Regular Article

Estimation of the basic reproduction number of COVID-19 from the incubation period distribution

Lasko Basnarkov[1,2,a], Igor Tomovski[2,b], and Florin Avram[3,c]

[1] Faculty of Computer Science and Engineering, SS Cyril and Methodius University, 1000 Skopje, Macedonia
[2] Macedonian Academy of Sciences and Arts, 1000 Skopje, Macedonia
[3] Laboratoire de Mathématiques Appliqués, Université de Pau, 64000 Pau, France

Abstract The estimates of the future course of spreading of the SARS-CoV-2 virus are frequently based on Markovian models in which the duration of residence in any compartment is exponentially distributed. Accordingly, the basic reproduction number R_0 is also determined from formulae where it is related to the parameters of such models. The observations show that the start of infectivity of an individual appears nearly at the same time as the onset of symptoms, while the distribution of the incubation period is not an exponential. Therefore, we propose a method for estimation of R_0 for COVID-19 based on the empirical incubation period distribution and assumed very short infectivity period that lasts only few days around the onset of symptoms. We illustrate this venerable approach to estimate R_0 for six major European countries in the first wave of the epidemic. The calculations show that even if the infectivity starts 2 days before the onset of symptoms and stops instantly when they appear (immediate isolation), the value of R_0 is larger than that from the classical, SIR model. For more realistic cases, when only individuals with mild symptoms spread the virus for few days after onset of symptoms, the respective values are even larger. This implies that calculations of R_0 and other characteristics of spreading of COVID-19 based on the classical, Markovian approaches should be taken very cautiously.

1 Introduction

The ongoing pandemic which resulted from spread of the SARS-CoV-2 virus has recruited scientists from various disciplines to apply their knowledge and ideas in understanding and modeling the epidemic spreading, as support to the effort that the healthcare workers are doing in preventing and mitigating the effects of the disease. The key modeling tools come from the mathematical epidemiology that dates back to Daniel Bernoulli [1] and has been built upon the cornerstones from the works by Ross [2–4] and Kermack and McKendrick [5]. The most famous SIS and SIR models are well known and applied even in other fields like spreading of rumor and computer viruses [6–8]. Their popularity, as well as that of their extensions, among others might be attributed to the mathematical framework used as ground—systems of ordinary differential equations. The simpler cases allow for analytical treatment to large extent, while numerical solutions are

applied for the more complex ones. The solutions of such models provide a perspective of the development of an epidemic under various constraints and active measures. For example, for the COVID-19, various extensions have been developed to estimate the numbers of infected, hospitalized, or casualties under different scenarios of the epidemic [9–15]. Common feature of these models is the Markovian assumption, which technically means that the transitions from one state of the disease to another, or the related infectiousness of the individuals, are independent on the past. This might be appropriate for diseases in which an infected person can start spreading the disease in very short period after contracting the pathogen. In general, this assumption is not always empirically supported. Particularly, COVID-19 is a disease which has been found to have particularly long period from infection to onset of symptoms—incubation period [16–18]. It was also observed that the spreading ability of an infected individual becomes significant only when the incubation period is near its end [19], thus questioning the reliability of the results obtained with the classical Markovian models. Although this delayed onset of infectiousness can be somehow addressed with introducing the state of Exposed (infected, but not yet infectious), in the Markovian framework, the infectiousness period would

[a] e-mail: lasko.basnarkov@finki.ukim.mk (corresponding author)
[b] e-mail: igor@manu.edu.mk
[c] e-mail: florin.avram@univ-pau.fr

have distribution with certain form that is not general. However, besides this observation, the inherent mathematical tractability makes the Markovian models dominant in the contributions related to the epidemic spreading, while the more general, non-Markovian ones are still very rare [20–24]. One should note that description and prediction of the development of the COVID-19 epidemic are not based solely on the compartmental models. For example, contributions from physical community apply other approaches like fractal interpolation method [25, 26] and generalized fractal dimension [27].

