

Two chemical engineers look at the COVID-19 pandemic

Alex De Visscher  | Paõlla Chrystine Pinheiro Patrício

Department of Chemical and Materials Engineering, Gina Cody School of Engineering and Computer Science, Concordia University, Montreal, Quebec, Canada

Correspondence

Alex De Visscher, Department of Chemical and Materials Engineering, Gina Cody School of Engineering and Computer Science, Concordia University, Montreal, QC Canada.
Email: alex.devisscher@concordia.ca

Abstract

Chemical engineering involves a skill set that is transferrable to a broad range of other areas. A case in point is the work that is being done by chemical engineers to better understand and fight the COVID-19 epidemic. In this study, we consider a problem that has eluded the COVID-19 research community, which is nevertheless very tractable with a chemical engineering mindset: the true or intrinsic mortality rate of COVID-19, that is, the fraction or percentage of COVID-19-infected people that die of the disease. We solve this problem in two locations (Spain and the state of New York) for the epidemic's first wave using a combination of daily death data, a fit of a computer simulation of an epidemiological model with adjustable parameters, and independent results of immunological blood testing on a random sample of the population. Parallels are drawn with the problem of determining the turnover frequency of a catalyst based on a similar combination of data and approaches. It is concluded from the study that the intrinsic mortality rate of COVID-19 was $1.45 \pm 0.45\%$ during the first wave, a number that reflects OECD countries. By incorporating data on the age dependence of the mortality rate, a relationship $f_{\text{mort}} = (3.0 \pm 0.7) \times 10^{-5} \exp(0.1a)$, where a is the age in years, is tentatively put forward for the mortality rate as a fraction.

KEYWORDS

chemical engineering, epidemiological model, mortality rate, SARS-CoV-2

1 | INTRODUCTION

COVID-19 is a disease caused by the SARS-CoV-2 virus. It was first observed in Wuhan, China, in late 2019 and spread across the world by mid-2020.

The computer simulation of the spread of an epidemic is a problem that chemical engineers are well placed to tackle because of the similarities between the transfer of a disease and a chemical reaction. The basis of many epidemiological models is the SIR model,^[1] which divides the total population (P) into a susceptible (uninfected) group (S), an infected group (I), and a recovered (immune) group (R). One can immediately see the concept of an autocatalytic reaction in the structure of this model, as will be clear in further sections.

An overview of epidemiological models is given by Martcheva.^[2]

A number of COVID-19 modelling studies have appeared in the chemical engineering literature. Manenti et al.^[3] compared the evolution of a COVID-19 epidemic with a batch reactor. However, they modelled the disease transfer with a rate constant that varies sigmoidally with time instead of modelling it as an autocatalytic reaction. Willis et al.^[4] used an extension of the SIR model that accounts for nonpharmaceutical interventions. Both studies used regression techniques to calibrate the models to available data. Geng et al.^[5] studied the spatial spread of COVID-19 and analyzed the scaling properties. They were able to make projections on how the reopening of urban areas after a lockdown has a longer-lasting effect

on the spread of the disease than a reopening of rural areas. Epidemiological modelling has even been used in a problem (P7-8, p. 458) in Fogler's classic reaction engineering textbook.^[6]

An epidemiological model for COVID-19 has been fitted successfully to mortality data from Italy, France, and Iran in previous work.^[7,8] The model distinguishes four illness phases: infected, sick, seriously sick, and better. In the model, sick people either get better (90%) or get seriously sick (10%). Seriously sick people either get better or die.

One of the main uncertainties that has hampered decision making in the response to the COVID-19 epidemic is the uncertainty about the mortality of the disease. The case mortality rate (CMR) is often used as a surrogate for intrinsic mortality. A quick look at one of the COVID-19 tracking websites indicates that the CMR ranged from 1% to 10%, depending on the country or state, in the early stages of the pandemic. Peer-reviewed values for this period range from 2% to 6.3%.^[9–12] These are likely overestimates of the real mortality because of underreporting of cases, but the extent of the overestimation is unclear to date, particularly because a significant proportion of the population that carries the virus remains asymptomatic. Random testing of the population for COVID-19 antibodies can potentially resolve this issue by calculating the infection mortality rate (IMR). An example is the controversial Santa Clara study that estimated the IMR at 0.12%–0.2%.^[13] However, a substantial bias is introduced this way when the disease is spreading rapidly due to the time lag between infection and death. An additional bias is caused by false positives in the antibody data, particularly when the incidence of COVID-19 is low, as was the case in the Santa Clara study.^[13] These biases lead to underestimates of the true mortality of the virus. Simulations indicate that the bias caused by the time lag alone can be a factor of 10 or more when the virus spreads unchecked.^[7,8] This leaves a huge uncertainty about the real mortality of the disease.

The true mortality rate of COVID-19 is an important input parameter in any epidemiological model of the disease that needs to be known or estimated before meaningful simulations can be run. This again brings the problem into the realm of chemical engineers' expertise. It is quite common in chemical engineering to combine experimental data, computer simulations, and independent data to arrive at estimates of properties that are impossible to determine if one of these three elements is missing. An example is the determination of the turnover frequency of active sites on a catalyst. By fitting kinetic models to experimental data of a catalytic process, it is possible to determine the reaction rate on the catalyst in moles per kilogram catalyst and per second. Independently, the concentration of catalytically active sites on

the catalyst, in moles of active sites per kilogram catalyst, can be determined experimentally with chemisorption studies. By dividing the reaction rate by the concentration of active sites, we obtain a property with units s^{-1} , known as the turnover frequency. It is the number of times per second that molecules react with the average active site on the catalyst.

In the same vein, the intrinsic mortality of a disease can be obtained by fitting an epidemiological model to death data from the disease and combining it with independent results of immunological and serological studies aimed at determining the proportion of the population that has been exposed to the SARS-CoV-2 virus. In this case, the immunological testing plays the same role as the determination of the active site concentration in the catalysis example. The number of reported cases is not a reliable piece of information in this respect due to underreporting, as discussed above.

