



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Undetected infectives in the Covid-19 pandemic

Maurizio Melis^{a,*}, Roberto Littera^b

^aAART (Association for the Advancement of Research on Transplantation), Cagliari, Italy

^bComplex Structure of Medical Genetics, R. Binaghi Hospital, ASL Cagliari, ATS Sardegna, Italy



ARTICLE INFO

Article history:

Received 28 September 2020

Received in revised form 4 January 2021

Accepted 4 January 2021

Keywords:

Epidemiology

Covid-19

SARS-CoV-2

SIR model

Undetected cases

ABSTRACT

Objectives: Epidemiological investigations and mathematical models have revealed that the rapid diffusion of Covid-19 can mostly be attributed to undetected infective individuals who continue to circulate and spread the disease: finding their number would be of great importance in the control of the epidemic.

Methods: The dynamics of an infection can be described by the SIR model, which divides the population into susceptible (*S*), infective *I*, and removed *R* subjects. In particular, we exploited the Kermack-McKendrick epidemic model, which can be applied when the population is much larger than the fraction of infected subjects.

Results: We proved that the fraction of undetected infectives, compared to the total number of infected subjects, is given by $1 - \frac{1}{R_0}$, where R_0 is the basic reproduction number. The mean value $R_0 = 2.10$ – 2.11 for the Covid-19 epidemic in three Italian regions yielded a percentage of undetected infectives of 52.4% (52.2%–52.6%) compared to the total number of infectives.

Conclusions: Our results, straightforwardly obtained from the SIR model, highlight the role of undetected carriers in the transmission and spread of the SARS-CoV-2 infection. Such evidence strongly recommends careful monitoring of the infective population and ongoing adjustment of preventive measures for disease control until a vaccine becomes available for most of the population.

© 2021 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

A critical issue in the control of an epidemic is to know the exact number of infective subjects. Current estimates of SARS-CoV-2 infection are significantly hampered by the difficulty to perform large-scale diagnostic tests, despite the current awareness that the spread of the Covid-19 pandemic is mostly caused by undetected carriers.

The speed at which an epidemic grows cannot be explained if we only consider the number of recorded infected patients who, supposedly, are immediately removed from the circulating population by hospitalization or isolation at home.

Undetected infectives can be classified into two categories: 1) paucisymptomatic or asymptomatic individuals, who never develop overt symptoms during the course of infection; 2) presymptomatic subjects, who will eventually develop symptoms. Undetected infectives are largely responsible for the rapid increase

of the epidemic. To reliably detect their presence, it would be necessary to test the entire population and not just the symptomatic cases.

The dynamics of an epidemic can be described by an epidemiological model known as the SIR model, which divides the whole population into three classes of subjects: susceptible (*S*), infective (*I*), and removed (*R*) individuals. Kermack and McKendrick (1933) developed a SIR model for the study of epidemics in populations much larger than the infected fraction. Under this assumption, which is fully verified in the Covid-19 epidemic, we proved that the total number of infectives, when an epidemic occurs, is approximately $R_0 \cdot R$, where $R_0 > 1$ is the basic reproduction number of the infection and *R* is the number of infectives who have been removed because of recovery, isolation, hospitalization or death. The number of undetected infectives is then $(R_0 - 1) \cdot R$. The fractions of removed and undetected infectives, in comparison to the total number of infectives, are $\frac{1}{R_0}$ and $1 - \frac{1}{R_0}$, respectively.

By applying the model described above to the data available on the Covid-19 epidemic in Italy, we calculated that the mean value of the basic reproduction number in three Italian regions (Lombardy, Emilia-Romagna, and Sardinia) was $R_0 = 2.10$ (95%

* Corresponding author.

E-mail addresses: maurizio.melis@gmail.com (M. Melis), roby.littera@gmail.com (R. Littera).

confidence interval, 2.09–2.11). Consequently, the number of undetected cases turned out to be about $R_0 - 1 = 1.1$ times the number of removed cases. More specifically, we found that the percentage of undetected infectives was about $1 - \frac{1}{R_0} = 52.4\%$ (95% confidence interval, 52.2%–52.6%) of the total number of infectives.