The basic reproduction number R_0 is the most popular quantity in epidemic spreading studies, because it bears key information. When $R_0 < 1$, an outbreak of epidemic is expected to fade away, while if $R_0 > 1$, it would grow. In the latter case, R_0 determines also the final size of the epidemic and the herd immunity level, achieved naturally or through vaccination, which expressed as fraction of the total population is given as $1 - 1/R_0$ [28]. The basic reproduction number is defined as expected number of newly infected persons by one infectious individual introduced in a completely susceptible population. The more general, effective reproduction number, R, corresponds to the situation when certain fraction of the population is already infected. The individual that passes the pathogen to the others is known as primary case, or infector, while, those who get the pathogen from him or her, as secondary cases, or infectees. In calculations of R_0, even in the earliest works in mathematical epidemiology, it was applied the general, non-Markovian framework. It is thus considered that the contagiousness ability of an infector depends on the time passed, since she or he become infected—the age of infection τ. The infectivity potential of the infector combined with the social behavior that further determines the spreading effectiveness during her or his contacts with the others while being in different stages of the disease is associated with a statistical quantity known in epidemiology as generation interval distribution $g(\tau)$. It represents the probability density of the time interval between the moments of infection of the primary and secondary case [29]. Although the approach which relies on the generation interval distribution considers general form of its shape and is thus likely most appropriate, its use for estimation of R_0 for given epidemic could be rather complicated due to the difficulties in precise determination of the moments of infection. Accordingly, various approximate alternatives are applied. Sometimes, as a substitution one uses the serial interval distribution that is the probability density of the period between the onset of symptoms at the primary and secondary case. It is more reliable to estimate this distribution, since it is easier to spot the moment, or at least the day of onset of symptoms, than the day of infection of an individual. One interesting observation is that the serial interval might be negative for significant fraction of infector–infectee pairs for COVID-19 [30], while the generation interval is strictly positive. This makes the estimates with this distribution as a proxy of the generation interval distribution questionable. In other approaches, only the mean

and variance of the generation interval distribution are estimated and then these are applied for determination of R_0. We emphasize that while using the generation interval distribution or its proxies one does not need to have any particular epidemic spreading model to calculate R_0. However, when some Markovian model is applied for studying an epidemic, the basic reproduction number is determined through certain relationship that involve the parameters of that model.

For the COVID-19, one can find estimates of the basic reproduction number based on different approaches. There are contributions that rely on the classical compartmental models by direct analysis of the models [31, 32], or by applying the next generation matrix approach [33–35]. In the non-Markovian setting, there can be found works that use the mean and variance of the generation interval distribution [36], or the serial interval distribution as its proxy [37–39] and so on. To circumvent the difficulties in direct estimate of the generation interval distribution, we propose a non-Markovian approach that aims to model it, using the incubation period distribution as a basis. The latter is further combined with a window function of infectiousness of an individual—the period when she or he can infect the others. The incubation period distribution is much easier to estimate than the generation interval distribution. For example, one needs to know only the days of onset of symptoms of infected individuals which attended a major social event, which is very likely a place where they have contracted the pathogen. Collecting such data is more reliable than deducing possible day of infection of infector–infectee pairs. The proposed method for design of the generation interval distribution is used to estimate the basic reproduction number for the first wave of epidemics in the spring, 2020, in the six most populous countries in Europe, where the countrywide epidemic with large number of cases appeared earlier than in other large countries. Our estimates from the most conservative scenarios which were considered are larger than those obtained with the Markovian approach in the classical SIR model. The other estimates based on more realistic assumptions are even larger. The observation that the estimates of R_0 with more realistic models like the one proposed here are larger than those with the classical approach suggests that the results from the latter should be considered very carefully. This holds for the results they provide and for the use of the Markovian assumption they rely on, which is not always empirically supported. We also believe that this framework could be further refined to obtain a complementary tool for determination of the generation interval distribution and the basic reproduction number of an epidemic.

The paper is organized as follows. In Sect. 2, we provide the theoretical basis of the formula used for estimation of R_0. In Sect. 3, it is elaborated on the generation interval distribution that is used for estimates of R_0. In the next, Sect. 4 presents the results and a discussion on them and we finish the paper with the conclusions.

2 Methods

The derivation of the formula that we use for estimation of the basic reproduction number for arbitrary generation interval distribution can be found in various works in the literature (for example in [29, 40, 41]). We stress that similar reasoning to its derivation appears already in the earliest studies in demography by Böckh [42] and in epidemiology by Lotka [43]. For completeness, we present here an approach at population level, since it is more appropriate for study of the available data. Most of the analysis will be made in discrete time as more appropriate, but continuous-time version will be given for completeness.