A reliable determination of the true or intrinsic mortality rate of COVID-19 has eluded the scientific community so far, in spite of the fact that the problem is a very tractable one for anyone with a chemical engineering mindset and expertise. The objective of this study is to determine the intrinsic mortality rate as a fraction or percentage of the number of people infected by the virus. For the purpose of this study, the mortality rate is treated as an intrinsic property of the disease. In doing so, this study establishes a baseline for mortality at a time when clinical practices were relatively ineffective due to a limited understanding of the disease. The study is necessarily limited to the first wave of COVID-19 because immunity to COVID-19 is of limited duration, so that later immunological studies will unavoidably underestimate the total proportion of the population that has been exposed to the virus. This problem will be easy to solve once accurate kinetic determinations of immunity loss to COVID-19 have been conducted.

2 | METHODOLOGY

2.1 | Model description

The model is an extension of the SIR model.^[1] A full description of the model is given by De Visscher,^[8] and a summary of the essential aspects is given here and in the Appendices. The model assumes a constant population (P) that can be divided into uninfected (U); infected, pre-symptomatic (I); symptomatic (S); seriously sick (SS); recovering or 'better' (B); recovered (R); and deceased (D). Transitions between the stages occur and are described like chemical reactions, each with their own rate constant k , as shown in Figure 1.

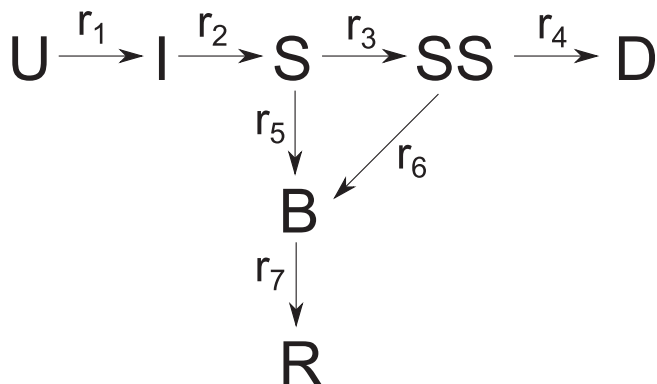
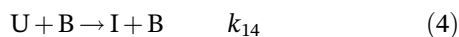
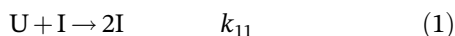


FIGURE 1 Mechanism of the COVID-19 model. B, better; D, deceased; I, infected; R, recovered; r_1 , r_2 , etc., rates of transitions (people per day); S, sick; SS, seriously sick; U, uninfected

The first transition ($U \rightarrow I$) results from four parallel processes with rate constants k_{1i} :



The transition $U + I \rightarrow 2I$ has the form of an autocatalytic reaction. It is assumed to occur with a pseudo-first-order rate constant, k_{11} , and reaction kinetics:

$$r_{11} = k_{11} I \frac{U}{P} \quad (5)$$

The division by P (total population) is needed to make the calculation scale independent.

The rate constant k_{11} is obtained by correcting its initial value, $k_{11,0}$, by a time-dependent factor that incorporates the effect of non-pharmaceutical interventions (NPIs). At the time of each NPI, a smooth transition of k_{11} from its pre-NPI to its post-NPI value is introduced (see Appendix B and De Visscher^[8] for details). All rate constants are given in De Visscher.^[8] In order to reduce the number of adjustable parameters, the ratios between the k_{1i} rate constants are fixed as follows:

$$k_{12} = \frac{k_{11}}{2} \quad (6)$$

$$k_{13} = \frac{k_{11}}{3} \quad (7)$$

TABLE 1 Parameter values for the mathematical model fitted to the death data for Spain. For the meaning of E_i and t_i , see Equation (B8) in Appendix B

Parameter	Spain	Spain
f_{mort}	0.0097	0.016
$k_{11,0}$ (day^{-1})	0.47	0.47
t_1	15 March	15 March
E_1	0.8755	0.881
t_2	1 April	1 April
E_2	0.05	0.05
t_3	12 May	12 May
E_3	0	0
t_4	21 June	21 June
E_4	−0.15	−0.15
Correction	7.7×10^{-4}	5.0×10^{-4}
Population	46.7×10^6	46.7×10^6
COVID positive	2.884×10^6	1.738×10^6

$$k_{14} = 0 \quad (8)$$

In addition to the initial conditions, only $k_{11,0}$ and its corrections through NPIs are adjustable when the mortality rate is known. The model allows for spikes in the infection rate, but spikes were not used in the analysis of any of the data discussed in this paper. The mortality is set by adjusting the rate constant k_4 of the process $SS \rightarrow D$. The default value, $k_4 = 0.012 \, 23 \, \text{day}^{-1}$, corresponds with a mortality rate of 1.5% (see Appendix A). The simulation was developed in MATLAB. The full set of differential equations is given in Appendix B. The source code is given in the Appendix S1.

2.2 | Death data and model fitting

Two locations were investigated: Spain and the state of New York.

The death data up to and including 18 July 2020 were collected from the Worldometer website on 19 July 2020.^[14] They are shown in the Table S1 in Appendix S1. Cumulative daily death data for New York up to and including 27 July 2020 were collected from *The New York Times* COVID-19 repository on 28 July 2020.^[15] The data are shown in the Table S2 in Appendix S1.

These data were used to calibrate the model summarized above.^[8] The calibration was done manually by trial and error. The model incorporated multiple NPIs. A reopening is represented as an NPI with negative effectiveness. The starting date of each intervention was based

TABLE 2 Parameter values for the mathematical model fitted to the death data for New York (NY), New York City (NYC), and the County of Monroe, New York (Monroe)