Previous investigations found that the percentages of asymptomatic infectives (i.e., subjects without fever, cough, or any other symptoms) were: 43.2% (32.2%–54.7%) in Vo', a small town near Padua in Italy (Lavezzo et al., 2020); 50.5% (46.5%–54.4%) onboard the Diamond Princess cruise ship in Yokohama, Japan (Mizumoto et al., 2020); 47% (38%–56%) in mainland China (Li et al., 2020b) and 52.0% (including paucisymptomatic infectives) in a large sample (64,660 subjects) of the Italian population (Italian National Institute of Statistics, 2020).

The data provided by the Italian Ministry of Health and Civil Protection Department (2020) up to the 3rd of June 2020 reported about 233,800 removed cases in Italy, including either patients hospitalized or isolated at home or recovered or dead. Based on the result found in the present study, the total number of paucisymptomatic, asymptomatic, and presymptomatic infectives had to be almost 491,000 up to that date. This means that 257,200 individuals were not diagnosed as infected, although they continued to circulate and spread the virus.

This study confirms that undetected infectives can be considered the key culprits for the rapid spread of SARS-CoV-2 within the population. Consequently, interventions to control the infection will need to be maintained until the complete disappearance of the epidemic.

Methods

In the SIR epidemic model, the population is divided into three distinct classes (Murray, 2002): the susceptible subjects, S , who can catch the disease; the undetected infectives, I , who have the disease and can transmit it; and the removed infected subjects, R , namely those with a laboratory diagnosis who are either hospitalized, isolated at home, dead or recovered.

We assume that all the individuals diagnosed as infected – either by nasopharyngeal swab or serological test – are immediately isolated, thus passing from the class of infectives I to that of the removed infectives R . On the contrary, the infected subjects without a positive diagnosis are classified as undetected infectives (U), who are either still infective (I) or infected but no longer contagious (U_0). The total number of undetected infectives U is then given by: $U = I + U_0$.

At any time t , the total number of infected subjects $I_{tot}(t)$ is the sum of the number of removed infectives $R(t)$ and undetected infectives $U(t)$: $I_{tot}(t) = R(t) + U(t)$. As discussed in the Introduction, the undetected infective individuals can be asymptomatic, paucisymptomatic, or presymptomatic.

The progression of an individual from the susceptible compartment S to the total infected class I_{tot} is represented by the scheme in Figure 1.

If $S(t)$ is the number of susceptible individuals at time t and N is the size of the population, the total number of infected subjects $I_{tot}(t)$ turns out to be

$$I_{tot}(t) = N - S(t)$$

By manipulating the differential equations which define the SIR model (Appendix A) and assuming that the initial number S_0 of susceptible individuals is close to N , i.e., $S_0 \cong N$, one obtains

$$\frac{dS}{dR} \cong -\frac{R_0}{N} \cdot S \Rightarrow S(t) \cong N \cdot e^{-R_0 \cdot R(t)/N}$$

where R_0 is the basic reproduction number (discussed in Appendix B). Under the assumption $R_0 \cdot R(t)/N \ll 1$ (a condition which is certainly verified if the population size is much larger than the number of infected subjects), we can approximate $S(t)$ in the following form:

$$S(t) \cong N \cdot \left[1 - \frac{R_0}{N} \cdot R(t) + \frac{1}{2} \left(\frac{R_0}{N} \right)^2 \cdot R(t)^2 \right]$$

The total number of infected subjects $I_{tot}(t)$ at time t then becomes

$$I_{tot}(t) = N - S(t) \cong R_0 \cdot R(t) \cdot \left[1 - \frac{R_0}{2N} \cdot R(t) \right]$$

while the undetected infectives $U(t)$ at time t turn out to be

$$U(t) = I_{tot}(t) - R(t) \cong (R_0 - 1) \cdot R(t) \cdot \left[1 - \frac{R_0^2}{2(R_0 - 1)N} \cdot R(t) \right]$$

The ratio between the removed infected subjects $R(t)$ and $I_{tot}(t)$ at time t is

$$\frac{R(t)}{I_{tot}(t)} \cong \frac{1}{R_0 \cdot \left[1 - \frac{R_0}{2N} \cdot R(t) \right]}$$

while the ratio between the undetected infectives $U(t)$ and $I_{tot}(t)$ at time t is

$$\frac{U(t)}{I_{tot}(t)} = 1 - \frac{R(t)}{I_{tot}(t)} \cong \frac{R_0 - 1}{R_0} \cdot \frac{1 - \frac{R_0^2}{2(R_0 - 1)N} \cdot R(t)}{1 - \frac{R_0}{2N} \cdot R(t)}$$

Being $R_0 \cdot R(t)/N \ll 1$, the previous four equations can be approximated as

$$I_{tot}(t) \cong R_0 \cdot R(t), \quad U(t) \cong (R_0 - 1) \cdot R(t),$$

$$\frac{R(t)}{I_{tot}(t)} \cong \frac{1}{R_0}, \quad \frac{U(t)}{I_{tot}(t)} \cong 1 - \frac{1}{R_0}$$

These results are obtained under the assumption $S_0 \cong N$, which implies $R_0 > 1$, i.e., that an epidemic ensues.