As is the case of COVID-19, when new cases are reported on daily basis, it is convenient to have a function $I_d(t)$ that represents the fraction between confirmed cases for day t and the total population. When the epidemic is in inception phase, the number of newly infected individuals grows exponentially and one has the following exponential form:

$$I_d(t) = I_0 e^{\lambda t}, \tag{1}$$

where λ is a parameter for the growth rate, while I_0 is a constant. The growth rate λ is related with the period of doubling with $T_d = \ln 2 / \lambda$. For many diseases, the individuals are not able to infect the others immediately, but after certain period has passed and this has to be accounted for. Before doing that, we remind first that the generation interval distribution $g(\tau)$ represents the probability distribution of the time period between infection of the primary and secondary case. Another interpretation is that $g(\tau)$ is the likelihood that some secondary case will appear for time τ later after the primary case has contracted the pathogen. At population level, it quantifies the fraction of new infectees, that have become infected due to contacts with unit fraction of infectors with age of infection τ. We emphasize that it is assumed that the generation interval distribution depends only on the time that has passed since contracting the disease, or the age of infection τ, but not on when that happened. This time-invariance assumption is mathematically plausible and is applicable when the pathogen does not mutate very fast, or condition for its spread does not change significantly. For that reason, the time t does not appear in the generation interval distribution g. We further assume that $g(\tau)$ has finite support T which means that $g(\tau) = 0$, for $\tau > T$. More details on the shape of this function will be provided later on. Denote with $I(t, \tau)$ the fraction of the total population of the infectees that have become infected at moment t, by having a contact with infectors that have acquired the pathogen τ time units earlier. Assuming homogeneous mixing of the population, $I(t, \tau)$, will be proportional to the fraction of susceptibles $S(t)$, but also to the fraction of the infectors with age τ, $I_d(t-\tau)$, and the intensity of their infectiousness encapsulated in the generation interval distribution $g(\tau)$. Thus, one has

the following relationship:

$$I(t, \tau) = R_0 S(t) I_d(t - \tau) g(\tau), \tag{2}$$

where the constant of proportionality R_0 will be shown to be exactly the basic reproduction number. The fraction of the new infectees within given moment will be obtained as sum that accounts for the contributions from the infectors of all ages

$$I_d(t) = \sum_{\tau=1}^{T} I(t, \tau), \tag{3}$$

which will further result in the following recurrent relationship:

$$I_d(t) = R_0 S(t) \sum_{\tau=1}^{T} I_d(t - \tau) g(\tau). \tag{4}$$

By applying the exponential form for the function of newly infected individuals (1), in the last relationship, one will have

$$I_0 e^{\lambda t} = I_0 R_0 S(t) e^{\lambda t} \sum_{\tau=1}^{T} e^{-\lambda \tau} g(\tau), \tag{5}$$

from where it follows that:

$$1 = R_0 S(t) \sum_{\tau=1}^{T} e^{-\lambda \tau} g(\tau). \tag{6}$$

Since the basic reproduction number corresponds to the inception of the epidemic $S(t) \approx 1$, for the parameter R_0, one has

$$R_0 = \frac{1}{\sum_{\tau=1}^{T} e^{-\lambda \tau} g(\tau)}. \tag{7}$$

We emphasize that the last relationship holds for generation interval distribution with general form. Finally, if one has obtained R_0 from (7) and it is assumed to be constant during the course of the epidemic, then the herd immunity threshold expressed in terms of the susceptible fraction S_{th} when the further spreading will stop, and which corresponds to unity reproduction number $R = 1$, can be obtained from (6) as

$$S_{th} = \frac{1}{R_0}. \tag{8}$$

It is worth noting that in derivation of the equation for the basic reproduction number (7), we did not use any of the well-known deterministic epidemic spreading models like SIS, SIR, or SEIR. Thus, if one needs only to estimate R_0, than it is not needed to refer to any specific model. However, a noticeable similarity in

the reasoning in those models and the approach presented here lies in the determination of the fraction of new infectees (2). In those models as well as in this framework, the homogeneous mixing makes the fraction of new infectees to be proportional to the existing infectors and susceptibles. For that reason, in this approach, the infected individuals present the Exposed ones when they do not infect yet, while they correspond to the Infected compartment when they do transmit the pathogen to the others. The Quarantine, Vaccinated, Recovered and other compartments are not considered in this setting, since they are not needed.

