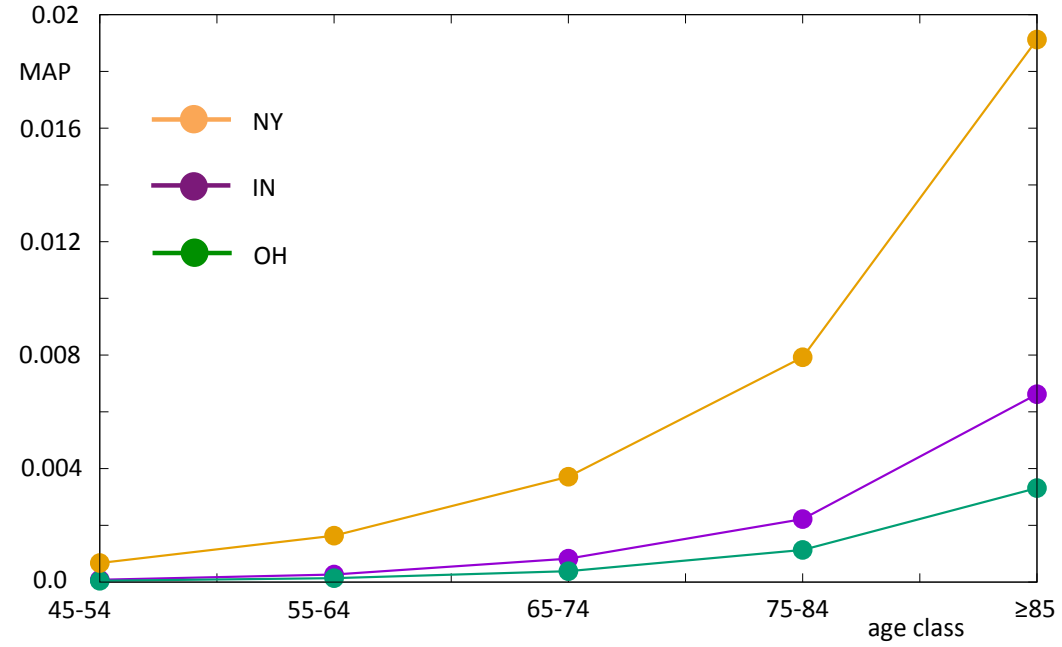
CDC	US Centers for Disease Control and Prevention
COVID-19	coronavirus disease 2019, caused by the SARS-CoV-2 virus
IBF	immune barrier function (the inverse of the IDF)
IDF	immune decline function (parameterised by the rate α =0.44 /decade)
MAP	mortality age profile
NI	NMAP index
NMAP	normalised mortality age profile
PCF	pre-conditioning field
PEC	pre-existing conditions
SIR	susceptible-infected-recovered (epidemiology framework)
UHF	United Health Foundation
US	United States
US25	the set of 25 worst-affected States within the US as of July 4th 2020
ViDe	viral detritus
WHO	World Health Organisation

List of figures and data tables

Figures:

Data tables:

Figures



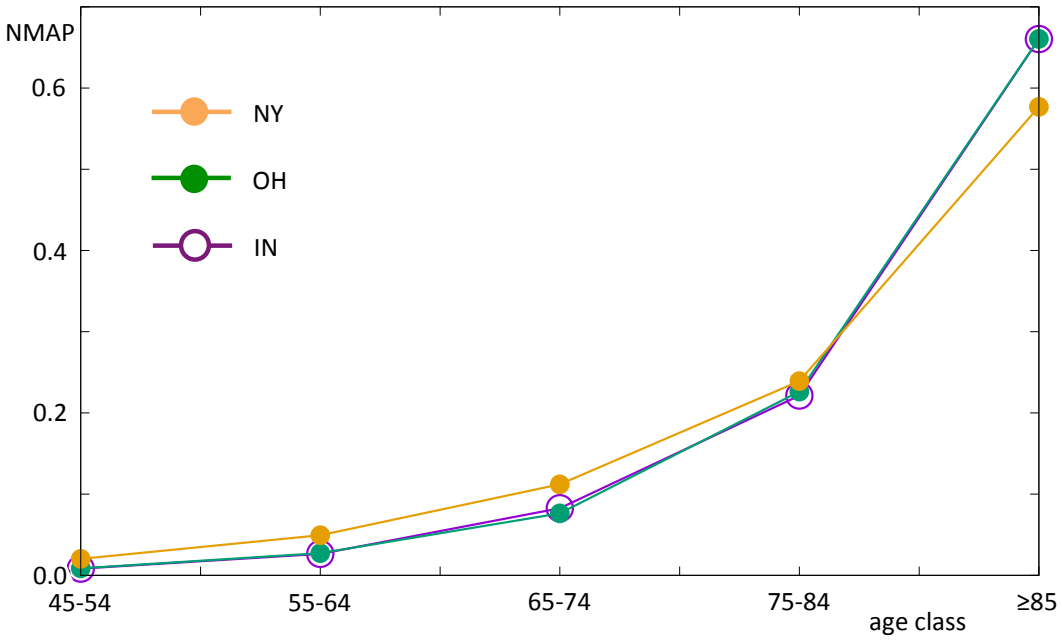
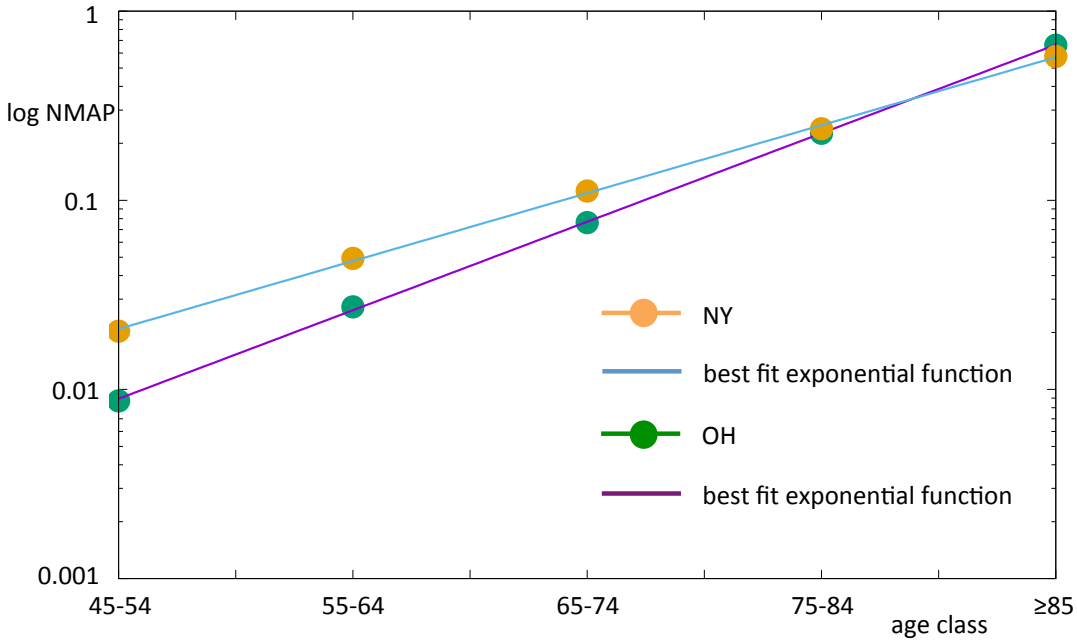


Figure 1a: Upper panel shows MAPs for the States of NY, IN and OH. The curve for NY is significantly higher due to the high mortality rates in that State. Lower panel shows NMAPs for the same three States. Note that the NMAPs for OH and IN are almost identical, meaning their NMAP indices will be very close (cf. Table 3). The NMAP for NY is flatter by comparison.



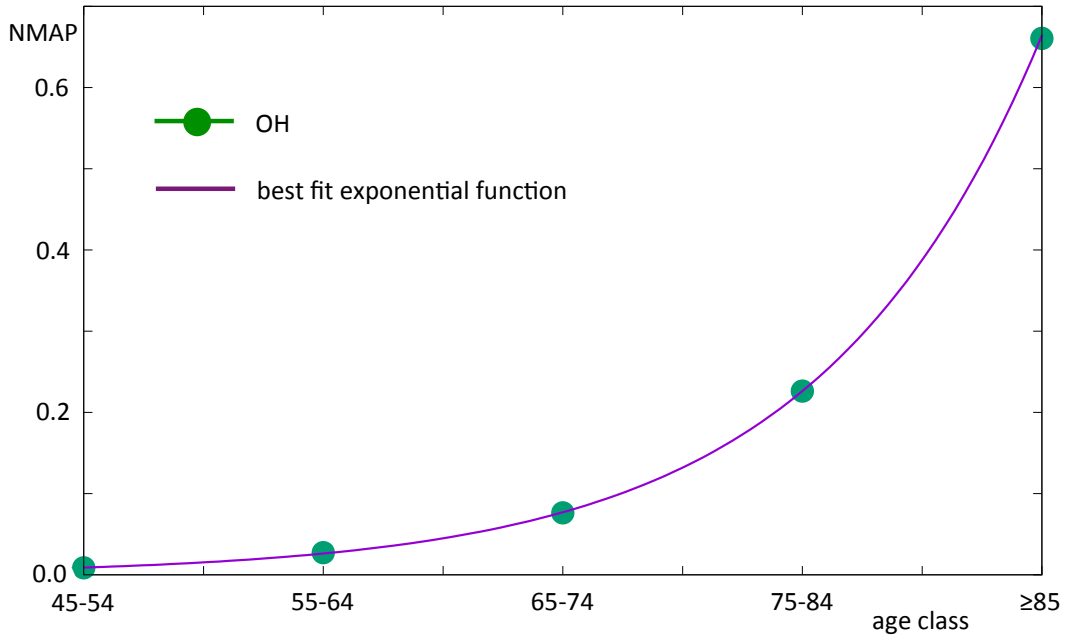
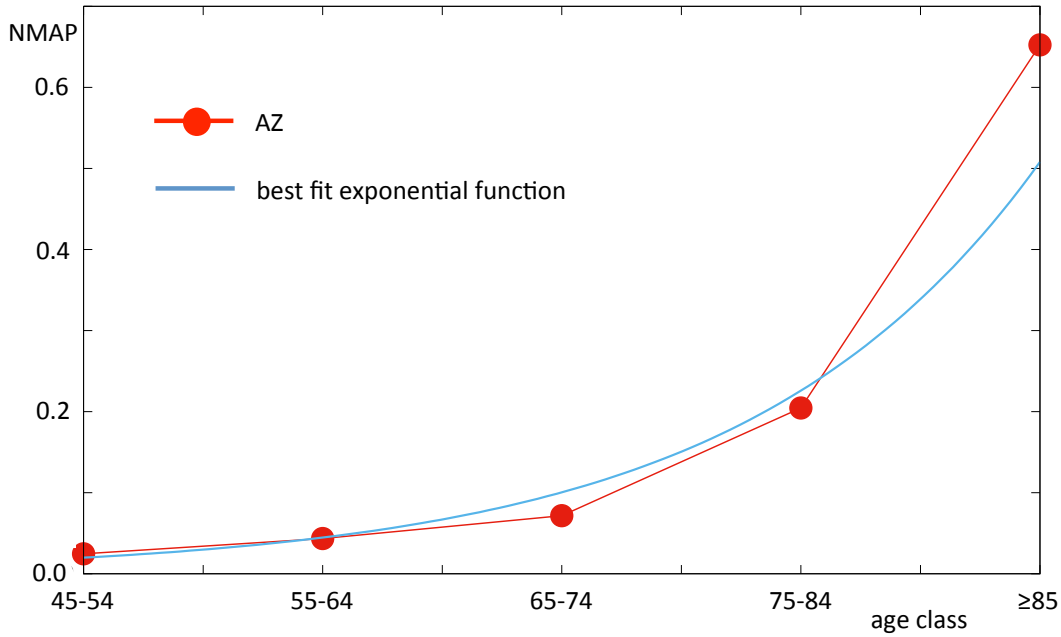


Figure 1b: Upper panel shows NMAPs for the States of NY and OH on a log-linear scale. The straightness of the curves indicates that the NMAPs are well approximated to exponentials. The additional lines are best fits to exponential curves, the slopes of which provide the NMAP indices (see Table 3). Lower panel shows the OH NMAP along with its exponential fit on a linear scale to stress the excellence of fit to an exponential function.