The fraction of undetected infectives U , compared to the total infectives I_{tot} , has been derived straightforwardly from the SIR epidemic model and only depends on the basic reproduction number R_0 .

Appendix A provides further details on the SIR model, Appendix B shows how the evaluation of the basic reproduction number R_0 was performed, Appendix C presents the numerical fit of the data, Appendix D shows how the effective reproduction number $R_{eff}(t)$ was computed, Appendix E describes the evaluation of the constants of the fit from the data at the peak of new infectives and Appendix F shows the application of the model to stratified groups.

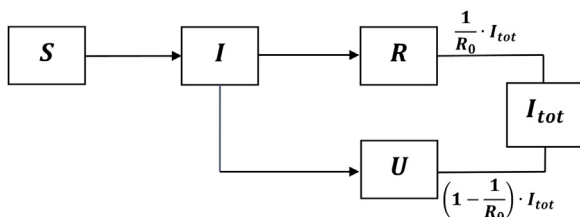


Figure 1. Scheme representing the progression of an individual from the susceptible compartment S to the total infected class I_{tot} . The fractions of I_{tot} in the R and U compartments are $\frac{1}{R_0}$ and $1 - \frac{1}{R_0}$, respectively.

Results

The data provided by the Italian Ministry of Health and Civil Protection Department (2020), updated to the 3rd of June 2020, were fitted for three Italian regions using a specific code written with Wolfram Mathematica 12.1 (Wolfram, 2020) and based on the Kermack-McKendrick model (Kermack and McKendrick, 1933).

Lombardy, in the north of Italy, has been the region with the highest number of Covid-19 infections, followed by Emilia-Romagna (at the second place from the 29th of February to the 24th of April, at the third place in the other periods of the epidemic). On the contrary, the Island of Sardinia, in the South of

Italy, was one of the regions with the lowest number of documented Covid-19 infections and deaths. The population size N in these regions, updated to the 1st of January 2019, were: Lombardy $N = 10,060,574$, Emilia-Romagna $N = 4,459,477$, Sardinia $N = 1,639,591$ (data from the Italian National Institute of Statistics).

In these three Italian regions, our epidemiological model yielded the mean value $R_0 = 2.10$ ($2.09 - 2.11$) for the basic reproduction number R_0 (Appendix B).

Time t was expressed in days since t_0 ($t = 0$), the day before the date of the first diagnosed patient: 19th of February in Lombardy, 20th of February in Emilia-Romagna, and 2nd of March in Sardinia.