2.1 Self-consistency relationship for R_0

The basic reproduction number represents an estimated multiplicative factor determining the number of newly infected individuals that will contract the spreading agent in a contact with certain infected individual. Here, we use it to denote the growth factor that determines the new infections that will arise from the sub-population of all individuals that have become infected at the same time. Thus, form one side, the fraction of all infectees that have contracted the pathogen from infectors that become infected at certain moment t should be $R_0 I_d(t)$. From another side, this fraction can be represented as forward-time sum of future infectees as follows:

$$R_0 I_d(t) = \sum_{\tau=1}^{T} I(t + \tau, \tau). \quad (9)$$

Using (2), one has

$$R_0 I_d(t) = R_0 \sum_{\tau=1}^{T} S(t + \tau) I_d(t) g(\tau). \quad (10)$$

For slowly growing epidemic, one might assume that the fraction of susceptibles does not change significantly in the considered period $S(t) \approx 1$. Then, from the last expression (10), one can see that the constant R_0 in the relationship (2) will be the basic reproduction number, if the generation interval distribution is properly normalized $\sum_{\tau=1}^{T} g(\tau) = 1$.

2.2 Continuous-time formula

Although analysis of the data in this work is based on discrete time, for completeness, we provide the continuous-time version that can be found in the literature in similar forms as given here. Denote with $I(t)$ the fraction of the population that has become infected within infinitesimal interval $(t, t + dt)$, and assume that it grows exponentially at the onset of epidemics $I(t) = I_0 e^{\lambda t}$. The new infectees have appeared from contacts with others that have been infected in the past $I(t - \tau)$. We consider the same meaning of the generation interval distribution $g(\tau)$ on which we

also impose finite support $(0, T)$. Now, the formula for newly infected population will read

$$I(t) = R_0 S(t) \int_0^T I(t - \tau) g(\tau) d\tau, \quad t \geq T, \quad (11)$$

where again R_0 is a parameter and $S(t)$ is the fraction of susceptibles at moment t. One should note that similar relationship has appeared earlier in the works of Ross and Hudson [44]. By plugging in the exponential form of the newly infected individuals, one will obtain

$$I_0 e^{\lambda t} = R_0 S(t) \int_0^T I_0 e^{\lambda(t-\tau)} g(\tau) d\tau, \quad (12)$$

which will reduce to similar relationship as that for the discrete time

$$1 = R_0 S(t) \int_0^T e^{-\lambda\tau} g(\tau) d\tau. \quad (13)$$

At the onset of an epidemic $S(t) \approx 1$, and for the basic reproduction number will be obtained the Lotka–Euler equation [45, 46]

$$R_0 = \frac{1}{\int_0^T e^{-\lambda\tau} g(\tau) d\tau}. \quad (14)$$

3 Shape of the generation interval distribution

The function $g(\tau)$ should model the generation of new infections by the fraction of the population that become infected within the same unit interval, at some later moment τ after contracting the spreading agent. Its shape, in a direct approach, should be deduced from epidemiological tracing of infector–infectee pairs. This means that one has to determine the dates of infection of all infectees for given infectors for which one also knows the dates of contracting the pathogen, that is far from trivial task for diseases like the COVID-19. If one succeeds in such a task, the resulting histograms would be conveniently fit with unimodal functions like the log-normal or the Weibull distribution as ones with support on the positive real numbers. To circumvent the difficulties in obtaining such epidemiological data, in our approach, we aim to deduce the shape of $g(\tau)$ using other characteristics of the respective disease. To get to its final form, we first note that its shape depends on the fractions of the individuals that will not recover within given time interval and thus is related to the function that describes the healing process. However, since the COVID-19 is a disease which has long period of recovery, we consider that all detected individuals have not recovered yet in the period under study. Furthermore, the infectiousness ability of a subject, besides

being delayed for certain incubation period, depends on the age of infection through intensity of the viral load and the contacts between individuals which depend on the health status of the infector, and also on her, or his awareness about being possibly positive on the virus. Another important observation is that the infectiousness of COVID-19 starts even before onset of symptoms [19]. We consider here that the period of the effective infectiousness (when new infections result from given subject) is short and thus for simplicity assume that it is constant and equal for all individuals before onset of symptoms. After appearance of symptoms, the more severe cases would immediately reduce the contacts with the others and thus would not be significant infectors any more, while those with mild symptoms, that constitute about 80% of the cases [47], would continue with their normal daily life and would spread the virus with equal intensity. Therefore, we consider an infectivity window function $w(t)$ which is non-zero only in the period $t \in [t_{\mathrm{init}}; t_{\mathrm{end}}]$. It means that the infectivity starts at certain initial moment t_{init} before the onset of symptoms, and ends at t_{end} after they appear. Thus, the shape of the infectivity window function $w(t)$ is one consisting of two steps

$$w(t) = \begin{cases} 1, & t_{\mathrm{init}} \leq t < 0 \\ 0.8, & 0 \leq t \leq t_{\mathrm{end}}. \end{cases} \quad (15)$$