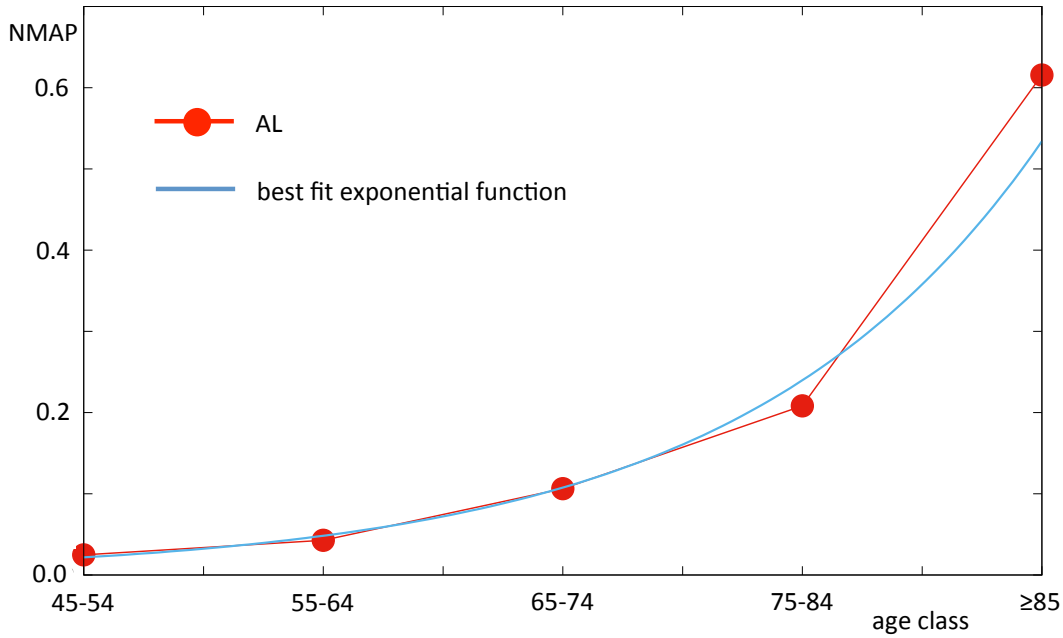


Figure 1c: NMAPs for the States of AZ (upper panel) and AL (lower panel) along with best fits to an exponential function. The poorness of these fits is reflected in the large errors in the value of the NMAP indices, greater than the cut-off of 5% (cf. Table 3). The NMAPs for these two States are not well-fitted to a pure exponential function and cannot be described by a single parameter.

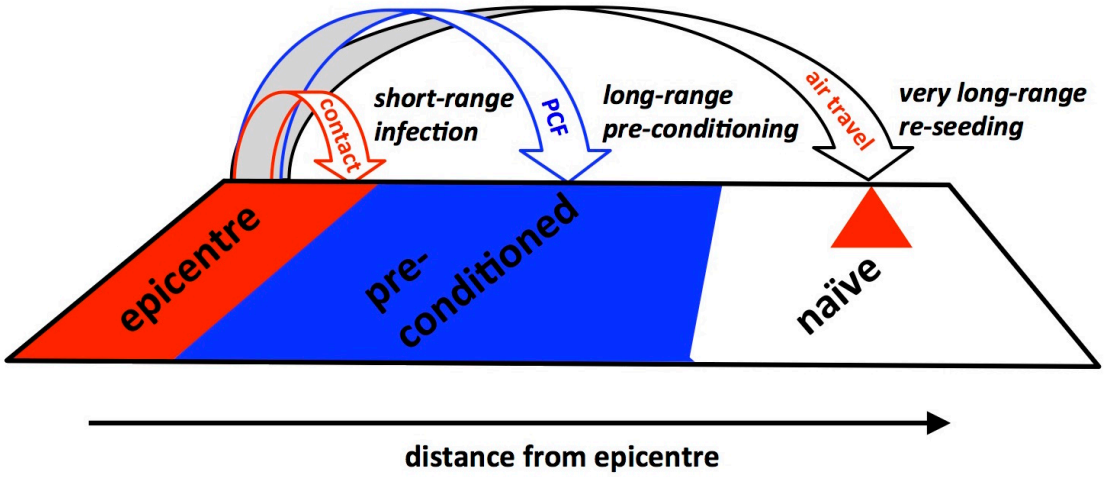
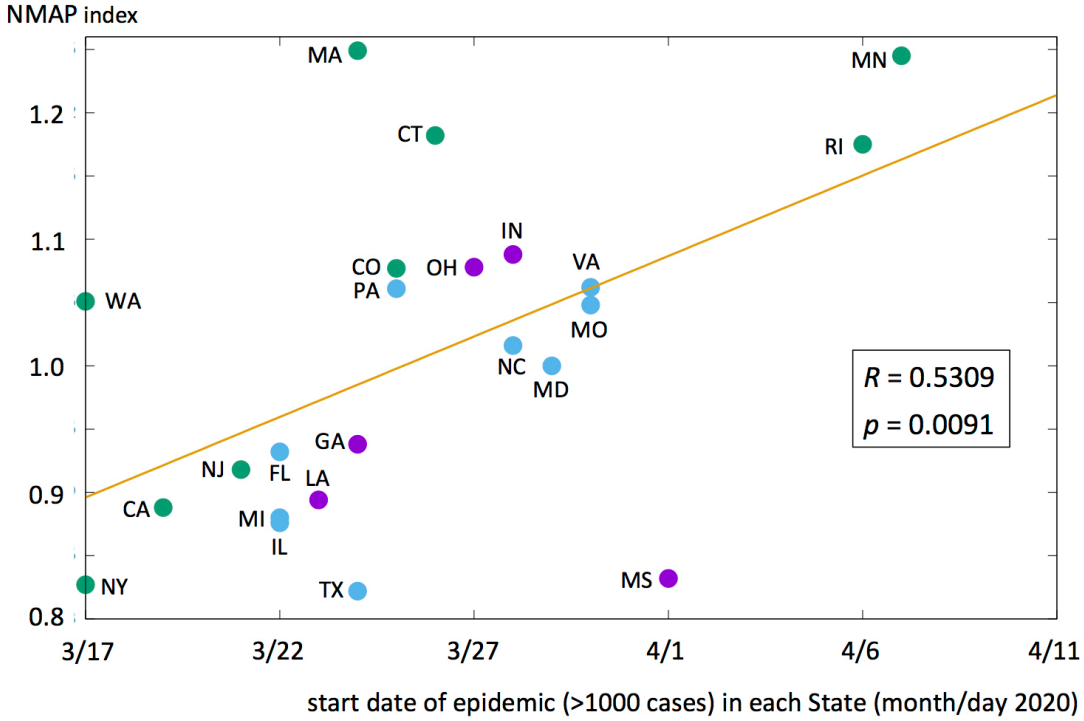


Figure 2: Schematic of the spatial effects of the PCF. In regions directly adjacent to the epicentre, local contact infection will dominate. Within the PCF hypothesis, more distant regions will experience immunological pre-conditioning prior to subsequent infection and experience a milder form of the epidemic. Very distant regions are likely to be seeded with new epidemics due to long-distance air travel prior to significant beneficial effects of the PCF. In interpreting this figure, it is helpful to think of the viral detritus (ViDe) instantiation of the PCF. *(This figure originally appeared as Figure 2 in Solaravus Technical Report 003, Newman 2020.)*



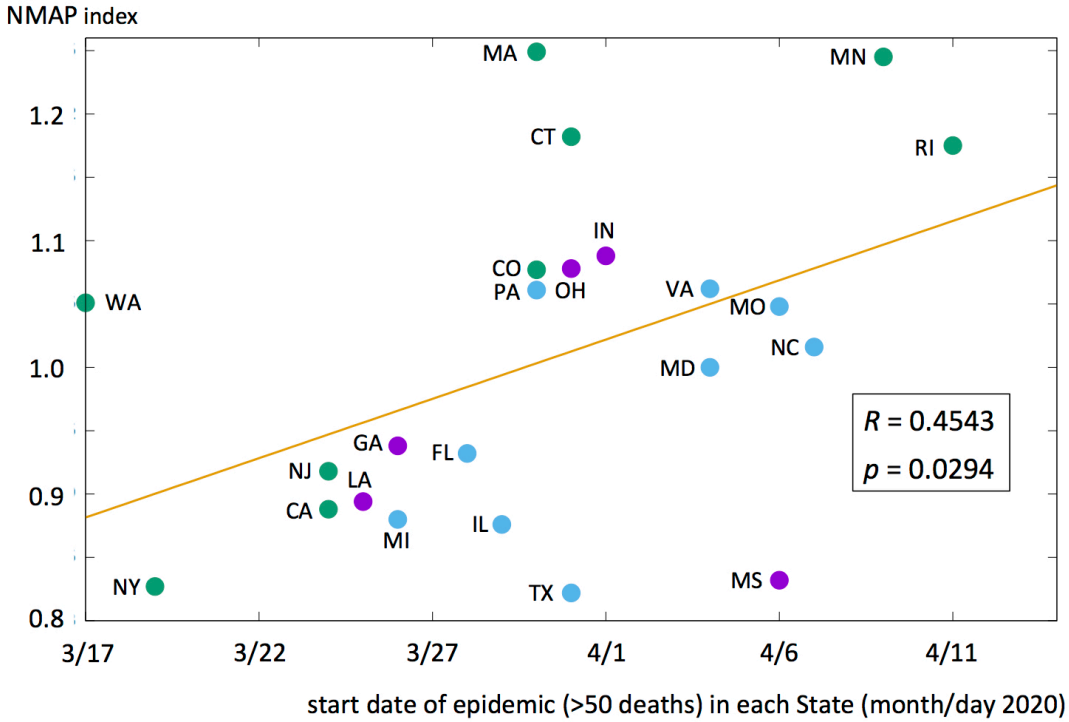
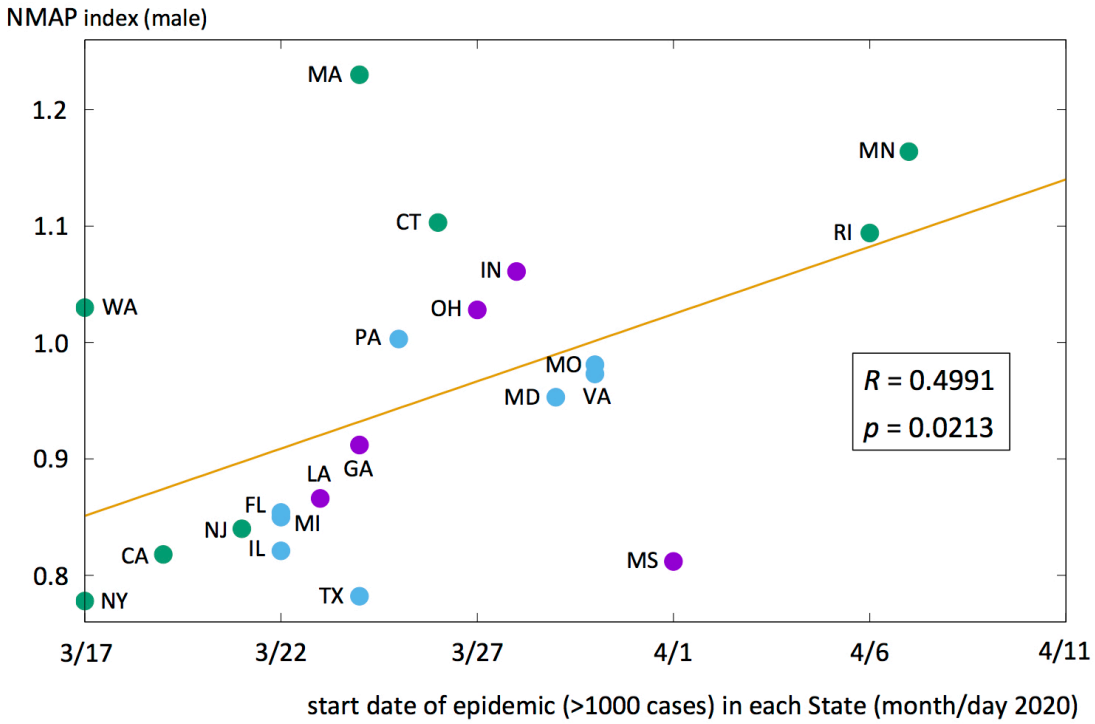


Figure 3a: NMAP indices plotted against start dates of epidemic according to definitions 1 (upper panel) and 2 (lower panel) for US25 (excluding AL and AZ whose indices have errors in excess of 5%, cf. Table 3). Colours green, blue, purple correspond to health indices in the categories H, M, L respectively (cf. Table 1). Best fit straight lines are shown in light brown. The correlation coefficient and *p*-value are boxed.