Parameter	NY	NY	NYC	NYC	Monroe	Monroe
f_{mort}	0.011	0.015	0.0125	0.015	0.01	0.015
$k_{11,0}$ (day ⁻¹)	0.5	0.5	0.5	0.5	0.35	0.35
t_1	13 March	13 March	13 March	13 March	13 March	13 March
E_1	0.145	0.1495	0.136	0.14	0.098	0.1
t_2	22 March	22 March	22 March	22 March	22 March	22 March
E_2	0.35	0.35	0.4	0.4	0.31	0.31
t_3	28 March	28 March	28 March	28 March	28 March	28 March
E_3	0.45	0.45	0.415	0.415	0.29	0.29
t_4	20 April	20 April	20 April	20 April	20 April	20 April
E_4	-0.035	-0.035	-0.03	-0.03	0.135	0.135
t_5	15 May	15 May	15 May	15 May	15 May	15 May
E_5	-0.095	-0.095	-0.14	-0.14	0	0
Correction	6.8×10^{-6}	5.0×10^{-6}	5.2×10^{-6}	4.3×10^{-6}	42×10^{-6}	28×10^{-6}
Population	19.6×10^6	19.6×10^6	8.4×10^6	8.4×10^6	0.745×10^6	0.745×10^6
COVID positive	2.712×10^6	1.999×10^6	1.719×10^6	1.430×10^6	19 768	13 037
False positive	0.0734%	4.177%	0.6765%	4.797%	1.281%	2.188%

on media reports of country-wide or state-wide initiatives to contain the epidemic. The dates chosen for the Spain model were 15 March and 1 April 2020 for the NPIs and 21 June for the reopening. An additional NPI with positive effectiveness was included to ensure a good fit with the data. The chosen date was 12 May. The 12 May NPI did not improve the agreement with the reported death data, and so the effectiveness of this NPI was set equal to 0. The dates chosen for the New York model were 13 March, 22 March, and 28 March 2020 for NPIs, and 15 May for the reopening. An additional NPI with negative effectiveness was needed to obtain a good fit with the data. The chosen date was 20 April.

The starting date of the simulation ($t = 0$) is set to 1 February 2020. The initial conditions are 100 infected, 10 sick, and 1 seriously sick, each multiplied by an adjustable factor labelled correction in Tables 1 and 2. The remaining adjustable parameters are the basic infection rate, $k_{11,0}$, in the absence of NPIs; the efficiency of each NPI as a fraction of $k_{11,0}$; and the true mortality of the disease, f_{mort} . The parameter f_{mort} is discussed in the next section.

2.3 | Infection data and mortality rate calibration

To write the kinetic parameters in terms of more tangible properties of the disease, two fractions are defined: f_{ss}

(the fraction of COVID-19-infected people that pass through the SS [seriously sick] state) and f_{mort} (the fraction of COVID-19-infected people that die). The relationships between those two fractions and the kinetic parameters are the following:

$$f_{\text{ss}} = \frac{k_3}{k_3 + k_5} \quad (9)$$

$$f_{\text{mort}} = \frac{k_3 k_4}{(k_3 + k_5)(k_4 + k_6)} \quad (10)$$

These fractions are set by adjusting kinetic parameters k_3 and k_4 . Hence:

$$k_3 = k_5 \frac{f_{\text{ss}}}{1 - f_{\text{ss}}} \quad (11)$$

$$k_4 = k_6 \frac{f_{\text{mort}}}{f_{\text{ss}} - f_{\text{mort}}} \quad (12)$$

Derivations of these equations are given in Appendix A. The values of k_5 (0.198 day⁻¹) and k_6 (0.0693 day⁻¹) were chosen based on clinical data of disease progression.^[8]

The fraction of COVID-19-positive people in Spain was taken from Pollán et al.^[16] The fraction of COVID-19-positive people in the state of New York was based on a news article.^[17]

Between 27 April–11 May 2020, an antibody test was conducted in Spain.^[16] In this study, over 60 000 randomly selected people were tested with a point-of-care test, and over 50 000 of those tested had blood samples taken for a laboratory serological test. Of those tested, 4.6% and 5.0% were positive on the point-of-care test and the laboratory test, respectively. Based on the overlap between the results, the researchers concluded that the actual number of antibody carriers is between 3.7% and 6.2% of the population. Assuming a population of 46.7 million, this leads to 1.73–2.90 million infected people in Spain at the time of the tests. In a separate test of confirmed positive cases, it was found that antibodies could be detected in 90% of confirmed infected people after 14 days, indicating that the false-negative rate of the tests was 10%.

On 22 April 2020, Governor Andrew Cuomo of New York announced that 13.9% of the population of New York state tested positive for COVID-19 antibodies in a random sampling of the population.^[17] Based on a population of 19.6 million, this suggests that 2.72 million people in the state of New York had been infected by this time. The percentages of the population that tested positive for COVID-19 antibodies in this study differed when considering New York City and upstate New York, which included 19 counties. While New York City had a 21% rate, upstate New York had only 3.9% of the population tested positive.^[17]

In a first test simulation, the death data are fitted to the model with default values of $f_{ss} = 0.1$ and $f_{mort} = 0.015$. The predicted number of infected people is

then compared with the observed numbers from the above studies.

Changing the mortality rate f_{mort} (i.e., changing k_4) requires an approximately inverse change in the initial number of infected people in the model and a near-negligible change of the efficiencies of the NPI to obtain identical predictions of the death data. Changing the mortality rate thus changes the predicted number of infected without affecting the fit of the death data. By finding the value of f_{mort} that leads to a correct prediction of the number of infected people, the mortality rate was estimated, subject to corrections as discussed below.

2.4 | Effect of false positives

It is assumed that there are no false negatives in the testing for the purpose of this calculation. The effect of false negatives is evaluated separately.

In this calculation, x is defined as the fraction of the actual number of positive cases over the population, y is the fraction of the number of false positives over the actual negative cases, and z is the fraction of the apparent positive cases over the population, which represents the sum of actual and false positives. Hence:

$$z = x + (1 - x)y \quad (13)$$

Solving for x (actual positives) leads to:

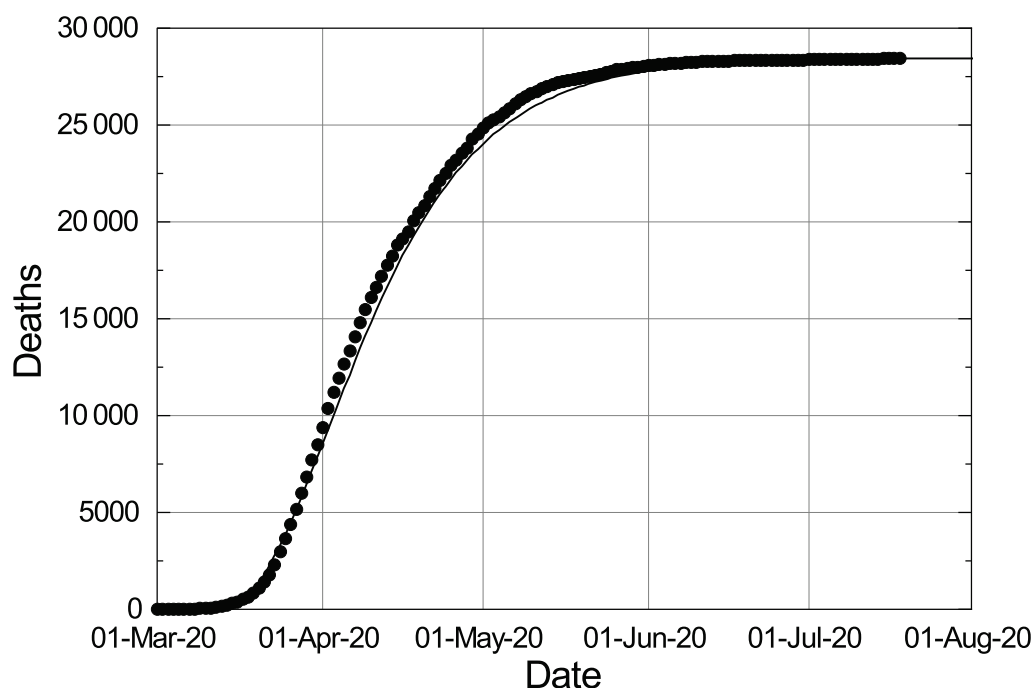


FIGURE 2 Cumulative death data by COVID-19 in Spain (circles) with model fit (solid line)

$$x = \frac{z-y}{1-y} \quad (14)$$

Solving for y (false positives) leads to:

$$y = \frac{z-x}{1-x} \quad (15)$$

For the Spain data, Pollán et al.^[16] provided a range of real positives based on the results obtained with two different tests. Their numbers were used without further processing.

The simulation of the New York data was run with different values of the intrinsic mortality, f_{mort} . For each value of f_{mort} , the model was fitted to the data, and the number of SARS-CoV-2-positive people as of 20 April 2020 was calculated. For each simulation, the proportion of false positives consistent with the simulation result was calculated with Equation (15). The relationships are plotted in Figure 3. Simulations were made for the entire state of New York, for New York City, and for the County of Monroe, NY (representing upstate New York). The lines representing the three sets of simulations have very nearly coinciding intersections, indicating that there is a unique combination of f_{mort} and the proportion of false positives that is consistent with the three data sets.

2.5 | Effect of age distribution

The age distributions of Spain and New York state were taken from <https://www.populationpyramid.net/spain/2019/>^[18] and https://www.census.gov/data/tables/2019/s36/chart_age.html^[19] respectively. For the New York data, the highest age category in the data is 85+ years, whereas for the Spain data, it is 100+ years. For consistency, all 85+ categories in the Spain data were lumped into a single group and assigned a representative age of 90 years. For Spain, this led to the same result as using the entire distribution with a representative age of 102 years for the 100+ years data. Based on works by Bonanad et al.^[20] and Goldstein and Lee,^[21] the mortalities in each age group are estimated as proportional to $\exp(0.1a)$, where a is the age in years. Hence, the fraction of the total population in each age group was calculated and multiplied with $\exp(0.1a)$, and the results were summed. The sums were 653 and 398 for Spain and New York, respectively. These numbers represent the overall mortality rate of a population divided by the extrapolated mortality at age zero. These numbers are proportional to the overall mortality rate of the two populations if we assume that both populations have

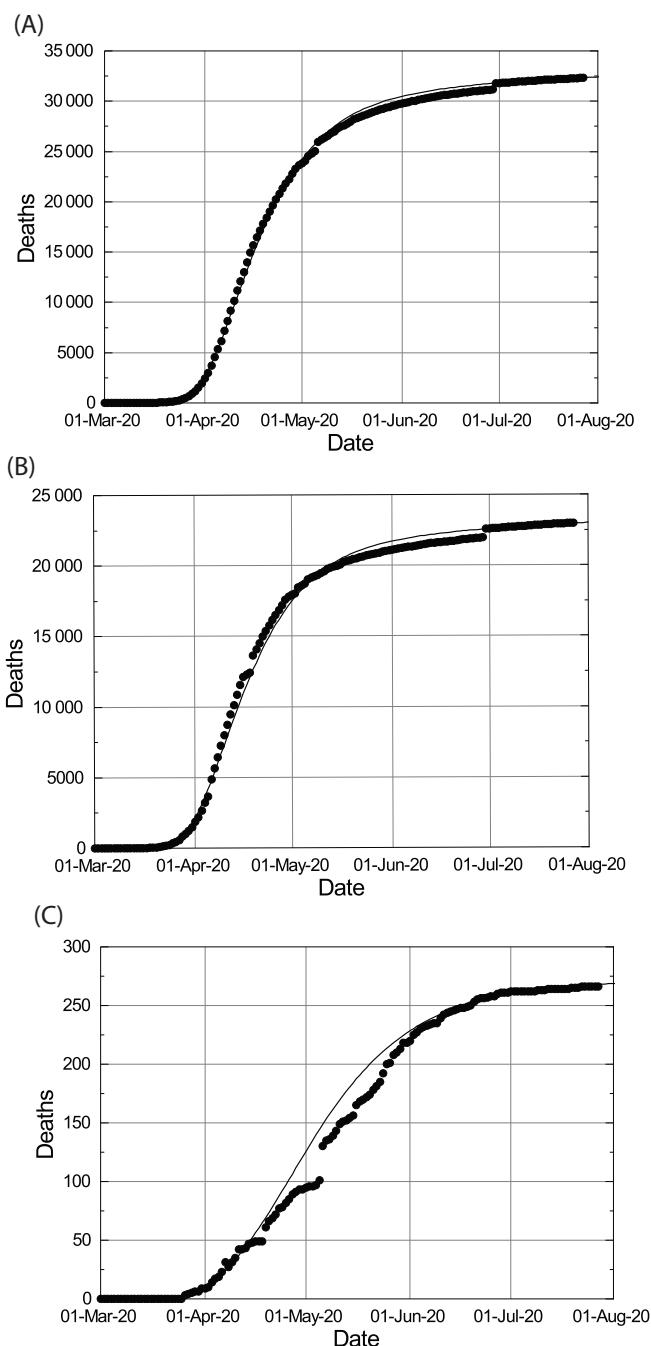


FIGURE 3 Cumulative death data by COVID-19 in (A) the state of New York (B) New York City, and (C) the County of Monroe, New York (circles) with model fit (solid line)