We further assume that the shape of the infectivity window does not depend on the infection age τ when the symptoms appear. Thus, the generation interval distribution will be convolution of the incubation period distribution $\beta(\tau)$ and the infectivity window $w(\tau)$

$$g(\tau) = C(\beta * w)(\tau) = C \sum_{\nu} \beta(\tau - \nu)w(\nu), \quad (16)$$

where C is a constant determined from the normalization condition $\sum_{\tau=1}^{T} g(\tau) = 1$. Therefore, the differences between the individuals related to the different response to the virus and the related onset of infectivity are encapsulated in the incubation period distribution, while the differences in the severity of the symptoms and the related infectiousness intensity in the infectivity window function.

4 Results and discussion

We have used the data from the Our World in Data database for the numbers of daily cases of COVID-19 in the first wave in the spring 2020 for six major European countries. As data window for study, we have taken the period that stars at the moment from which new cases were reported every day, until the day when lockdown measures were introduced. Two exceptions are France and Spain, for which the period of study ends with the day when no cases were reported. We have opted for this choice for these two countries, since the number for

the next day accounts for the cases of the last 2 days and does not represent a daily count. We note that this approach for France and Spain does not shorten significantly the considered period. Such choice for the range of data was made under the assumption that in the given period, the virus SARS-CoV-2 was spreading almost freely in the population, with only positive cases being isolated. From those numbers, we have fit an exponential curve and estimated the growth rate factor λ. The description of the considered period and the growth rate estimates are summarized in Table 1.

We have used the estimates of the growth factor λ in the expression for the calculation of the basic reproduction number (7). The generation interval distribution function $g(\tau)$ was obtained from the convolution (16) of windows with variable width, and incubation period distributions obtained from two sources in the literature [16, 17].

The onset of infectiousness was considered to be two [19] or 1 day before the appearance of symptoms. We have also assumed that after the onset of symptoms only the mild cases, that are about 80% [47], can be considered as further infectors. As the end of the infectiousness period was considered either the same day or 2 days after the onset of the symptoms. We have thus chosen three separate cases combined with the two incubation period distributions, as described in more detail in Table 2. The first case is very conservative and is related to the assumption that all infectors, even those with mildest symptoms, would become very cautious in their contacts with the others and would not spread the disease further. The other two cases are less conservative and likely a more realistic scenario, where those with mild symptoms do not change their daily routines and spread the virus freely before receiving positive test. Extending the infectivity window beyond the point of 2 days, following the onset of symptoms, would relate to sporadic cases only, if one considers the severe measures imposed by the governments in preventing the spread of COVID-19. In that sense, we consider that the analyzed six scenarios are sufficient to illustrate the dynamics of spreading in reality.

The estimates of R_0 for the different scenarios for all six European countries are presented in Table 3. One can easily note that the estimate is very sensitive on the start of the infectiousness, in relation to the onset of symptoms, as evidenced from the results for scenarios 3

Table 1 Countries under study and the estimated exponential growth factor

Country	Start date	End date	λ
Germany	February 25	March 12	0.298
France	February 25	March 11	0.322
Italy	February 21	March 9	0.235
Russia	March 12	March 30	0.171
Spain	February 25	March 11	0.343
United Kingdom	February 23	March 22	0.237

Table 2 Description of different scenarios for calculation of basic reproduction number, based on the non-Markovian approach

No.	Description	Function	Parameters
S1	Start 2 days before onset, stop immediately	Log-normal[a] [16]	$P_1 = 1.62; P_2 = 0.42$
S2	Start 2 days before onset, stop 2 days after	Log-normal [16]	$P_1 = 1.62; P_2 = 0.42$
S3	Start 1 day before onset, stop 2 days after	Log-normal [16]	$P_1 = 1.62; P_2 = 0.42$
S4	Start 2 days before onset, stop immediately	Weibull[b] [17]	$P_1 = 2.04; P_2 = 0.103$
S5	Start 2 days before onset, stop 2 days after	Weibull [17]	$P_1 = 2.04; P_2 = 0.103$
S6	Start 1 day before onset, stop 2 days after	Weibull [17]	$P_1 = 2.04; P_2 = 0.103$