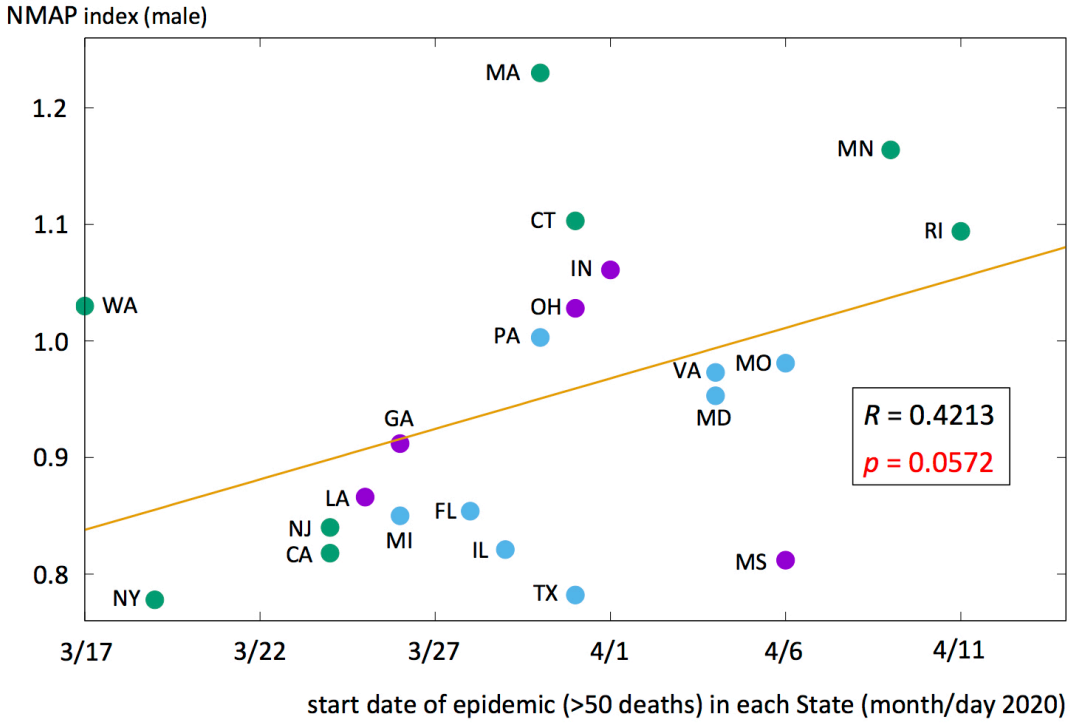
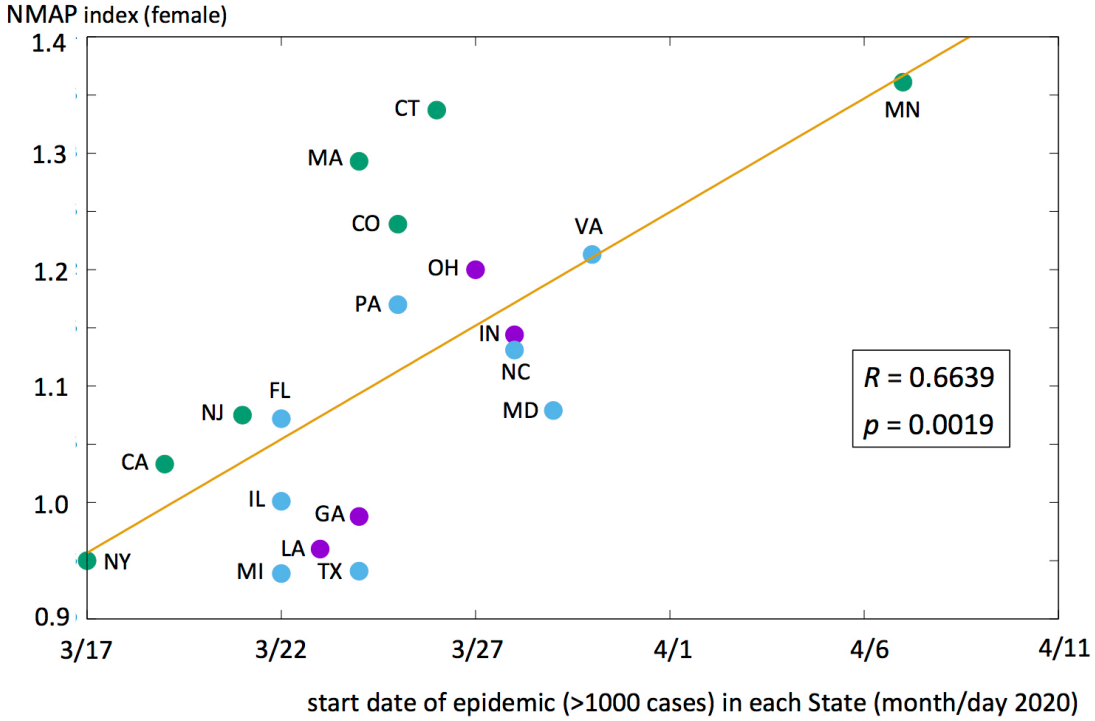


Figure 3b: NMAP indices for male sub-populations, plotted against start dates of epidemic according to definitions 1 (upper panel) and 2 (lower panel) for US25 (excluding AL, AZ, CO and NC whose indices have errors in excess of 5%, cf. Table 3). Colours green, blue, purple correspond to health indices in the categories H, M, L respectively (cf. Table 1). Best fit straight lines are shown in light brown. The correlation coefficient and *p*-value are boxed.



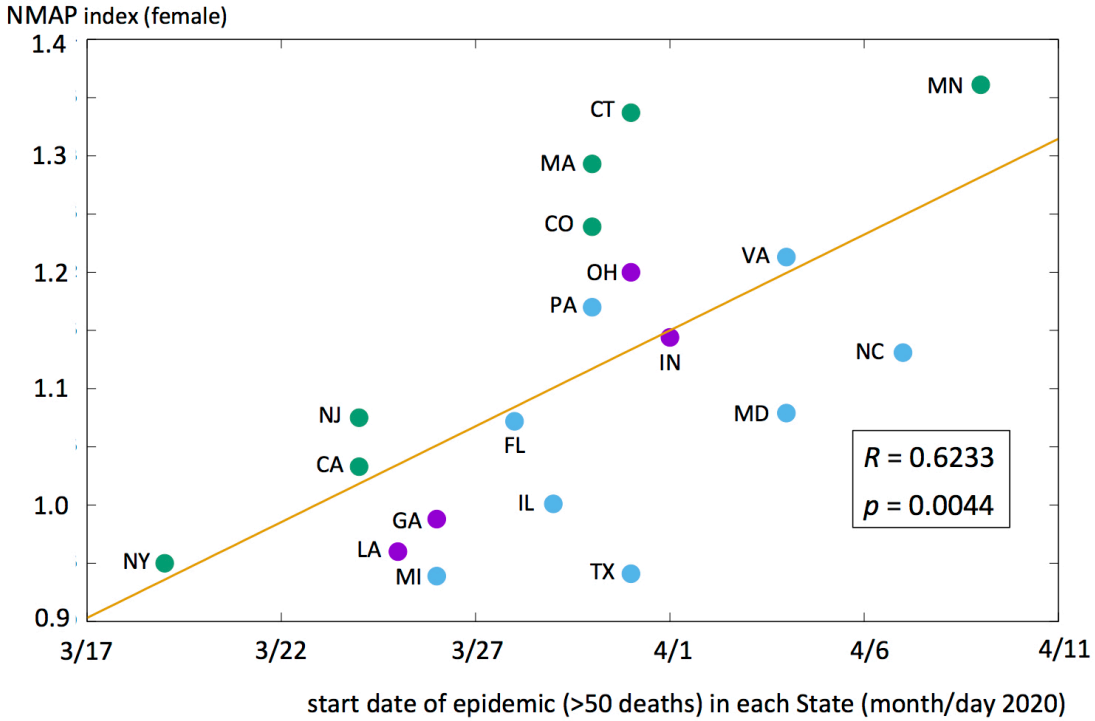


Figure 3c: NMAP indices for female sub-populations, plotted against start dates of epidemic according to definitions 1 (upper panel) and 2 (lower panel) for US25 (excluding AL, AZ, MO, MS, RI and WA whose indices have errors in excess of 5%, cf. Table 3). Colours green, blue, purple correspond to health indices in the categories H, M, L respectively (cf. Table 1). Best fit straight lines are shown in light brown. The correlation coefficient and *p*-value are boxed.

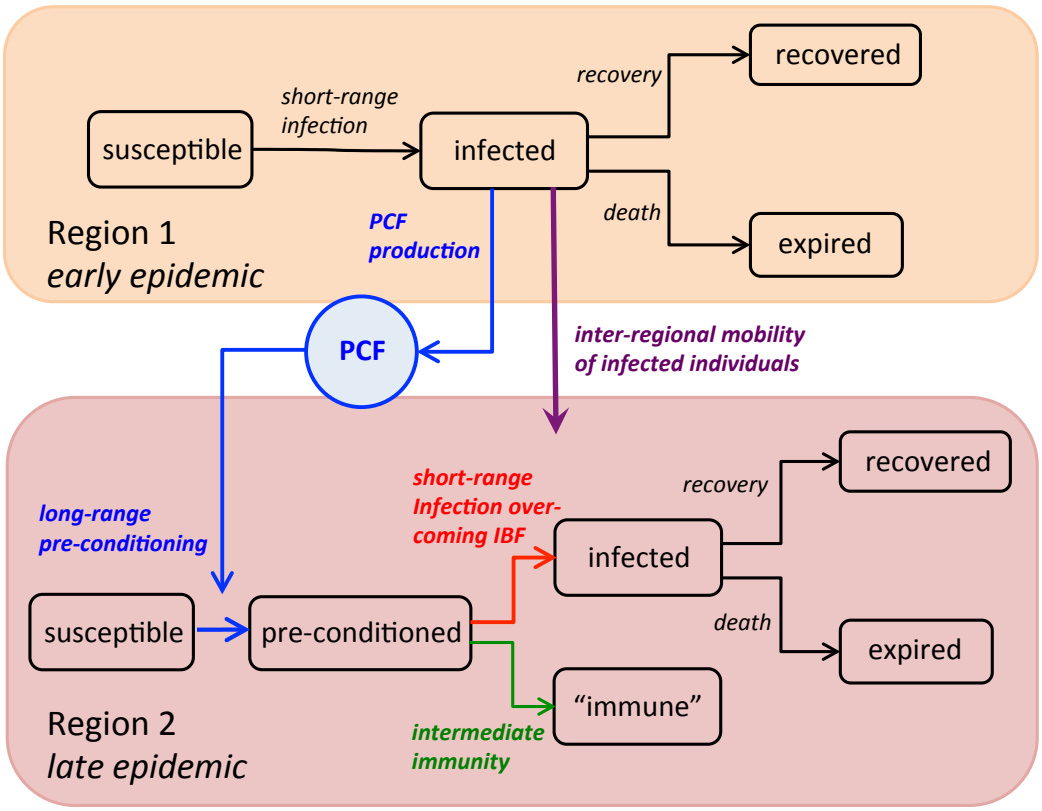


Figure 4: Schematic of the inter-regional dynamics of the PCF within the SIR framework of epidemiology. (*Intra*-regional dynamics are discussed at length in Newman 2020). The blue arrows represent the PCF dynamics, spreading to Region 2 from Region 1, whilst the purple arrow represents mobility of infected individuals, spreading the infection from Region 1 to Region 2. The red arrow represents infection of pre-conditioned individuals in Region 2. This creates an additional step in the infection dynamics in Region 2 that was absent in Region 1. This new step involves an additional immunological barrier for the virus to overcome, described in the main text by the immune barrier function (IBF). It is hypothesised to be responsible for the sharp differences in NMAP indices between "earlier States" and "later States": those individuals (typically the elderly) with weaker immune systems will experience less protection from pre-conditioning and will be disproportionately infected (relative to Region 1), thus steeping the NMAP and increasing the NMAP index. For those individuals in which the infection fails to overcome the IBF, they will have effectively achieved a form of immunity, which we term "intermediate", as it is unlikely to be permanent (cf. discussion of the additional process of reversion in Newman 2020).