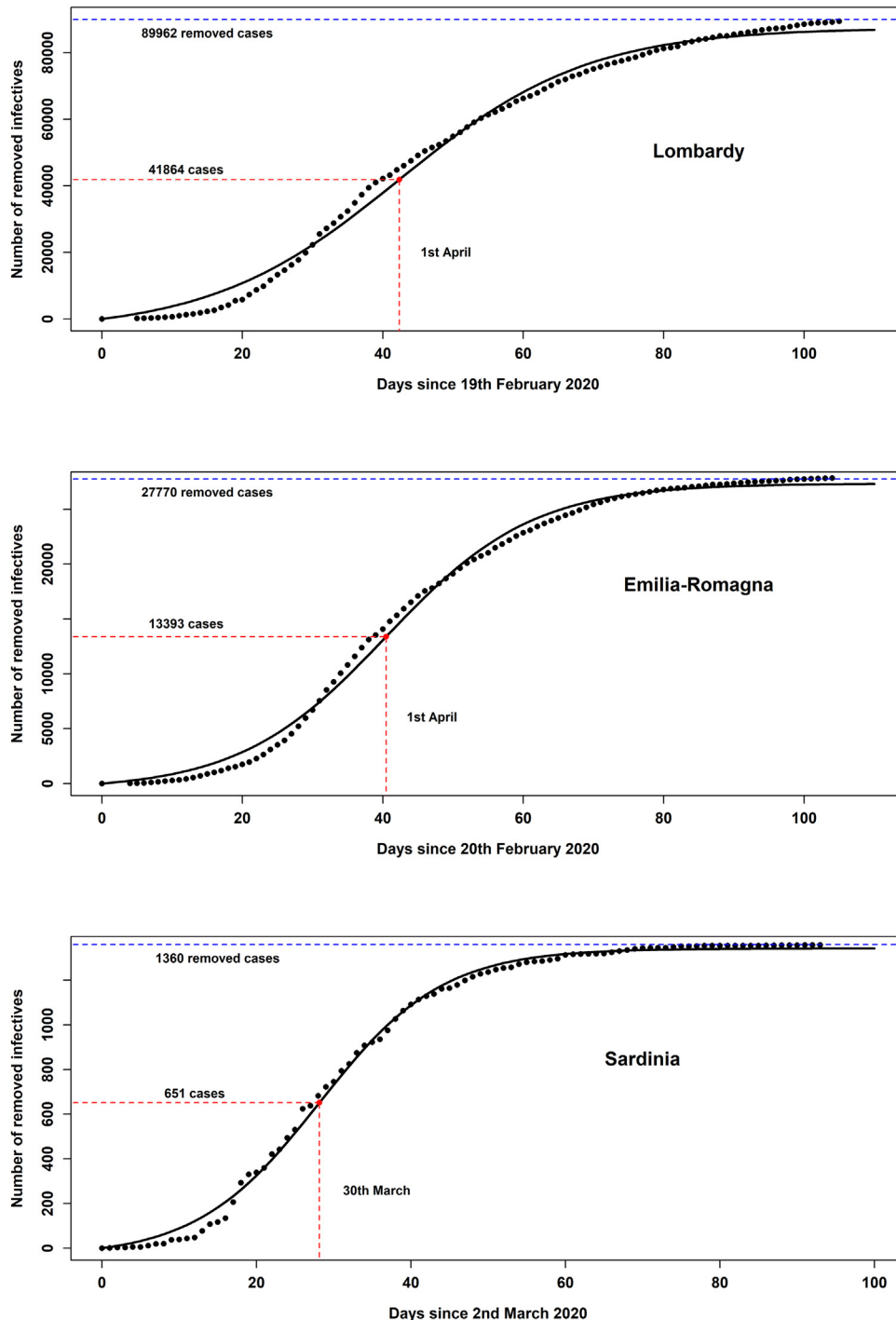


Figure 2. The fit of the number of removed infectives $R(t)$ according to the Kermack-McKendrick model in three Italian regions.

At any time t , the mean percentage of removed infectives $R(t)$ in comparison to the total number of infectives $I_{tot}(t)$ was about $\frac{1}{R_0} \cong 47.6\%$ (47.4% – 47.8%), while the mean percentage of undetected infectives $U(t)$ was about $1 - \frac{1}{R_0} \cong 52.4\%$ (52.2% – 52.6%).

Based on the data provided by the Italian Ministry of Health and Civil Protection Department (2020), Figure 2 represents the number of removed infectives $R(t)$ in Lombardy, Emilia-Romagna, and Sardinia, fitted by the equation $R(t) = c_1 \cdot [\tanh(c_2 t - c_3) + \tanh(c_3)]$. Further details are discussed in Appendix C.

In Appendix D, the Kermack-McKendrick model was also used to compute the effective reproduction number $R_{eff}(t)$ and to evaluate the time corresponding to the threshold $R_{eff}(t) = 1$ at which the epidemic starts to decline.

Table 1 reports the main epidemic parameters of the Covid-19 epidemic in Lombardy, Emilia-Romagna, and Sardinia: the basic reproduction number R_0 , the final numbers (for $t \rightarrow \infty$) of the removed (R), undetected (U) and total I_{tot} infectives, the percentages U/I_{tot} and R/I_{tot} , the day t_0 when the epidemic started, the time t_{peak} (both in days, since t_0 , and according to calendar date) of the maximum rate $R'(t_{peak})$ of new cases per day, with the corresponding number of removed infectives $R(t_{peak})$, the constants c_1, c_2, c_3 in the equation $R(t) = c_1 \cdot [\tanh(c_2 t - c_3) + \tanh(c_3)]$, determined by fitting the data on the Covid-19 epidemic with Wolfram Mathematica 12.1 (Wolfram, 2020).