[a] For the log-normal distribution, P_1 is the mean and P_2 is the standard deviation of the logarithm of the distribution
[b] For the Weibull distribution, P_1 is the scale, while P_2 is the shape parameter

and 6 versus the other ones. Next, only scenarios 1 and 4, as the most conservative ones provide estimates that are roughly comparable to those found in the literature for the European countries [31, 36]. The more realistic assumptions in the other cases produce even larger values. In reality, there would have been cases for even longer period of infectivity than 2 days after the onset of symptoms, and this could be the true for significant fraction of infected population. Although we have not tried to model this, the logic suggests that the longer the infectivity period, after the onset of symptoms, the

larger R_0 is. This implies that better estimates of the generation interval distribution are essential in providing better assessment of the basic reproduction number. One such approach could be based on a better model of the infectivity window function $w(\tau)$, that is the period when an infected person is infectious, in relation to the onset of symptoms. The situation can be even more complicated if the infectivity window function is dependent on the incubation period duration. It could be possible for those that develop symptoms earlier to correspond one infectivity window function, while for those that develop them after longer period—another one. Thus, one has to be very cautious with using the estimated R_0, particularly when the incubation period of the related disease has pronounced duration and when the respective generation interval distribution is known only vaguely.

As a comparison, we have also made an estimate based on the Markovian approach with the classical SIR model expressed with the well-known set of differential equations [48]

$$
\begin{aligned}
\dot{S} &= -\beta S(t)I(t), \\
\dot{I} &= \beta S(t)I(t) - \gamma I(t), \\
\dot{R} &= \gamma I(t),
\end{aligned}
\tag{17}
$$

where $S(t)$, $I(t)$, and $R(t)$ are the fractions of the susceptible, infectious, and recovered individuals at given moment, while β and γ are parameters. The basic reproduction number is obtained from the parameters as $R_0 = \beta/\gamma$ [48]. To find R_0, we have numerically integrated the SIR model (17) and calculated the new infections at day k with the integral

$$
I_{\text{daily}}^{\text{SIR}}(k) = \beta \int_{k-1}^{k} S(t)I(t)dt.
\tag{18}
$$

The simulations were run with only one infected individual at the initial moment $t = 0$. The values of the parameters β and γ were obtained by least-squares error function between the logarithms of daily cases from observations and from simulations with

$$
\epsilon = \text{argmin}_{s,\beta,\gamma} \left\{ \frac{1}{N} \sum_{k=1}^{N} \left[\ln(I_{\text{data}}(k)) - \ln(I_{\text{daily}}^{\text{SIR}}(k+s)) \right] \right\}.
\tag{19}
$$

Table 3 Estimated basic reproduction number R_0 for six European countries for the six non-Markovian scenarios and the classical SIR model

Country	S1	S2	S3	S4	S5	S6	SIR
France	3.22	4.06	5.02	5.37	6.72	8.01	3.06
Germany	3.00	3.73	4.53	4.93	6.10	7.17	2.15
Italy	2.45	2.95	3.42	3.85	4.60	5.22	2.09
Russia	1.97	2.27	2.51	2.87	3.29	3.60	1.45
Spain	3.43	4.38	5.49	5.78	7.31	8.81	2.67
United Kingdom	2.47	2.97	3.45	3.89	4.65	5.28	1.70

Table 4 Parameters of the classical SIR model

Country	Population (2020 est.)	β	γ	Initial day
France	67 million	0.49	0.16	7
Germany	83 million	0.56	0.26	7
Italy	60 million	0.46	0.22	19
Russia	144 million	0.55	0.38	18
Spain	47 million	0.56	0.21	5
United Kingdom	67 million	0.56	0.33	8

In the last equation, s is the first day from the simulations which is assumed to coincide to the first day of observed data for each country, while N is the number of days for which the analysis is made. The estimated values of β and γ were used to calculate R_0 and the results are given in the last column in the Table 4. It is noticeable that the values of R_0 based on the classical Markovian SIR model for the COVID-19 seem to be significant underestimates.

5 Conclusions

We have estimated the basic reproduction number R_0 using the more general non-Markovian framework, that besides being known from the emergence of mathematical epidemiology, has not been widely applied. The approach was used to determine the value of R_0 in six major countries in Europe during the first wave of the COVID-19 epidemic. The onset of infectiousness, instead of starting immediately after contraction of the pathogen, was taken to be related to the onset of symptoms, for which the empirical evidence suggests that is not distributed exponentially as the Markovian assumption implies. The incubation period distribution was further combined with an infectiousness window function which was considered to have short period—1 or 2 days before and finish up to 2 days after the onset of symptoms. From both functions, we have constructed the generation interval distribution that uniquely determines R_0. In all scenarios we have considered, the calculated value for R_0 was obtained to exceed the one from the classical relationship R_0^c. This suggests that the calculations with the classical, Markovian approach, should be taken rather cautiously.