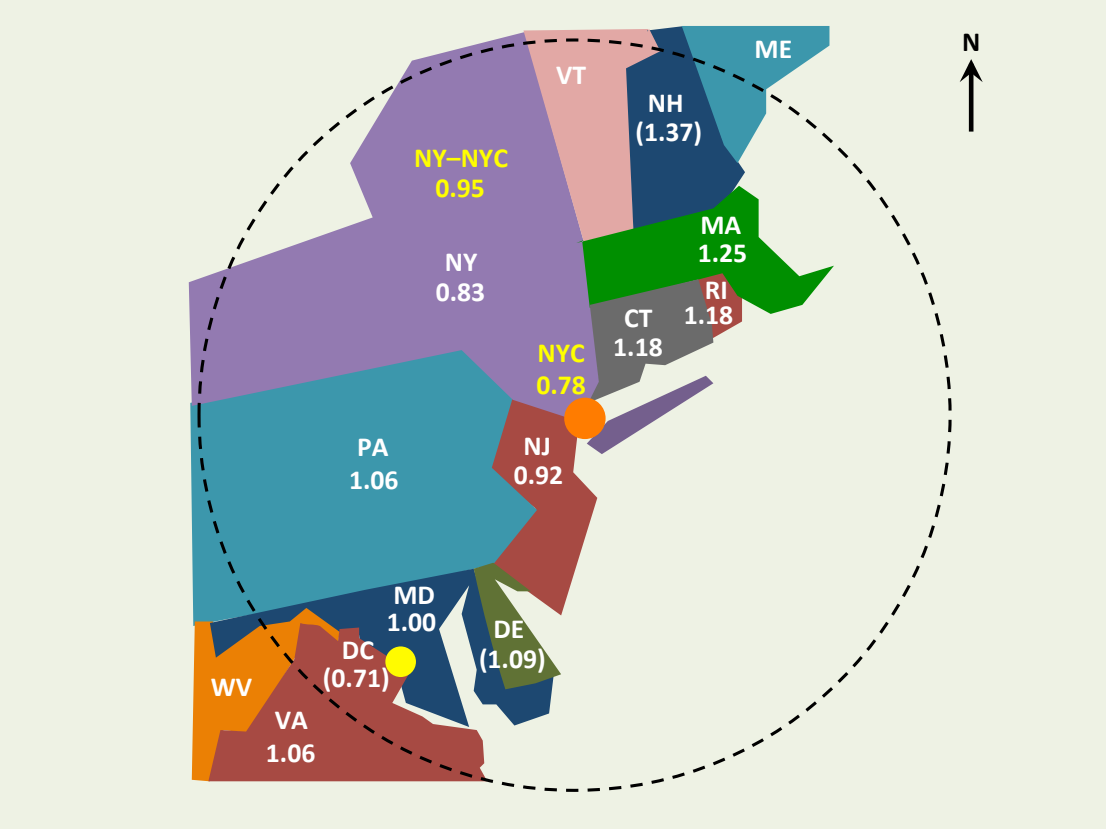


Figure 5: Schematic of the north-eastern US States (and the District of Columbia) with New York City (NYC) at the centre. For each State the NI is given. Note, the NI for NY State as a whole (including NYC) is 0.83 (cf. Table 3) whilst (shown in yellow font) the NI for NYC is 0.78 and the NI for NY State excluding NYC is 0.95 (cf. Table 5). Those NIs in parentheses are given for completeness, but are less statistically robust (cf. Tables 3 and 5). NIs cannot be robustly calculated for West Virginia, Vermont and Maine due to the relatively low number of deaths in those States. Note the strong spatial correlations of NIs between those States east of NYC (values around 1.2) and separately between those States south-west and west of NYC (values around 1.0).

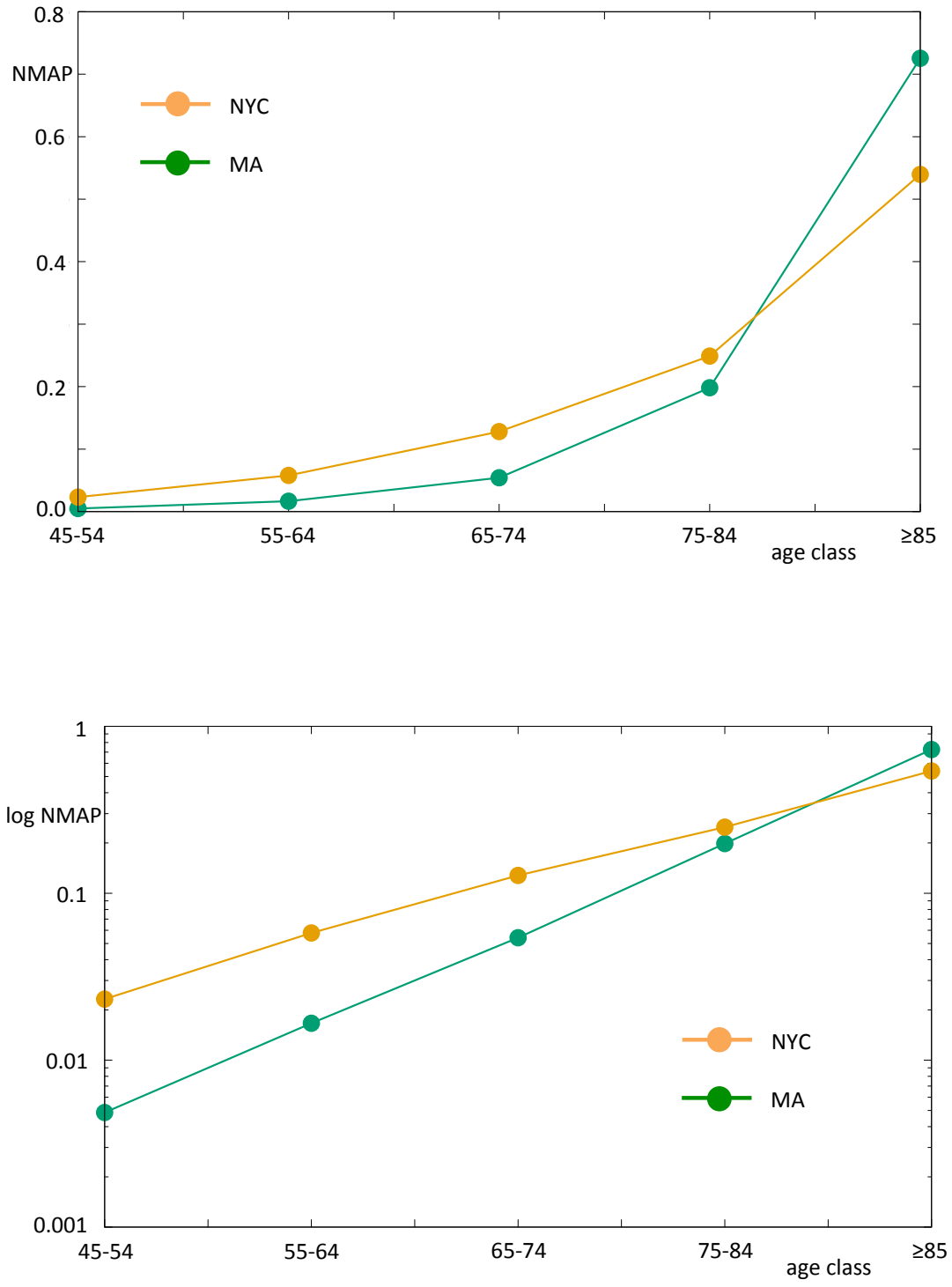


Figure 6a: NMAPs for MA and NYC; linear scale (upper panel) and log-linear scale (lower panel)

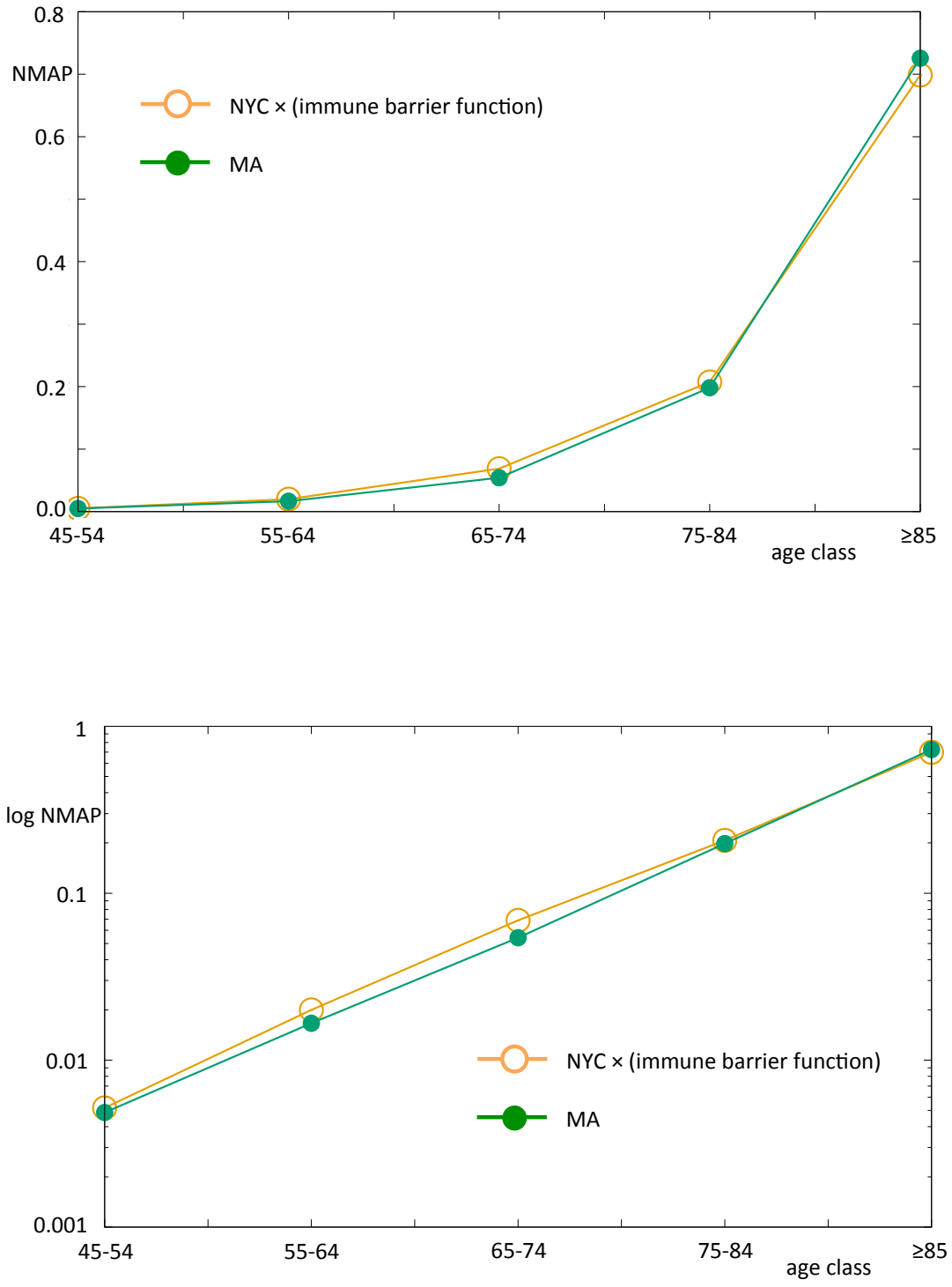


Figure 6b: NMAPs comparing MA to the NYC data multiplied by the immune barrier function (IBF): $\exp(\alpha \times \text{age class})$ where $\alpha = 0.44/\text{decade}$ (this product is subsequently normalised to ensure unit area beneath the curve as with all NMAPs); linear scale (upper panel) and log-linear scale (lower panel). Note the high degree of overlap, requiring open circles for the NYC data to allow the MA data to be visible.