The validity of our model was tested by determining the constants c_1, c_2, c_3 of the fit from the data on $R(t_{peak})$ and $R'(t_{peak})$, as discussed in Appendix E. The results obtained only differ by a maximum of 3% from the values in Table 1, thereby confirming that our model provides reliable estimates of the main epidemic parameters.

Figure 3 shows the number of newly recorded infectives per day in Lombardy, Emilia-Romagna, and Sardinia. These curves plot the equation $\frac{dR}{dt} = c_1 \cdot c_2 \cdot [\text{sech}(c_2 t - c_3)]^2$ (Appendix C), which yields the rate of newly removed infectives in the Kermack-McKendrick model.

Figure 4 compares the percentages of asymptomatic infectives found in three previous investigations, conducted in Vo' (Lavezzo et al., 2020), Japan (Mizumoto et al., 2020), and China (Li et al., 2020b), with the percentage of undetected infectives in Lombardy, Emilia-Romagna and Sardinia obtained in this study through the SIR model.

The serological investigation conducted in Italy on 64,660 subjects from the 15th of May to the 15th of July 2020 revealed that the percentage of paucisymptomatic infectives was 24.7% and that of asymptomatic infectives was 27.3%. Therefore, the total

percentage of paucisymptomatic and asymptomatic infectives was 52.0%, as discussed in the preliminary report released by the Italian National Institute of Statistics (2020).

The result obtained with the SIR model (shown in Figure 4) seems to be affected by a relatively small error compared to the errors of other studies. The reason is that the 95% confidence interval associated with our finding only represents the uncertainty intrinsic to the mathematical model, excluding the error in the data provided by the Italian Ministry of Health and Civil Protection Department (2020) for removed infectives $R(t)$ at time t . These data were probably underestimated because of the difficulty to administer swabs or serological tests to all the suspect cases or even to subjects with overt symptoms. However, we only considered the errors associated with the statistical goodness of fit in our model, being unable to evaluate the uncertainty of the data on removed infectives.

Figure 5 shows the plots against time t of the number of removed, undetected and total infectives in three Italian regions according to the Kermack-McKendrick model, i.e., $R(t) = c_1 \cdot [\tanh(c_2 t - c_3) + \tanh(c_3)]$ (with c_1, c_2, c_3 given in Table 1), $U(t) = (R_0 - 1) \cdot R(t)$ and $I_{tot}(t) = R_0 \cdot R(t)$, respectively.

The assumption that the population size must be larger than the number of infected subjects corresponds to an approximated relative error $\frac{R_0 \cdot R(t)}{2N}$ on the undetected fraction of infectives, i.e., a percent error lower than 0.9% in Lombardy, 0.6% in Emilia-Romagna, and 0.1% in Sardinia.

Our model can be applied to a sample of infected individuals stratified into two groups by a specific characteristic, e.g., age or gender (Appendix F). The stratification of the undetected infectives U in the two groups turns out to approach that of the whole sample of infected subjects I_{tot} , i.e., $\frac{U_1}{U_2} \cong \frac{I_{tot,1}}{I_{tot,2}}$. This conclusion was confirmed by exploiting the data of the investigation conducted in Vo' (Lavezzo et al., 2020).

Discussion

The speed at which an infection spreads is strongly influenced by the number of undetected infected individuals who contribute to disseminate the virus without being diagnosed as positive. This study proved that in any epidemic the fraction of undetected infectives, compared to the total number of infections, is given by the approximated expression $1 - \frac{1}{R_0}$, which only depends on the basic reproduction number R_0 .

The analytical expression of R_0 found in Appendix B was exploited to compute the basic reproduction number in three Italian regions (Lombardy, Emilia-Romagna, and Sardinia); the

Table 1
Main epidemic parameters from the fit of the removed infectives $R(t)$ in Lombardy, Emilia-Romagna, and Sardinia. The 95% confidence intervals are reported between brackets.