identical relationships between mortality rate and age. Hence, a ratio of $653/398 = 1.64$ is expected between Spain and New York. Next, the best-fitting mortality rate obtained in Spain and New York, given this ratio as a constraint, is divided by the factors calculated here (653 and 398, respectively), to obtain the coefficient A in the following equation for a universal age-mortality rate relationship:

$$f_{\text{mort}} = A \exp(Ba) \quad (16)$$

where $B = 0.1 \text{ year}^{-1}$.

3 | RESULTS

3.1 | Spain data

Figure 2 shows a fit of the mathematical model to the death data for Spain.^[14] The parameter values for the model are shown in Table 1. Seroepidemiological data of Pollán et al.^[16] is used to obtain a plausible range of mortality rates. Pollán et al.^[16] found that the seroprevalence of SARS-CoV-2 in Spain was 3.7%–6.2% between 27 April and 11 May 2020. Assuming a population of 46.7 million, this leads to 1.73–2.90 million people infected with the virus. Model fits were conducted with a range of mortalities, and the predicted number of infected people as of 5 May was compared with the range above. As Table 1 indicates, mortality rates of 0.97% and 1.6% correspond with 2.89 and 1.74 million infected, respectively. This leads to an uncorrected estimate of $1.285 \pm 0.315\%$ for the IMR. This estimate already incorporates the effect of false positives of the tests. Two more corrections are still needed.

First, the two tests used in the Spanish study were evaluated for the detectability of antibodies in people who tested positive 2 weeks prior. Both tests produced approximately 90% positive results, that is, a false-negative rate of 10%. To account for this, mortality rates should be multiplied by 0.9. A zero uncertainty is assigned to this factor.

Second, the excess death numbers in Spain, that is, the difference between the total number of deaths of all causes in Spain in 2020 and the average of the number of deaths of all causes in the preceding years, far exceed the reported COVID-19 deaths.^[22] It is expected that part of the excess death rate is due to the displacement of people with non-COVID illnesses by COVID patients in hospitals, leading to nonoptimal health outcomes and excess deaths in non-COVID patients. Hence, it is reasonable to assume that the reported number of COVID deaths represents a lower limit of the actual number, whereas the number of excess deaths represents an upper limit. For the period 11 March–5 May 2020, the cumulative excess death number is 44 072, 72% more than the reported COVID-19 deaths in Spain over the same period (25 613). Hence, a correction factor of 1.36 ± 0.36 is applied to the mortality rate.

Applying the two corrections, a true IMR of $(1.285 \pm 0.315\%) \times (1.36 \pm 0.36) \times (0.9 \pm 0) = 1.57 \pm 0.80\%$ is obtained. The error margins are obtained by adding the

relative errors, that is, by assuming that there is no correlation between the sources of error.

3.2 | New York data

Figure 3A shows a fit of the mathematical model to the death data for the state of New York.^[15] The adjustable parameter values for the model are shown in Table 2. The parameter values were adjusted with an assumed mortality rate of 1.1% and the process was repeated with an assumed mortality rate of 1.5%. The simulation results are identical for the two fits. The model predicts 2.71 million positive cases as of April 20, 2020, in the case of a 1.1% mortality rate and 1.99 million positive cases in the case of a 1.5% mortality rate. As indicated above, the actual number of infected people deduced from a serological study is 2.72 million.^[17] This number contains false positives, whereas the modelled number estimates only real infections. From the number of real cases in a given model run, and the corresponding number of measured cases in the serological study, a false-positive rate can be estimated (see Methodology section). The results are 0.07% and 4.17%, respectively, for the New York state model runs with assumed mortality rates of 1.1% and 1.5%, respectively. The relationship between the false-positive rate and the assumed mortality rate consistent with both the model and the data is shown in Figure 4. The predicted mortality rate is only weakly dependent on the proportion of false-positive cases because of the large incidence of the disease among the population.

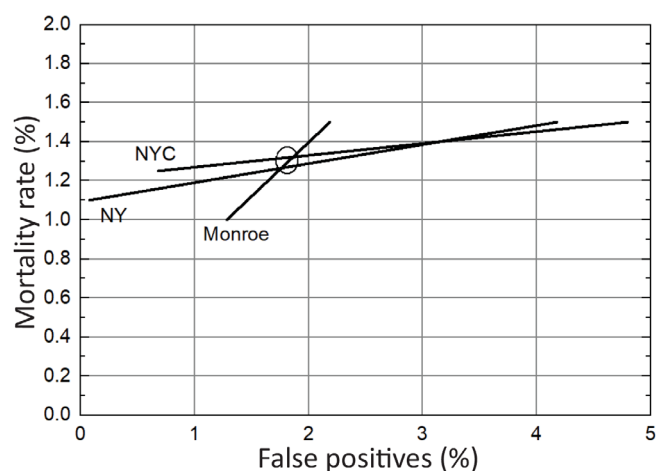


FIGURE 4 Infection mortality rate inferred by model fits as a function of assumed false-positive rate in immunological data in the state of New York (NY), New York City (NYC), and the County of Monroe, New York (Monroe) indicating parameter space consistent with the three data sets (circle)

The same procedure was followed for New York City. According to the serological study,^[11] the incidence of COVID-19 in New York City was 21% in April, or 1.76 million people (real and false positives combined), assuming a total population of 8.4 million. Model fits are shown in Figure 3B. Corresponding parameter estimates are given in Table 2. Simulations with mortality rates of 1.25% and 1.5% led to predicted numbers of positive cases of 1.72 and 1.43 million, respectively. These numbers are consistent with false-positive proportions of 0.68% and 4.80% in the serological tests, respectively. This relationship is shown in Figure 4. Again, there is a weak relationship between the false-positive proportion and the corresponding mortality rate.