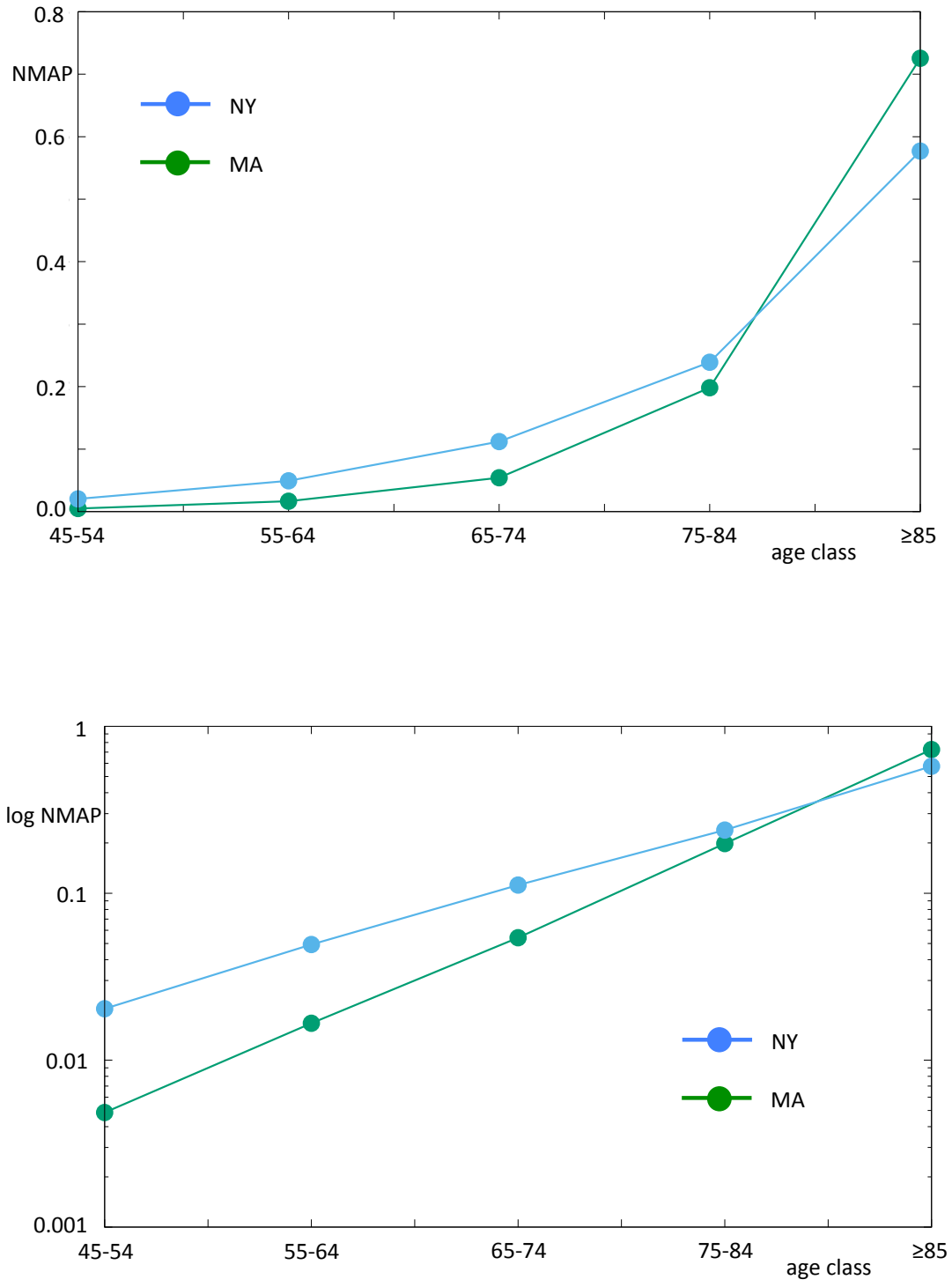


Figure 6c: NMAPs for MA and NY State; linear scale (upper panel) and log-linear scale (lower panel)

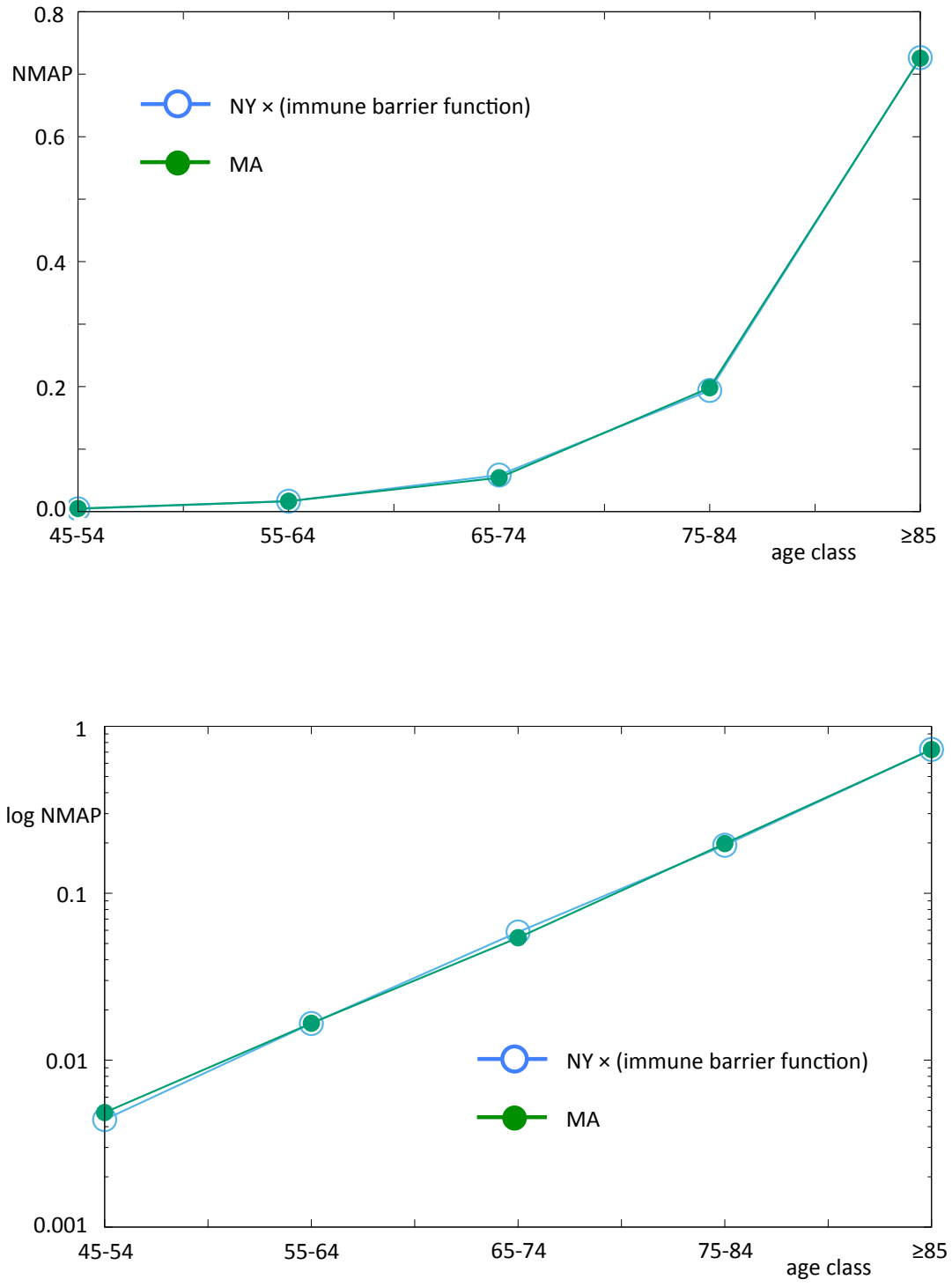


Figure 6d: NMAPs comparing MA to the NY State data multiplied by the immune barrier function (IBF): $\exp(\alpha \times \text{age class})$ where $\alpha = 0.44/\text{decade}$ (this product is subsequently normalised to ensure unit area beneath the curve as with all NMAPs); linear scale (upper panel) and log-linear scale (lower panel). Note the remarkable degree of overlap, requiring open circles for the NY State data to allow the MA data to be visible.

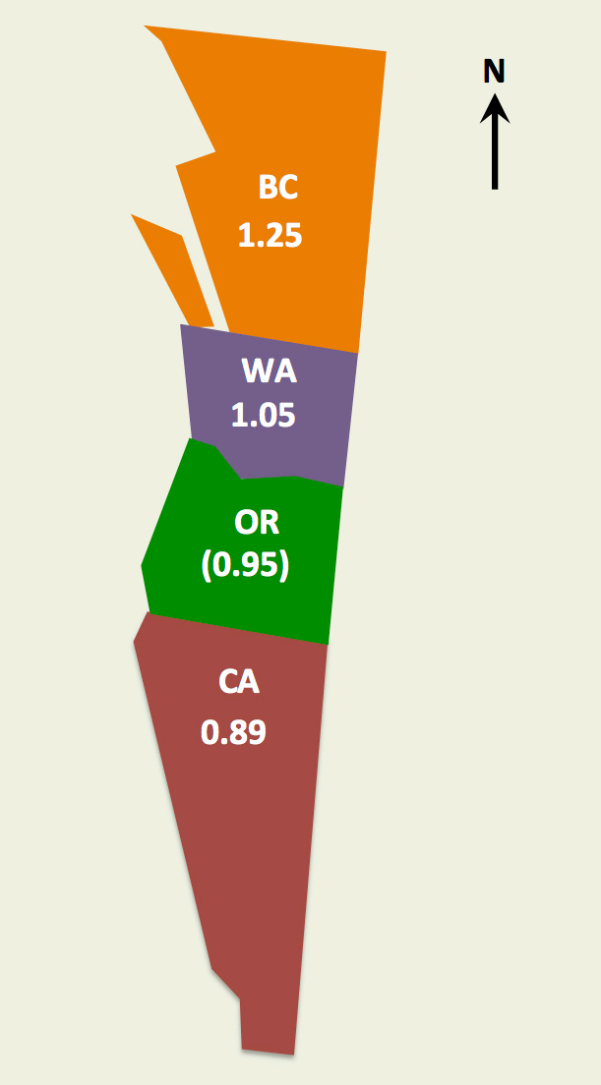


Figure 7: Schematic of the west coast of the US and southern Canada. For each State/Province the NMAP index is given. The NMAP index for Oregon is in parentheses as it is less statistically robust due to the relatively low number of deaths in that State (cf. Table 5). Note the trend of increasing NMAP indices from south to north.