Parameters	Lombardy	Emilia-Romagna	Sardinia
R_0	2.07 (2.06–2.08)	2.10 (2.09–2.11)	2.13 (2.12–2.14)
$R(t \rightarrow \infty)$	87472 (84982–89962)	27348 (26926–27770)	1341 (1323–1360)
$U(t \rightarrow \infty)$	92036 (88976–95097)	29619 (29022–30216)	1508 (1479–1537)
$I_{tot}(t \rightarrow \infty)$	179508 (173957–185059)	56967 (55948–57986)	2849 (2802–2897)
U/I_{tot} (%)	51.3% (51.1%–51.4%)	52.0% (51.9%–52.1%)	52.9% (52.8%–53.1%)
R/I_{tot} (%)	48.7% (48.6%–48.9%)	48.0% (47.9%–48.1%)	47.1% (46.9%–47.2%)
t_0 (date)	19th February 2020	20th February 2020	2nd March 2020
t_{peak} (days)	42.3 (37.2–47.5)	40.5 (36.7–44.2)	28.2 (25.8–30.5)
t_{peak} (date)	1 Apr (27 Mar–7 Apr)	1 Apr (28 Mar–4 Apr)	30 Mar (28 Mar–2 Apr)
$R'(t_{peak})$	1697 (1560–1835)	669 (631–708)	43 (41–46)
$R(t_{peak})$	41864 (40308–43421)	13393 (13135–13650)	651 (639–663)
c_1	4.561 (4.467–4.654)·10 ⁴	1.396 (1.379–1.412)·10 ⁴	6.901 (6.837–6.964)·10 ²
c_2	0.037 (0.035 – 0.039)	0.048 (0.046 – 0.050)	0.063 (0.060 – 0.066)
c_3	1.576 (1.478–1.673)	1.942 (1.852–2.032)	1.772 (1.694–1.850)

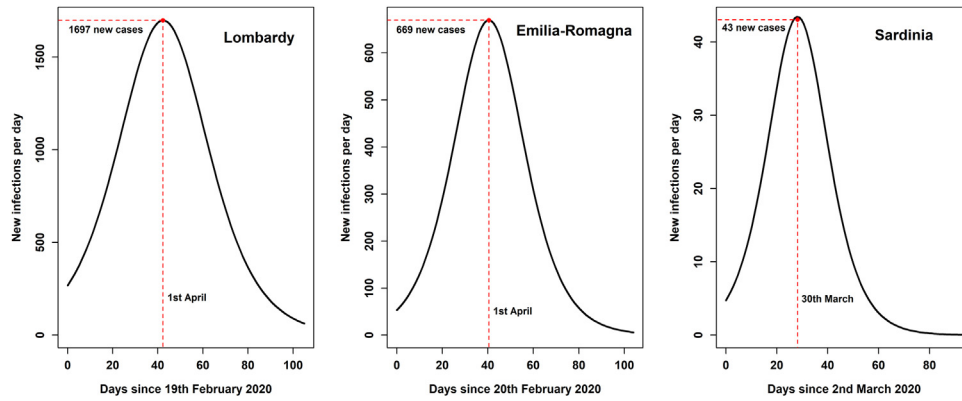


Figure 3. Rate $\frac{dR}{dt}$ of newly removed infectives per day according to the Kermack-McKendrick model in three Italian regions.

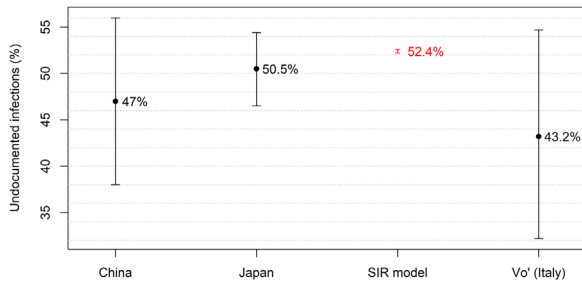


Figure 4. Comparison between the percentage of undetected infectives obtained using the SIR model and the percentages of asymptomatic infectives in three previous investigations conducted in China, Japan, and Vo' (Italy). The error bars represent the 95% confidence intervals.

corresponding mean value $R_0 = 2.10$ (95% confidence interval, 2.09–2.11) overlaps well with the result $R_0 = 2.2$ (1.4 – 3.9) found in China (Li et al., 2020a) and the result $R_0 = 2.28$ (2.06–2.52) obtained in Japan from the data collected onboard a cruise ship (Zhang et al., 2020).

In Appendix D, the Kermack-McKendrick model was also used to compute the effective reproduction number $R_{eff}(t)$, as previously defined, e.g., by Nishiura and Chowell (2009).

By exploiting the aforesaid mean value of R_0 , we found that the percentage of undetected infectives was $1 - \frac{1}{R_0} \approx 52.4\%$ (95% confidence interval, 52.2%–52.6%) of the total infectives.