In contrast, a seroprevalence of 3.9% was found in Upstate New York. In this population, false positives have a disproportionate effect on the estimated mortality. This property is used to drastically narrow down the set of plausible false-positive rates and corresponding mortality rates. The county of Monroe, NY, was used as a representative data set with an average incidence of COVID-19. The model fit to the Monroe, NY, death data is shown in Figure 3C. The corresponding parameter estimates are given in Table 2. A mortality rate of 1% corresponds with a false-positive rate of 1.28%, whereas a mortality rate of 1.5% corresponds with a false-positive rate of 2.19%. This relationship is shown in Figure 4.

Comparing the data of New York state, New York City, and Monroe, NY, in Figure 4, a small parameter region can be identified that is in general agreement with the three data sets. It will be defined somewhat arbitrarily as $1.3 \pm 0.1\%$ for the mortality rate and $1.8 \pm 0.2\%$ for the false-positive rate.

For an unbiased estimate of the true IMR, some corrections are needed. First, there is the effect of underreporting of COVID-19 deaths. Again, the reported deaths are a lower limit of the actual number of COVID-19 deaths. The excess death rate for all causes provides an upper limit of the death rate due to COVID-19.

Excess deaths during the COVID crisis are tracked in *The New York Times*.^[23] For New York City, the cumulative number of deaths during the COVID crisis between 15 March and 2 May is 23 000, which is 27.8% greater than the reported number of COVID-19 deaths during the same time period. It follows that the mortality is 1–1.278 times the estimated value, or a correction factor of 1.139 ± 0.139 .

In addition, false negatives in testing may lead to an underestimation of the number of people who contracted the virus. In the absence of specific information, a range of 0%–20% false negatives is assumed. It follows that the IMR should be multiplied by a correction factor 0.9 ± 0.1 .

Applying the two corrections, a true IMR of $(1.3 \pm 0.1\%) \times (1.139 \pm 0.139) \times (0.9 \pm 0.1) = 1.33 \pm 0.41\%$ is obtained.

4 | DISCUSSION

For Spain and New York state, true IMRs of $1.57 \pm 0.80\%$ and $1.33 \pm 0.41\%$ are obtained. The values are not significantly different and remarkably similar given the differences between the populations and the differences in the methodologies of the studies.

The numbers are also in agreement with the preliminary analysis of serological data in the Netherlands,^[8] where a lower limit of the true IMR of 1.06% was found. Because no information is available about the false-positive rate of the serological test, no upper limit could be determined in that study. However, IMRs of 1.57% and 3.12% had been estimated under the assumption of 1% and 2% false-positive rates, respectively.

In the Spanish study, where a point-of-care test was used in conjunction with an immunoassay, 3.7% of the population tested positive on both tests, whereas 6.2% of the population tested positive on at least one test. Assuming that the 2.5% testing positive on one test were false positives, the average false-positive rate for the two tests is 1.25%, less than the rate estimated for the New York test, but likely of the same order as the selectivity of the test used in the Netherlands. Taking into account that 10% of people with a prior infection may have tested negative in each of the two tests, the number of people testing positive on one test may be indicative of a false-positive rate as low as 0.85%.

Averaging the results of Spain and New York state, a value of $1.45 \pm 0.45\%$ is obtained, assuming no correlation between the errors of both studies, that is, a range of 1.0% to 1.9%. This value is for the first wave of the COVID-19 pandemic, that is, for mid-2020, and is representative of OECD countries. Improvement in understanding of the disease as a blood clotting disease rather than a purely respiratory disease and associated improvement in the recommended treatment of the disease may have reduced the mortality rate of the disease since the first wave. Also, the mortality rate only applies to unvaccinated people. Hence, care should be taken to extrapolate the results obtained in this study to situations where best practices exceed the prevailing practice during the first COVID-19 wave. In developing countries with limited medical resources, the mortality rate we obtained may still apply and even be exceeded. On the other hand, developing countries may have lower mortality rates due to differences in the age distribution (see Section 4.1).

By comparison, seasonal influenza has a CMR of 0.166% in the United States.^[24] For unvaccinated people, COVID-19 is roughly 10 times as deadly as the seasonal flu.

4.1 | Effect of age distribution

Further refinements of the analysis are possible when age distributions are accounted for. It was shown that COVID-19 mortality increases exponentially with age.^[20,21] As an average of these studies,^[20,21] a proportional relationship with $\exp(0.1a)$ is found, where a is the age in years. In what follows, it will be assumed without justification that the incidence of COVID-19 is independent of age, that is, the number of cases in every age group is proportional to the number of people in that age group. Hence, the results of this section are more tentative and prone to bias than the rest of this paper.

By assuming a mortality rate proportional to $\exp(0.1a)$ and the age distribution of Spain^[22] and New York state,^[19] it is calculated that the IMR in Spain should be 1.64 times as high as in New York, assuming all else is equal. This is much more pronounced than the ratio of 1.18 found in the preceding sections. If the confidence intervals of 0.77%–2.37% and 0.92%–1.74% are accepted for Spain and New York, respectively, and a mortality ratio of 1.64 is forced on the results, then the confidence intervals are narrowed to 1.51%–2.37% for Spain and 0.92%–1.45% for New York state. This would lead to mortalities of roughly $1.94 \pm 0.43\%$ for Spain and $1.18 \pm 0.26\%$ for New York state. Returning to the age distribution, this is consistent with the following equation for mortality as a function of age:

$$f_{\text{mort}} = A \exp(Ba) \quad (17)$$

where f_{mort} is the mortality rate as a fraction, $A = (3.0 \pm 0.7) \times 10^{-5}$, $B = 0.1 \text{ year}^{-1}$, and a is age in years. To obtain f_{mort} as a percentage, $A = (3.0 \pm 0.7) \times 10^{-3}\%$ can be used.