Better estimates of R_0 would be obtained with empirical function of the infectiousness window, or direct estimation of the shape of the generation interval distribution. This needs more involving epidemiological tracing, that is not an easy task. However, since the calculations are strongly dependent on the generation interval distribution, we believe that it will lead to more intensive work for gathering epidemiological data and also increase the awareness that to this non-Markovian setting in the epidemic models should be given more attention.

References

1. K. Dietz, J.A.P. Heesterbeek, Math. Biosci. **180**(1–2), 1–21 (2002)
2. R. Ross, Proc. Math. Phys. Eng. Sci. **92**(638), 204–230 (1916)
3. R. Ross, H.P. Hudson, Proc. Math. Phys. Eng. Sci. **93**(650), 212–225 (1917)
4. R. Ross, H.P. Hudson, Proc. Math. Phys. Eng. Sci. **93**(650), 225–240 (1917)
5. W.O. Kermack, A.G. McKendrick, Proc. Math. Phys. Eng. Sci. **115**(772), 700–721 (1927)
6. C. Moore, M.E.J. Newman, Phys. Rev. E **61**, 5678–5682 (2000)
7. R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. **86**(14), 3200–3203 (2001)
8. S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, Y. Moreno, Eur. Phys. Lett. **89**(3), 38009 (2010)
9. G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Nat. Med. **26**(6), 855–860 (2020)
10. A. Vespignani, H. Tian, C. Dye, J.O. Lloyd-Smith, R.M. Eggo, M. Shrestha, S.V. Scarpino, B. Gutierrez, M.U.G. Kraemer, J. Wu, K. Leung, G.M. Leung, Nat. Rev. Phys. **2**(6), 279–281 (2020)
11. S.S. Musa, S. Qureshi, S. Zhao, A. Yusuf, U.T. Mustapha, D. He, Infect. Dis. Model. **6**, 448–460 (2021)
12. Y. Tang, T.D.A. Serdan, A.L. Alecrim, D.R. Souza, B.R.M. Nacano, F.L.R. Silva, E.B. Silva, S.O. Poma, M. Gennari-Felipe, P.N. Iser-Bem, L.N. Masi, S. Tang, A.C. Levada-Pires, E. Hatanaka, M.F. Cury-Boaventura, F.T. Borges, T.C. Pithon-Curi, M.C. Curpertino, J. Fiamoncini, C.G. Leandro, R. Gorjao, R. Curi, S.M. Hirabara, Sci. Rep. **11**(1), 16400 (2021)
13. R. Gopal, V. Chandrasekar, M. Lakshmanan, Eur. Phys. J. Spec. Top. (2022). https://doi.org/10.1140/epjs/s11734-022-00426-8
14. P.S. Rana, N. Sharma, Eur. Phys. J. Spec. Top. (2022). https://doi.org/10.1140/epjs/s11734-022-00534-5
15. E.V.M. dos Reis, M.A. Savi, Eur. Phys. J. Spec. Top. **231**, 893–904 (2022)
16. S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng, H.R. Meredith, A.S. Azman, N.G. Reich, J. Lessler, Ann. Intern. Med. **172**(9), 577–582 (2020)
17. J. Qin, C. You, Q. Lin, T. Hu, S. Yu, X.-H. Zhou, Sci. Adv. **6**(33), 1202 (2020)
18. C. McAloon, A. Collins, K. Hunt, A. Barber, A.W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D.E.A. McEvoy, BMJ Open **10**(8), 039652 (2020)
19. X. He, E.H.Y. Lau, P. Wu et al., Nat. Med. **26**(5), 672–675 (2020)