Data tables

State	Abb'n	Epidemic start (1000 cases)	Epidemic start (50 deaths)	No. of deaths (as of 07/04/20)	Health index (L, M, H)
Alabama	AL	04/01	04/06	935	L (–0.838)
Arizona	AZ	03/30	04/04	1,259	M (–0.105)
California	CA	03/19	03/24	5,172	H (0.431)
Colorado	CO	03/25	03/30	1,520	H (0.588)
Connecticut	CT	03/26	03/31	3,473	H (0.799)
Florida	FL	03/22	03/28	2,969	M (–0.087)
Georgia	GA	03/24	03/26	2,054	L (–0.394)
Illinois	IL	03/22	03/29	5,973	M (0.060)
Indiana	IN	03/28	04/01	2,375	L (–0.432)
Louisiana	LA	03/23	03/25	2,613	L (–1.021)
Maryland	MD	03/29	04/04	3,220	M (0.306)
Massachusetts	MA	03/24	03/30	7,217	H (0.866)
Michigan	MI	03/22	03/26	5,217	M (–0.194)
Minnesota	MN	04/07	04/09	1,309	H (0.665)
Mississippi	MS	04/01	04/06	906	L (–1.010)
Missouri	MO	03/30	04/06	855	M (–0.345)
New Jersey	NJ	03/21	03/24	12,843	H (0.460)
New York	NY	03/17	03/19	29,530	H (0.476)
North Carolina	NC	03/28	04/07	859	M (–0.191)
Ohio	OH	03/27	03/31	2,256	L (–0.424)
Pennsylvania	PA	03/25	03/30	6,600	M (–0.014)
Rhode Island	RI	04/06	04/11	822	H (0.382)
Texas	TX	03/24	03/31	2,221	M (–0.286)
Virginia	VA	03/30	04/04	1,723	M (0.305)
Washington	WA	03/17	03/17	1,107	H (0.584)

Table 1: Data for the 25 worst-affected States (in terms of total mortality from COVID-19) as of 04/07/20: starting dates (MM/DD/2020) of epidemic for each State, measured by 1000 cases (column 3) or by 50 deaths (column 4), total number of COVID-19 deaths as of 04/07/20, and health categories of low (< –0.35), medium (between –0.35 and +0.35) and high (> +0.35), with the associated indices from the United Health Foundation dataset (cf. subsection 2.5 for details).

Alabama						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	1,395,626	302,901	309,774	227,090	106,746	27,134
pop'n (F)	1,384,237	323,272	342,398	266,306	144,279	58,108
pop'n (M/F)	2,779,863	626,173	652,172	493,396	251,025	85,242
deaths (M)	13	43	69	136	133	106
deaths (F)	16	27	57	101	103	131
deaths (M/F)	29	70	126	237	236	237
NMAP (M)	1.52×10^{-3}	2.32×10^{-2}	3.64×10^{-2}	9.78×10^{-2}	0.203	0.638
NMAP (F)	3.20×10^{-3}	2.31×10^{-2}	4.61×10^{-2}	0.105	0.198	0.625
NMAP (M/F)	2.31×10^{-3}	2.48×10^{-2}	4.28×10^{-2}	0.106	0.208	0.616

Arizona						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	2,149,919	422,450	413,360	340,427	186,987	51,986
pop'n (F)	2,043,369	430,012	453,433	387,529	211,421	80,753
pop'n (M/F)	4,193,288	852,462	866,793	727,956	398,408	132,739
deaths (M)	41	60	101	121	200	175
deaths (F)	20	30	61	103	150	197
deaths (M/F)	61	90	162	224	350	372
NMAP (M)	3.67×10^{-3}	2.73×10^{-2}	4.70×10^{-2}	6.84×10^{-2}	0.206	0.648
NMAP (F)	2.70×10^{-3}	1.92×10^{-2}	3.71×10^{-2}	7.32×10^{-2}	0.196	0.672
NMAP (M/F)	3.39×10^{-3}	2.46×10^{-2}	4.35×10^{-2}	7.16×10^{-2}	0.205	0.652

California						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	12,302,593	2,522,257	2,327,186	1,530,227	719,421	271,869
pop'n (F)	11,737,663	2,545,769	2,454,240	1,756,234	931,875	457,711
pop'n (M/F)	24,040,256	5,068,026	4,781,426	3,286,461	1,651,296	729,580
deaths (M)	136	237	467	699	737	674
deaths (F)	43	86	190	372	551	980
deaths (M/F)	179	323	657	1,071	1,288	1,654
NMAP (M)	2.59×10^{-3}	2.20×10^{-2}	4.70×10^{-2}	0.107	0.240	0.581
NMAP (F)	1.20×10^{-3}	1.10×10^{-2}	2.53×10^{-2}	6.92×10^{-2}	0.193	0.700
NMAP (M/F)	2.08×10^{-3}	1.78×10^{-2}	3.84×10^{-2}	9.10×10^{-2}	0.218	0.633

Table 2: demographic and COVID-19 mortality data by gender and age class for each State in US25, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

Colorado						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	1,792,237	356,963	346,233	239,923	97,665	31,740
pop'n (F)	1,674,084	353,255	364,937	262,495	120,366	55,666
pop'n (M/F)	3,466,321	710,218	711,170	502,418	218,031	87,406
deaths (M)	24	50	104	183	240	249
deaths (F)	10	16	36	111	179	318
deaths (M/F)	34	66	140	294	419	567
NMAP (M)	1.16×10^{-3}	1.22×10^{-2}	2.61×10^{-2}	6.62×10^{-2}	0.213	0.681
NMAP (F)	7.69×10^{-4}	5.83×10^{-3}	1.27×10^{-2}	5.44×10^{-2}	0.191	0.735
NMAP (M/F)	1.06×10^{-3}	1.00×10^{-2}	2.12×10^{-2}	6.30×10^{-2}	0.207	0.698

Connecticut						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	987,301	238,346	247,014	160,264	78,283	29,325
pop'n (F)	964,432	256,296	266,129	182,796	104,108	58,371
pop'n (M/F)	1,951,733	494,642	513,143	343,060	182,391	87,696
deaths (M)	25	55	181	363	460	593
deaths (F)	13	19	95	226	451	992
deaths (M/F)	38	74	276	589	911	1,585
NMAP (M)	8.63×10^{-4}	7.86×10^{-3}	2.50×10^{-2}	7.72×10^{-2}	0.200	0.689
NMAP (F)	5.86×10^{-4}	3.22×10^{-3}	1.55×10^{-2}	5.37×10^{-2}	0.188	0.739
NMAP (M/F)	7.64×10^{-4}	5.87×10^{-3}	2.11×10^{-2}	6.73×10^{-2}	0.196	0.709

Florida						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	5,748,075	1,342,853	1,348,193	1,109,704	638,569	217,282
pop'n (F)	5,594,344	1,399,115	1,507,961	1,294,530	754,981	343,718
pop'n (M/F)	11,342,419	2,741,968	2,856,154	2,404,234	1,393,550	561,000
deaths (M)	50	85	192	379	502	419
deaths (F)	22	33	85	248	362	592
deaths (M/F)	72	118	277	627	864	1,011
NMAP (M)	2.66×10^{-3}	1.94×10^{-2}	4.35×10^{-2}	0.104	0.240	0.590
NMAP (F)	1.59×10^{-3}	9.52×10^{-3}	2.28×10^{-2}	7.73×10^{-2}	0.194	0.695
NMAP (M/F)	2.24×10^{-3}	1.52×10^{-2}	3.43×10^{-2}	9.22×10^{-2}	0.219	0.637

Table 2 (cont): demographic and COVID-19 mortality data by gender and age class for each State in US25, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

Georgia						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	3,195,777	681,648	609,190	407,128	178,755	47,792
pop'n (F)	3,188,813	714,682	672,937	489,820	237,925	95,008
pop'n (M/F)	6,384,590	1,396,330	1,282,127	896,948	416,680	142,800
deaths (M)	31	72	166	303	297	195
deaths (F)	27	49	105	196	284	329
deaths (M/F)	58	121	271	499	581	524
NMAP (M)	1.41×10^{-3}	1.54×10^{-2}	3.96×10^{-2}	0.108	0.242	0.594
NMAP (F)	1.60×10^{-3}	1.30×10^{-2}	2.95×10^{-2}	7.56×10^{-2}	0.226	0.655
NMAP (M/F)	1.53×10^{-3}	1.46×10^{-2}	3.57×10^{-2}	9.39×10^{-2}	0.235	0.619

Illinois						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	3,778,259	807,510	810,557	533,438	249,685	86,613
pop'n (F)	3,666,352	828,544	859,310	609,891	337,966	172,955
pop'n (M/F)	7,444,611	1,636,054	1,669,867	1,143,329	587,651	259,568
deaths (M)	151	253	509	832	843	709
deaths (F)	59	96	249	463	656	1,153
deaths (M/F)	210	349	758	1,295	1,499	1,862
NMAP (M)	2.83×10^{-3}	2.22×10^{-2}	4.45×10^{-2}	0.111	0.239	0.580
NMAP (F)	1.64×10^{-3}	1.18×10^{-2}	2.96×10^{-2}	7.76×10^{-2}	0.198	0.681
NMAP (M/F)	2.44×10^{-3}	1.85×10^{-2}	3.93×10^{-2}	9.80×10^{-2}	0.221	0.621

Indiana						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	1,991,702	415,745	425,989	287,314	130,823	44,400
pop'n (F)	1,933,002	426,005	448,289	326,385	180,047	82,177
pop'n (M/F)	3,924,704	841,750	874,278	613,699	310,870	126,577
deaths (M)	21	43	145	314	350	330
deaths (F)	16	27	87	194	340	508
deaths (M/F)	37	70	232	508	690	838
NMAP (M)	9.05×10^{-4}	8.87×10^{-3}	2.92×10^{-2}	9.38×10^{-2}	0.230	0.638
NMAP (F)	9.27×10^{-4}	7.10×10^{-3}	2.17×10^{-2}	6.66×10^{-2}	0.211	0.692
NMAP (M/F)	9.40×10^{-4}	8.29×10^{-3}	2.65×10^{-2}	8.26×10^{-2}	0.221	0.660

Table 2 (cont): demographic and COVID-19 mortality data by gender and age class for each State in US25, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

Louisiana						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	1,396,417	276,273	288,341	202,164	87,647	25,688
pop'n (F)	1,372,797	291,040	314,500	232,733	124,222	48,156
pop'n (M/F)	2,769,214	567,313	602,841	434,897	211,869	73,844
deaths (M)	47	96	212	356	413	268
deaths (F)	43	49	151	223	348	407
deaths (M/F)	90	145	363	579	761	675
NMAP (M)	1.87×10^{-3}	1.93×10^{-2}	4.08×10^{-2}	9.77×10^{-2}	0.261	0.579
NMAP (F)	2.43×10^{-3}	1.31×10^{-2}	3.72×10^{-2}	7.43×10^{-2}	0.217	0.656
NMAP (M/F)	2.17×10^{-3}	1.71×10^{-2}	4.03×10^{-2}	8.90×10^{-2}	0.240	0.611

Maryland						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	1,746,574	396,969	385,328	247,743	115,401	38,335
pop'n (F)	1,730,410	425,771	426,625	298,699	160,683	70,180
pop'n (M/F)	3,476,984	822,740	811,953	546,442	276,084	108,515
deaths (M)	74	91	251	335	483	407
deaths (F)	25	52	137	283	421	661
deaths (M/F)	99	143	388	618	904	1,068
NMAP (M)	2.48×10^{-3}	1.34×10^{-2}	3.81×10^{-2}	7.92×10^{-2}	0.245	0.622
NMAP (F)	1.07×10^{-3}	9.08×10^{-3}	2.39×10^{-2}	7.05×10^{-2}	0.195	0.701
NMAP (M/F)	1.91×10^{-3}	1.16×10^{-2}	3.20×10^{-2}	7.58×10^{-2}	0.219	0.659