To evaluate how representative the average between Spain and the state of New York is in terms of age distribution, the age-dependent mortality rate obtained here is applied to the age distributions of Europe, Northern America, and Oceania. Mortalities of 1.69%, 1.46%, and 1.11% are obtained, indicating that $1.45 \pm 0.45\%$ reflects OECD countries well. In contrast, the values for Africa, Asia, and South America are 0.292%, 0.710%, and 0.806%, respectively, and the world average is 0.777%. This suggests that the mortality rate calculated in this study is representative of OECD countries but not of most of the rest of the world.

5 | CONCLUSION

Characterizing the kinetic properties of the spread, disease progression, and mortality rate of COVID-19 has clear methodological similarities with chemical engineering kinetics. In this study, it was shown that it is possible to exploit these similarities to estimate the intrinsic mortality rate of the disease. Combining reported death data of COVID-19 in Spain and in the state of New York, as well as reported excess total death data, with epidemiological modelling and independent immunological data, it is concluded that the true IMR of COVID-19 was $1.45 \pm 0.45\%$ during the first wave, which occurred in mid-2020. This is a robust estimation that accounts for false positives and false negatives in immunological testing and underreporting of the actual number of COVID-19 deaths. This is an average that represents both jurisdictions. Similar numbers can be expected in other OECD countries as well.

Accounting for the different age distributions in Spain and New York, tentative values of the IMRs of $1.94 \pm 0.43\%$ and $1.18 \pm 0.26\%$ are put forward for Spain and New York, respectively, assuming that all ages have infection numbers proportionally to their size as a fraction of the population. These numbers are consistent with an age-dependent IMR (as a fraction) of $(3.0 \pm 0.7) \times 10^{-5} \exp(0.1a)$, where a represents age in years. These numbers apply to the first wave of COVID-19, in spring 2020, in OECD countries.

NOMENCLATURE

a	age (years)
A	coefficient of the universal age-mortality rate relationship
B	coefficient of the universal age-mortality rate relationship (year^{-1})
E	fractional reduction of k_{11} during intervention
f_{mort}	fraction of COVID-19-infected people that dies
f_{SS}	fraction of COVID-19 infected people that pass through the SS state
k	rate constant (day^{-1})
r	rate (people/day)
t	time (days)
x	fraction of the actual number of positive cases over the population
y	fraction of the number of false positive over the actual negative cases
z	fraction of the apparent positive cases over the population

AUTHOR CONTRIBUTIONS

Alex De Visscher: Conceptualization; investigation; methodology; software; supervision; writing – original

draft. **Paólla Chrystine Pinheiro Patrício:** Investigation; software; validation; writing – review and editing.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/cjce.24557>.

DATA AVAILABILITY STATEMENT

All relevant data related to the paper are included in the Supplementary Document.

ORCID

Alex De Visscher  <https://orcid.org/0000-0002-2464-7605>

REFERENCES

- [1] W. Kermack, A. McKendrick, *Proc. R. Soc. London, Ser. A* **1927**, 115, 700.
- [2] M. Martcheva, *Introduction to Mathematical Epidemiology*, Springer, New York **2015**.
- [3] F. Manenti, A. Galeazzi, F. Bisotti, K. Prifti, A. Dell'Angelo, A. Di Pretoro, C. Ariatti, *Chem. Eng. Sci.* **2020**, 227, 115918.
- [4] M. J. Willis, A. Wright, V. Bramfitt, V. H. Grizales Díaz, *Chem. Eng. Sci.* **2021**, 231, 116330.
- [5] X. Geng, F. Gerges, G. G. Katul, E. Bou-Zeid, H. Nassif, M. C. Boufadel, *Chem. Eng. J.* **2021**, 420, 127702.
- [6] H. S. Fogler, *Elements of Chemical Reaction Engineering*, 4th ed., Prentice Hall, Hoboken, NJ **2006**.
- [7] A. De Visscher, ArXiv preprint 2020, 2003.08824, <https://arxiv.org/abs/2003.08824> (accessed: November 2021).
- [8] A. De Visscher, *Nonlinear Dyn.* **2020**, 101, 1871.
- [9] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. Zhu, Y. Tai, C. Bai, T. Gao, J. Song, P. Xia, J. Dong, J. Zhao, F.-S. Wang, *Lancet Respir. Med.* **2020**, 8, 420.
- [10] P. Mehta, D. F. McAuley, M. Brown, E. Sanchez, R. S. Tattersall, J. J. Manson, *Lancet* **2020**, 395, 1033.
- [11] The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, *China CDC Weekly* **2020**, 2, 145.
- [12] G. J. B. Sousa, T. S. Garces, V. R. F. Cestari, R. S. Florêncio, T. M. M. Moreira, M. L. D. Pereira, *Epidemiol. Infect.* **2020**, 148, e123.
- [13] E. Bendavid, B. Mulaney, N. Sood, S. Shah, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra-Walker, J. Tedrow, A. Bogan, T. Kupiec, D. Eichner, R. Gupta, J. P. A. Ioannidis, J. Bhattacharya, *International Journal of Epidemiology* **2021**, 50, 410.
- [14] Worldometer Coronavirus Spain, <https://www.worldometers.info/coronavirus/country/spain/> (accessed: July 2020).
- [15] The New York Times COVID-19 repository, <https://github.com/nytimes/covid-19-data> (accessed: July 2020).
- [16] M. Pollán, B. Pérez-Gómez, R. Pastor-Barriuso, J. Oteo, M. A. Hernán, M. Pérez-Olmeda, J. L. Sanmartín, A. Fernández-García, I. Cruz, N. Fernández de Larrea, M. Molina, F. Rodríguez-Cabrera, M. Martín, P. Merino-Amador, J. León Paniagua, J. F. Muñoz-Montalvo, F. Blanco, R. Yotti, ENE-COVID Study Group, *Lancet* **2020**, 396, 535.
- [17] NYC: 1 in 5 Residents Infected with Coronavirus, Says Cuomo, <https://www.nydailynews.com/coronavirus/ny-coronavirus-cuomo-20200423-alxgtumui5hk3odbusu2yr6kxq-story.html> (accessed: April 2020).
- [18] PopulationPyramid – Spain 2019, <https://www.populationpyramid.net/spain/2019/> (accessed: November 2021).
- [19] CensusScope – New York Age Distribution 2000, https://www.censusscope.org/us/s36/chart_age.html (accessed: November 2021).
- [20] C. Bonanad, S. García-Blas, F. Tarazona-Santabalbina, J. Sanchis, V. Bertomeu-González, L. Fácila, A. Ariza, J. Núñez, A. C. D. Cordero, *Journal of the American Medical Directors Association* **2020**, 21, 915.
- [21] J. R. Goldstein, R. D. Lee, *Proc. Natl. Acad. Sci. U. S. A.* **2020**, 117, 22035.
- [22] Covid-19-excess-deaths-tracker, https://github.com/TheEconomist/covid-19-excess-deaths-tracker/blob/master/output-data/excess-deaths/spain_excess_deaths.csv (accessed: November 2021).
- [23] The New York Times, *The New York Times*, May 5, **2020**.
- [24] N.-A. M. Molinari, I. R. Ortega-Sanchez, M. L. Messonnier, W. W. Thompson, P. M. Wortley, E. Weintraub, C. B. Bridges, *Vaccine* **2007**, 25, 5086.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: A. De Visscher, P. C. Pinheiro Patrício, *Can. J. Chem. Eng.* **2022**, 1. <https://doi.org/10.1002/cjce.24557>