20. P. Van Mieghem, R. van de Bovenkamp, Phys. Rev. Lett. **110**, 108701 (2013)
21. I.Z. Kiss, G. Röst, Z. Vizi, Phys. Rev. Lett. **115**, 078701 (2015)
22. M. Feng, S.-M. Cai, M. Tang, Y.-C. Lai, Nat. Commun. **10**(1), 3748 (2019)
23. I. Tomovski, L. Basnarkov, A. Abazi, IEEE Trans. Netw. Sci. Eng. **9**(2), 552–563 (2022)
24. L. Basnarkov, I. Tomovski, T. Sandev, L. Kocarev, Chaos Solit. Fractals (2022). https://doi.org/10.1016/j.chaos.2022.112286
25. A. Gowrisankar, T.M.C. Priyanka, S. Banerjee, Eur. Phys. J. Plus **137**, 100 (2022)
26. C. Kavitha, A. Gowrisankar, S. Banerjee, Eur. Phys. J. Plus **136**, 596 (2021)
27. A. Gowrisankar, L. Rondoni, S. Banerjee, Eur. Phys. J. Plus **135**, 526 (2020)
28. P. Fine, K. Eames, D.L. Heymann, Clin. Infect. Dis. **52**(7), 911–916 (2011)
29. J. Wallinga, M. Lipsitch, Proc. R. Soc. B: Biol. Sci. **274**(1609), 599–604 (2007)
30. Z. Du, X. Xu, Y. Wu, L. Wang, B.J. Cowling, L.A. Meyers, Emerg. Infect. Dis. **26**(6), 1341 (2020)
31. M. D'Arienzo, A. Coniglio, Biosaf. Health **2**(2), 57–59 (2020)
32. G.G. Katul, A. Mrad, S. Bonetti, G. Manoli, A.J. Parolari, PLoS One **15**(9), 0239800 (2020)
33. A.G. Neves, G. Guerrero, Phys. D: Nonlinear Phenom. **413**, 132693 (2020)
34. D. Otoo, E.K. Donkoh, J.A. Kessie, Eur. J. Appl. Math. **14**(1), 135–148 (2021)
35. D. Ghosh, P. Santra, G. Mahapatra, A. Elsonbaty, A. Elsadany, Eur. Phys. J. Spec. Top. (2022). https://doi.org/10.1140/epjs/s11734-022-00537-2
36. I. Locatelli, B. Trächsel, V. Rousson, PLoS One **16**(3), 0248731 (2021)
37. S. Zhao, Q. Lin, J. Ran, S.S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He, M.H. Wang, J. Infect. Dis. **92**, 214–217 (2020)
38. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K.S. Leung, E.H. Lau, J.Y. Wong et al., N. Engl. J. Med. **382**, 1199–1207 (2020)
39. M. Aghaali, G. Kolifarhood, R. Nikbakht, H.M. Saadati, S.S. Hashemi Nazari, Transbound Emerg. Dis. **67**(6), 2860–2868 (2020)
40. O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, Public. Newton Inst. **5**, 95–115 (1995)
41. F. Brauer, Infect. Dis. Model. **2**(2), 113–127 (2017)
42. R. Böckh, *Statistisches Jahrbuch der Stadt Berlin* (Leonh, Simion, 1877)
43. A.J. Lotka, J. Wash. Acad. Sci. **9**(3), 73–77 (1919)
44. J.A.P. Heesterbeek, Acta. Biotheor. **50**(3), 189–204 (2002)
45. A.J. Lotka, Science **26**(653), 21–22 (1907)
46. S.W. Park, D. Champredon, J. Dushoff, J.R. Soc, Interface **17**, 20190719 (2020)
47. Z. Wu, J.M. McGoogan, JAMA **323**(13), 1239–1242 (2020)
48. H.W. Hethcote, SIAM Rev. **42**(4), 599–653 (2000)
49. K. Ejima, K. S. Kim, C. Ludema, A. I. Bento, S. Iwanami, Y. Fujita, H. Ohashi, Y. Koizumi, K. Watashi, K, Aihara, H, Nishiura, S. Iwami, Epidemics **35**, 100454 (2021). https://doi.org/10.1016/j.epidem.2021.100454.https://www.sciencedirect.com/science/article/pii/S1755436521000153
50. N. Néant, G. Lingas, Q.L. Hingrat, J. Ghosn, I. Engelmann, Q. Lepiller, A. Gaymard, V. Ferré, C. Hartard, J-C. Plantier, V. Thibault, J. Marlet, B. Montes, K. Bouiller, F-X. Lescure, J-F. Timsit, E. Faure, J. Poissy, C. Chidiac, F. Raffi, A. Kimmoun, M. Etienne, J-C. Richard, P. Tattevin, D. Garot, V.L. Moing, D. Bachelet, C. Tardivon, X. Duval, Y. Yazdanpanah, F. Mentré, C. Laouénan, B. Visseaux, J. Guedj, Proc. Natl. Acad. Sci. **118**(8), e2017962118 (2021). https://doi.org/10.1073/pnas.2017962118.https://www.pnas.org/doi/abs/10.1073/pnas.2017962118