Massachusetts						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	1,956,847	448,551	452,699	303,285	139,355	50,150
pop'n (F)	1,945,046	473,000	488,465	353,261	187,911	103,579
pop'n (M/F)	3,901,893	921,551	941,164	656,546	327,266	153,729
deaths (M)	40	80	307	662	1,056	1,256
deaths (F)	20	58	176	434	945	2,183
deaths (M/F)	60	138	483	1,096	2,001	3,439
NMAP (M)	5.73×10^{-4}	5.00×10^{-3}	1.90×10^{-2}	6.12×10^{-2}	0.212	0.702
NMAP (F)	3.70×10^{-4}	4.41×10^{-3}	1.29×10^{-2}	4.42×10^{-2}	0.181	0.757
NMAP (M/F)	4.99×10^{-4}	4.86×10^{-3}	1.66×10^{-2}	5.41×10^{-2}	0.198	0.726

Table 2 (cont): demographic and COVID-19 mortality data by gender and age class for each State in US25, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

Michigan						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	2,838,818	638,026	680,769	476,813	222,455	69,935
pop'n (F)	2,744,166	652,733	720,950	532,373	284,519	134,358
pop'n (M/F)	5,582,984	1,290,759	1,401,719	1,009,186	506,974	204,293
deaths (M)	80	187	385	722	769	582
deaths (F)	43	106	252	516	621	953
deaths (M/F)	123	293	637	1,238	1,390	1,535
NMAP (M)	1.99×10^{-3}	2.07×10^{-2}	3.99×10^{-2}	0.107	0.244	0.587
NMAP (F)	1.45×10^{-3}	1.51×10^{-2}	3.24×10^{-2}	9.00×10^{-2}	0.203	0.658
NMAP (M/F)	1.81×10^{-3}	1.86×10^{-2}	3.73×10^{-2}	0.101	0.225	0.617

Minnesota						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	1,670,328	347,261	373,337	245,473	114,663	42,918
pop'n (F)	1,606,130	342,419	383,070	265,013	148,112	72,455
pop'n (M/F)	3,276,458	689,680	756,407	510,486	262,775	115,373
deaths (M)	7	18	63	119	191	239
deaths (F)	–	–	32	86	174	368
deaths (M/F)	–	–	95	205	365	607
NMAP (M)	5.28×10^{-4}	6.52×10^{-3}	2.12×10^{-2}	6.10×10^{-2}	0.210	0.701
NMAP (F)	–	–	1.25×10^{-2}	4.87×10^{-2}	0.176	0.762
NMAP (M/F)	–	–	1.75×10^{-2}	5.60×10^{-2}	0.194	0.733

Mississippi						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	884,624	174,482	182,054	131,500	58,563	15,606
pop'n (F)	884,346	185,182	201,419	152,856	85,144	30,754
pop'n (M/F)	1,768,970	359,664	383,473	284,356	143,707	46,360
deaths (M)	17	33	69	118	125	72
deaths (F)	19	26	62	97	121	146
deaths (M/F)	36	59	131	215	246	218
NMAP (M)	2.33×10^{-3}	2.30×10^{-2}	4.60×10^{-2}	0.109	0.260	0.560
NMAP (F)	2.95×10^{-3}	1.93×10^{-2}	4.23×10^{-2}	8.73×10^{-2}	0.195	0.653
NMAP (M/F)	2.64×10^{-3}	2.13×10^{-2}	4.44×10^{-2}	9.82×10^{-2}	0.222	0.611

Table 2 (cont): demographic and COVID-19 mortality data by gender and age class for each State in US25, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

Missouri						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	1,775,158	369,432	401,755	276,416	137,039	43,365
pop'n (F)	1,729,982	383,922	431,129	320,207	176,607	81,440
pop'n (M/F)	3,505,140	753,354	832,884	596,623	313,646	124,805
deaths (M)	5	18	56	102	131	109
deaths (F)	–	–	34	68	103	213
deaths (M/F)	–	–	90	170	234	322
NMAP (M)	6.99×10^{-4}	1.21×10^{-2}	3.46×10^{-2}	9.16×10^{-2}	0.237	0.624
NMAP (F)	–	–	2.26×10^{-2}	6.09×10^{-2}	0.167	0.749
NMAP (M/F)	–	–	2.91×10^{-2}	7.66×10^{-2}	0.201	0.694

New Jersey						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	2,541,332	601,079	587,462	372,185	181,182	67,156
pop'n (F)	2,476,639	632,520	631,209	440,168	248,996	128,602
pop'n (M/F)	5,017,961	1,233,599	1,218,671	812,353	430,178	195,758
deaths (M)	257	502	1,112	1,556	1,778	1,645
deaths (F)	93	171	519	928	1,559	2,723
deaths (M/F)	350	673	1,631	2,484	3,337	4,368
NMAP (M)	2.45×10^{-3}	2.02×10^{-2}	4.58×10^{-2}	0.101	0.238	0.593
NMAP (F)	1.22×10^{-3}	8.81×10^{-3}	2.68×10^{-2}	6.87×10^{-2}	0.204	0.690
NMAP (M/F)	1.99×10^{-3}	1.56×10^{-2}	3.81×10^{-2}	8.72×10^{-2}	0.221	0.636

New York						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	5,624,636	1,242,488	1,243,082	832,067	402,125	146,915
pop'n (F)	5,551,953	1,314,850	1,353,135	979,528	565,716	285,714
pop'n (M/F)	11,176,589	2,557,338	2,596,217	1,811,595	967,841	432,629
deaths (M)	656	1,249	2,873	4,339	4,324	3,346
deaths (F)	240	474	1,369	2,386	3,347	4,927
deaths (M/F)	896	1,723	4,242	6,725	7,671	8,273
NMAP (M)	2.77×10^{-3}	2.38×10^{-2}	5.48×10^{-2}	0.124	0.255	0.540
NMAP (F)	1.60×10^{-3}	1.33×10^{-2}	3.75×10^{-2}	9.02×10^{-2}	0.219	0.638
NMAP (M/F)	2.42×10^{-3}	2.03×10^{-2}	4.93×10^{-2}	0.112	0.239	0.577

Table 2 (cont): demographic and COVID-19 mortality data by gender and age class for each State in US25, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

North Carolina						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	3,009,520	670,067	634,795	465,008	213,321	59,533
pop'n (F)	2,964,670	708,020	707,974	548,988	282,190	119,534
pop'n (M/F)	5,974,190	1,378,087	1,342,769	1,013,996	495,511	179,067
deaths (M)	19	29	47	101	137	116
deaths (F)	6	10	34	76	126	158
deaths (M/F)	25	39	81	177	263	274
NMAP (M)	2.15×10^{-3}	1.48×10^{-2}	2.53×10^{-2}	7.41×10^{-2}	0.219	0.665
NMAP (F)	1.03×10^{-3}	7.17×10^{-3}	2.44×10^{-2}	7.02×10^{-2}	0.227	0.671
NMAP (M/F)	1.80×10^{-3}	1.22×10^{-2}	2.59×10^{-2}	7.50×10^{-2}	0.228	0.657

Ohio						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	3,340,591	735,765	780,862	537,165	254,307	83,405
pop'n (F)	3,251,683	752,436	831,942	615,196	341,210	164,880
pop'n (M/F)	6,592,274	1,488,201	1,612,804	1,152,361	595,517	248,285
deaths (M)	23	46	152	291	344	338
deaths (F)	7	19	69	150	332	485
deaths (M/F)	30	65	221	441	676	823
NMAP (M)	1.11×10^{-3}	1.01×10^{-2}	3.13×10^{-2}	8.72×10^{-2}	0.218	0.652
NMAP (F)	5.04×10^{-4}	5.92×10^{-3}	1.94×10^{-2}	5.71×10^{-2}	0.228	0.689
NMAP (M/F)	9.07×10^{-4}	8.70×10^{-3}	2.73×10^{-2}	7.63×10^{-2}	0.226	0.661

Pennsylvania						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	3,552,410	819,094	879,291	614,944	298,548	107,333
pop'n (F)	3,451,209	840,447	932,240	696,215	409,022	206,307
pop'n (M/F)	7,003,619	1,659,541	1,811,531	1,311,159	707,570	313,640
deaths (M)	51	143	371	697	906	1,052
deaths (F)	20	61	215	497	770	1,817
deaths (M/F)	71	204	586	1,194	1,676	2,869
NMAP (M)	9.85×10^{-4}	1.20×10^{-2}	2.89×10^{-2}	7.77×10^{-2}	0.208	0.672
NMAP (F)	4.95×10^{-4}	6.20×10^{-3}	1.97×10^{-2}	6.09×10^{-2}	0.161	0.752
NMAP (M/F)	7.87×10^{-4}	9.54×10^{-3}	2.51×10^{-2}	7.07×10^{-2}	0.184	0.710

Table 2 (cont): demographic and COVID-19 mortality data by gender and age class for each State in US25, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

Rhode Island						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	296,955	67,924	71,466	48,982	22,203	8,370
pop'n (F)	289,151	72,082	77,092	54,289	30,712	18,089
pop'n (M/F)	586,106	140,006	148,558	103,271	52,915	26,459
deaths (M)	–	–	40	83	112	125
deaths (F)	–	–	19	67	109	246
deaths (M/F)	–	–	59	150	221	371
NMAP (M)	–	–	2.52×10^{-2}	7.62×10^{-2}	0.227	0.672
NMAP (F)	–	–	1.32×10^{-2}	6.62×10^{-2}	0.191	0.730
NMAP (M/F)	–	–	1.98×10^{-2}	7.25×10^{-2}	0.208	0.699

Texas						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	9,319,033	1,757,311	1,567,960	1,014,948	459,051	131,881
pop'n (F)	8,998,298	1,793,688	1,665,956	1,159,346	583,319	251,054
pop'n (M/F)	18,317,331	3,550,999	3,233,916	2,174,294	1,042,370	382,935
deaths (M)	66	122	263	348	302	231
deaths (F)	35	48	104	173	238	291
deaths (M/F)	101	170	367	521	540	522
NMAP (M)	2.36×10^{-3}	2.32×10^{-2}	5.60×10^{-2}	0.114	0.220	0.585
NMAP (F)	2.15×10^{-3}	1.48×10^{-2}	3.45×10^{-2}	8.25×10^{-2}	0.225	0.641
NMAP (M/F)	2.41×10^{-3}	2.09×10^{-2}	4.96×10^{-2}	0.105	0.226	0.596

Virginia						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	2,530,051	548,071	528,235	362,212	169,277	52,892
pop'n (F)	2,450,381	567,624	575,098	417,290	221,260	95,294
pop'n (M/F)	4,980,432	1,115,695	1,103,333	779,502	390,537	148,186
deaths (M)	16	47	120	192	261	226
deaths (F)	10	17	63	143	230	398
deaths (M/F)	26	64	183	335	491	624
NMAP (M)	9.49×10^{-4}	1.29×10^{-2}	3.41×10^{-2}	7.95×10^{-2}	0.231	0.641
NMAP (F)	7.16×10^{-4}	5.25×10^{-3}	1.92×10^{-2}	6.01×10^{-2}	0.182	0.732
NMAP (M/F)	8.52×10^{-4}	9.36×10^{-3}	2.71×10^{-2}	7.01×10^{-2}	0.205	0.687

Table 2 (cont): demographic and COVID-19 mortality data by gender and age class for each State in US25, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

Washington						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	2,297,145	469,013	468,346	338,061	147,359	46,935
pop'n (F)	2,183,080	461,908	492,112	372,720	177,791	81,121
pop'n (M/F)	4,480,225	930,921	960,458	710,781	325,150	128,056
deaths (M)	17	30	71	127	164	191
deaths (F)	4	16	34	90	138	225
deaths (M/F)	21	46	105	217	302	416
NMAP (M)	1.28×10^{-3}	1.11×10^{-2}	2.62×10^{-2}	6.50×10^{-2}	0.193	0.704
NMAP (F)	4.70×10^{-4}	8.89×10^{-3}	1.77×10^{-2}	6.20×10^{-2}	0.199	0.712
NMAP (M/F)	1.01×10^{-3}	1.06×10^{-2}	2.35×10^{-2}	6.57×10^{-2}	0.200	0.699