APPENDIX A: DERIVATION OF EQUATIONS FOR f_{ss} AND f_{mort}

First, we derive an equation for f_{ss} , the fraction of infected people that go through the seriously sick stage. We see from Figure 1 that sick people convert to seriously sick people with the reaction rate r_3 , whereas the overall reaction rate of sick people is $r_3 + r_5$. Hence, the fraction is:

$$f_{ss} = \frac{r_3}{r_3 + r_5} \quad (A1)$$

Substituting first-order kinetics into the reaction rates leads to:

$$f_{ss} = \frac{k_3 S}{k_3 S + k_5 S} = \frac{k_3}{k_3 + k_5} \quad (A2)$$

To calculate rate constant k_3 from a given value of f_{ss} , we can use:

$$k_3 = k_5 \frac{f_{SS}}{1 - f_{SS}} \quad (11)$$

Based on hospital admission data, a value of $f_{SS} = 0.1$ was put forward.^[8] Based on clinical data of disease progression, a value of $k_5 = \ln(2)/3.5 = 0.198 \text{ day}^{-1}$ was put forward.^[8] This leads to a value of $k_3 = 0.022 \text{ day}^{-1}$.

Next, we calculate the mortality rate, which can be defined as reaction rate r_4 divided by reaction rate r_2 . When S is at a steady state, we can assume:

$$r_2 = r_3 + r_5 \quad (A3)$$

Hence, the mortality rate is:

$$f_{\text{mort}} = \frac{r_4}{r_3 + r_5} \quad (A4)$$

Substituting the reaction kinetics leads to:

$$f_{\text{mort}} = \frac{k_4 SS}{k_3 S + k_5 S} \quad (A5)$$

We cannot eliminate S and SS without an additional assumption. That assumption is the pseudo-steady state approximation of the number of seriously sick people:

$$r_3 = r_4 + r_6 \quad (A6)$$

Substituting kinetics leads to:

$$k_3 S = k_4 SS + k_6 SS \quad (A7)$$

Solving for S leads to:

$$S = \frac{k_4 + k_6}{k_3} SS \quad (A8)$$

This is substituted in the equation for f_{mort} :

$$f_{\text{mort}} = \frac{k_4 SS}{(k_4 + k_6) SS + k_5 \frac{k_4 + k_6}{k_3} SS} \quad (A9)$$

This equation can be simplified to:

$$f_{\text{mort}} = \frac{k_3 k_4}{(k_3 + k_5)(k_4 + k_6)} \quad (A10)$$

To calculate the rate constant k_4 from the mortality rate, we can use:

$$k_4 = k_6 \frac{f_{\text{mort}}}{f_{SS} - f_{\text{mort}}} \quad (12)$$

Based on clinical information on the progression of the disease, a value of $k_6 = \ln(2)/10 = 0.0693 \text{ day}^{-1}$ was chosen.^[8] Assuming a mortality rate of 1.5%, this leads to $k_4 = 0.0693 \times 0.015/0.085 = 0.01223 \text{ day}^{-1}$.

APPENDIX B: DIFFERENTIAL EQUATIONS OF THE MODEL

Below are the differential equations that describe the dynamics of the number of uninfected (U), infected (I), sick (S), seriously sick (SS), deceased (D), recovering (B), and recovered immune (R) people out of a population P .

$$\frac{dU}{dt} = -(k_{11}I + k_{12}S + k_{13}SS + k_{14}B) \frac{U}{P} \quad (B1)$$

$$\frac{dI}{dt} = (k_{11}I + k_{12}S + k_{13}SS + k_{14}B) \frac{U}{P} - k_2 I \quad (B2)$$

$$\frac{dS}{dt} = k_2 I - k_3 S - k_5 S \quad (B3)$$

$$\frac{dSS}{dt} = k_3 I - k_4 SS - k_6 SS \quad (B4)$$

$$\frac{dB}{dt} = k_4 SS \quad (B5)$$

$$\frac{dB}{dt} = k_5 S + k_6 SS - k_7 B \quad (B6)$$

$$\frac{dR}{dt} = k_7 B \quad (B7)$$

where the rate constant k_{11} is calculated as:

$$k_{11} = k_{11,0} \left(1 - \sum_i E_i \left(\frac{1}{2} - \frac{1}{2} \text{erf}(t - t_i) \right) \right) \quad (B8)$$

In this equation, E_i is the incremental effectiveness of a nonpharmaceutical intervention mandated at time t_i . It is relative to the original infection rate constant $k_{11,0}$. A negative value of E_i indicates an increase in the infection rate, for instance, due to a reopening. The term

$$E_i \left(\frac{1}{2} - \frac{1}{2} \operatorname{erf}(t - t_i) \right)$$

ensures a smooth transition from zero incremental effectiveness shortly before t_i to incremental effectiveness E_i shortly after t_i . The other kinetic constants are discussed in the main text.