Table 2 (cont): demographic and COVID-19 mortality data by gender and age class for each State in US25, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

State	Abb'n	NMAP index		
		Male	Female	Total
Alabama	AL	0.835 (7.7%)	0.805 (5.9%)	0.801 (6.1%)
Arizona	AZ	0.781 (13.0%)	0.877 (8.3%)	0.811 (10.9%)
California	CA	0.818 (1.5%)	1.033 (4.5%)	0.888 (3.5%)
Colorado	CO	1.015 (5.1%)	1.239 (5.0%)	1.077 (5.0%)
Connecticut	CT	1.103 (2.0%)	1.337 (2.4%)	1.182 (1.8%)
Florida	FL	0.854 (0.9%)	1.072 (3.6%)	0.932 (2.5%)
Georgia	GA	0.912 (2.1%)	0.988 (3.3%)	0.938 (0.8%)
Illinois	IL	0.821 (2.3%)	1.001 (3.4%)	0.876 (3.0%)
Indiana	IN	1.061 (3.2%)	1.144 (0.7%)	1.088 (1.7%)
Louisiana	LA	0.866 (2.5%)	0.960 (4.2%)	0.894 (2.3%)
Maryland	MD	0.953 (3.2%)	1.079 (2.8%)	1.000 (2.4%)
Massachusetts	MA	1.230 (1.2%)	1.293 (3.4%)	1.249 (1.1%)
Michigan	MI	0.850 (3.1%)	0.939 (4.1%)	0.880 (3.3%)
Minnesota	MN	1.164 (1.4%)	1.361 (1.8%)	1.245 (2.2%)
Mississippi	MS	0.812 (2.1%)	0.857 (6.1%)	0.832 (3.5%)
Missouri	MO	0.981 (1.1%)	1.151 (7.6%)	1.048 (4.5%)
New Jersey	NJ	0.840 (1.6%)	1.075 (2.4%)	0.918 (2.6%)
New York	NY	0.778 (1.7%)	0.950 (2.1%)	0.827 (1.5%)
North Carolina	NC	0.977 (6.6%)	1.131 (1.2%)	1.016 (3.8%)
Ohio	OH	1.028 (2.0%)	1.200 (2.1%)	1.077 (0.8%)
Pennsylvania	PA	1.003 (2.9%)	1.170 (4.2%)	1.061 (3.9%)
Rhode Island	RI	1.094 (0.3%)	1.309 (5.6%)	1.175 (2.6%)
Texas	TX	0.782 (3.8%)	0.941 (2.8%)	0.822 (2.6%)
Virginia	VA	0.973 (2.1%)	1.213 (2.2%)	1.062 (2.4%)
Washington	WA	1.030 (5.0%)	1.118 (5.5%)	1.051 (4.7%)

Table 3: NMAP indices (NIs) measured from the slopes of log(NMAP) versus age class plots (with % error in parentheses) for each State in US25. In 63 of the 75 cases the error is ≤ 5.0% indicating the NMAP approximately follows an exponential dependency on age class, and can therefore be captured by a single parameter. Those NIs for which the error in log-linear fitting exceeds 5.0% are coloured red. These data are excluded from the analysis in Figure 3 as the NMAP in those cases does not approximately follow an exponential dependency on age class and requires more than one parameter for characterisation (cf. Figure 1 for examples). Data calculated from restricted datasets (arising from the CDC privacy policy on numbers of COVID deaths <10 in a given age class) are shaded light grey for reference.

Delaware						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	263,964	58,962	64,019	49,802	23,502	7,675
pop'n (F)	263,299	63,925	72,246	58,093	30,755	10,929
pop'n (M/F)	527,263	122,887	136,265	107,895	54,257	18,604
deaths (M)	–	–	26	56	57	58
deaths (F)	5	14	19	36	61	114
deaths (M/F)	–	–	45	92	118	172
NMAP (M)	–	–	3.53×10^{-2}	9.77×10^{-2}	0.211	0.656
NMAP (F)	1.40×10^{-3}	1.62×10^{-2}	1.94×10^{-2}	4.58×10^{-2}	0.147	0.771
NMAP (M/F)	–	–	2.62×10^{-2}	6.77×10^{-2}	0.173	0.734

District of Columbia						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	227,771	38,041	33,166	21,005	9,913	4,312
pop'n (F)	243,510	37,485	36,856	27,828	15,979	6,589
pop'n (M/F)	471,281	75,526	70,022	48,833	25,892	10,901
deaths (M)	17	31	52	102	70	50
deaths (F)	5	15	30	54	46	70
deaths (M/F)	22	46	82	156	116	120
NMAP (M)	2.87×10^{-3}	3.14×10^{-2}	6.04×10^{-2}	0.187	0.272	0.446
NMAP (F)	1.23×10^{-3}	2.40×10^{-2}	4.88×10^{-2}	0.116	0.173	0.637
NMAP (M/F)	2.28×10^{-3}	2.97×10^{-2}	5.71×10^{-2}	0.156	0.218	0.537

New Hampshire						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	359,127	94,079	104,134	70,969	30,029	11,104
pop'n (F)	349,370	95,639	108,953	75,630	37,704	19,720
pop'n (M/F)	708,497	189,718	213,087	146,599	67,733	30,824
deaths (M)	–	–	14	30	56	67
deaths (F)	–	–	–	19	41	93
deaths (M/F)	–	–	–	49	97	160
NMAP (M)	–	–	1.59×10^{-2}	5.00×10^{-2}	0.221	0.714
NMAP (F)	–	–	–	4.15×10^{-2}	0.180	0.779
NMAP (M/F)	–	–	–	4.80×10^{-2}	0.206	0.746

Table 4a: demographic and COVID-19 mortality data by gender and age class for six additional US States/Regions for use in the north-eastern US and US west coast analyses, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

New York City						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	2,520,663	504,974	465,360	306,443	154,206	52,287
pop'n (F)	2,576,520	550,900	534,851	393,493	230,971	108,080
pop'n (M/F)	5,097,183	1,055,874	1,000,211	699,936	385,177	160,367
deaths (M)	500	914	2,043	3,015	2,870	1,953
deaths (F)	171	362	966	1,645	2,117	2,550
deaths (M/F)	671	1,276	3,009	4,660	4,987	4,503
NMAP (M)	2.75×10^{-3}	2.51×10^{-2}	6.08×10^{-2}	0.136	0.258	0.517
NMAP (F)	1.68×10^{-3}	1.66×10^{-2}	4.58×10^{-2}	0.106	0.232	0.598
NMAP (M/F)	2.53×10^{-3}	2.32×10^{-2}	5.78×10^{-2}	0.128	0.249	0.540

New York State excl. New York City						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	3,103,973	737,514	777,722	525,624	247,919	94,628
pop'n (F)	2,975,433	763,950	818,284	586,035	334,745	177,634
pop'n (M/F)	6,079,406	1,501,464	1,596,006	1,111,659	582,664	272,262
deaths (M)	156	335	830	1,324	1,454	1,393
deaths (F)	69	112	403	741	1,230	2,377
deaths (M/F)	225	447	1,233	2,065	2,684	3,770
NMAP (M)	2.04×10^{-3}	1.84×10^{-2}	4.32×10^{-2}	0.102	0.238	0.597
NMAP (F)	1.22×10^{-3}	7.72×10^{-3}	2.59×10^{-2}	6.66×10^{-2}	0.194	0.705
NMAP (M/F)	1.73×10^{-3}	1.39×10^{-2}	3.61×10^{-2}	8.67×10^{-2}	0.215	0.647

Oregon						
age classes	<45	45-54	55-64	65-74	75-84	≥85
pop'n (M)	1,215,013	263,087	261,063	213,903	98,177	28,700
pop'n (F)	1,172,808	256,393	282,738	233,693	112,456	52,682
pop'n (M/F)	2,387,821	519,480	543,801	447,596	210,633	81,382
deaths (M)	–	–	16	32	26	26
deaths (F)	–	–	–	20	22	37
deaths (M/F)	–	–	–	52	48	63
NMAP (M)	–	–	4.44×10^{-2}	0.108	0.192	0.656
NMAP (F)	–	–	–	8.70×10^{-2}	0.199	0.714
NMAP (M/F)	–	–	–	0.104	0.204	0.692

Table 4a (cont): demographic and COVID-19 mortality data by gender and age class for six additional US States/Regions for use in the north-eastern US and US west coast analyses, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. A dash indicates restricted data arising from the CDC privacy policy on low numbers of COVID deaths (<10) in a given age class.

British Columbia						
age classes	40-49	50-59	60-69	70-79	80-89	≥90
pop'n	617,410	709,300	611,615	347,010	172,765	41,685
deaths	2	5	16	34	79	53
NMAP	1.74×10^{-3}	3.78×10^{-3}	1.40×10^{-2}	5.26×10^{-2}	0.245	0.682

Ontario					
age classes	<20	20-39	40-59	60-79	≥80
pop'n	3,019,640	3,475,990	3,855,065	2,505,545	592,260
deaths	1	11	115	736	1,889
NMAP	9.42×10^{-5}	9.00×10^{-4}	8.48×10^{-3}	8.35×10^{-2}	0.907

Quebec					
age classes	<20	20-39	40-59	60-79	≥80
pop'n	1,763,085	2,061,555	2,281,860	1,681,340	376,520
% deaths	–	0.2	2.2	24.0	73.7
deaths*	–	∼11	∼125	∼1360	∼4170
NMAP	–	4.60×10^{-4}	4.57×10^{-3}	6.76×10^{-2}	0.927

*The official Quebec data for deaths are provided as percentages of the total (5,666 as of July 18th 2020). The numbers of deaths are thereby inferred and rounded to integer values as a guide to the reader. The NMAP is calculated directly from the percentages data.

Table 4b: demographic and COVID-19 mortality data by age class for the three most populous Canadian Provinces, using mid-July data, along with the resulting NMAPs, calculated according to the method provided in Appendix A.1. Note the age classes for BC are decadal but shifted by 5 years relative to the US data, and the age classes for ON and QC are not decadal, but defined for twenty years periods. (The population data is from the last official census of 2016. Given NMAPs are normalised, the effect of population increases over the past four years will be small assuming demographic profiles across age groups have not changed significantly.)

State	Abb'n	NMAP index		
		Male	Female	Total
Delaware	DE	0.954 (5.0%)	0.975 (16.4%)	1.093 (8.0%)
D. of Columbia	DC	0.682 (10.7%)	0.782 (9.1%)	0.713 (7.2%)
New Hampshire	NH	1.290 (3.6%)	1.466 (0.0%)	1.371 (3.5%)
New York City	NYC	0.750 (3.8%)	0.879 (2.4%)	0.775 (3.1%)
NY excl. NYC	NY–NYC	0.866 (0.9%)	1.104 (2.9%)	0.946 (2.3%)
Oregon	OR	0.865 (10.1%)	1.052 (12.4%)	0.948 (16.7%)
Br. Columbia	BC	N/A	N/A	1.249 (4.1%)
Ontario	ON	N/A	N/A	1.144 (0.7%)
Quebec	QC	N/A	N/A	1.276 (2.4%)

Table 5: NMAP indices (NIs) measured from plots of log(NMAP) versus age class (with % error in parentheses) for the 6 additional States/Regions included in the north-eastern US and US west coast analyses, and the 3 most populous Canadian Provinces. Those NIs for which the error in log-linear fitting exceeds 5.0% are coloured red (8 out of 18 cases). These data are excluded from quantitative analysis as the NMAP in such cases does not approximately follow an exponential dependency on age class and requires more than one parameter for characterisation. Data calculated from restricted datasets (arising from the CDC privacy policy on numbers of COVID deaths <10 in a given age class) are shaded light grey for reference.