The assumption that the population size must be larger than the number of infected subjects corresponds to a percent error lower than 1% on the undetected fraction of infectives.

As shown in Figure 4, the percentage of undetected infectives obtained in this study overlaps well with the percentages of asymptomatic infectives found in previous investigations (Lavezzo et al., 2020; Mizumoto et al., 2020; Li et al., 2020b), confirming that the fraction of undetected infectives is considerable and is likely to have a strong influence on the dynamics of the epidemic.

In a study conducted in Vo' (Lavezzo et al., 2020), a small town in Veneto (Italy), most inhabitants were tested through nasopharyngeal swabs in two consecutive surveys; the mean percentage of asymptomatic infectives corresponded to 43.2% (32.2%–54.7%) of the total SARS-CoV-2 infections. Notably, the nasopharyngeal swabs performed in this study showed no statistically significant differences between the viral load of symptomatic and asymptomatic infections. Moreover, the viral load tended to peak in the large majority of participants around the day of symptom onset, suggesting an essential role for presymptomatic transmission in the spread of the virus. Therefore, both asymptomatic and presymptomatic transmission represent a major threat to epidemiologic control and containment of the infection (Lavezzo et al., 2020).

An investigation performed on the passengers of the Diamond Princess (Mizumoto et al., 2020), a cruise ship in Yokohama (Japan), revealed that from the start of the epidemic the percentage of asymptomatic infectives on board the ship was 50.5% (46.5%–54.4%) of the total infectives.

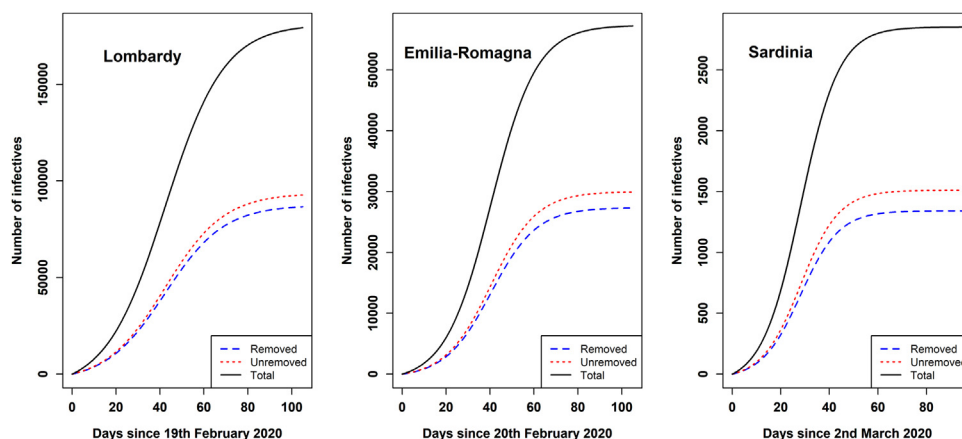


Figure 5. Plots against time t of the number of removed, undetected and total infectives in Lombardy, Emilia-Romagna, and Sardinia according to the Kermack-McKendrick model.

One of the first studies (Li et al., 2020b) to reveal the crucial role of undetected infections in the Covid-19 pandemic estimated the undetected fraction of infectives on the basis of a mathematical model connecting mobility data and observations of reported infections within China. The percentage of undetected infectives turned out to be $U = 86.2\%$ ($81.6\%–89.8\%$) of the total number of positive cases. However, in this study the transmission rate of undetected infectives was assumed to be $\mu = 55\%$ ($46\%–62\%$) of the transmission rate of symptomatic infectives. On the contrary, we assumed that all infected subjects – with or without symptoms – may present an elevated viral load and transmit the virus at the same rate, as confirmed by the investigation in Vo' (Lavezzo et al., 2020) and the systematic literature search conducted by Walsh et al. (2020). Under this assumption, the effective percentage U_{eff} of undetected infectives is given by $U_{\text{eff}} = \mu \cdot U = 47\%$ ($38\%–56\%$).

Another study (Yusef et al., 2020) investigated 350 attendees of a wedding in Jordan, 76 of whom tested positive for SARS-CoV-2. Among them, 36 individuals were asymptomatic, i.e., 47.4% ($35.8\%–59.2\%$) of the total number of infected subjects.

The studies by Lavezzo et al. (2020), Mizumoto et al. (2020) and Yusef et al. (2020) were based on laboratory tests performed in small communities (the inhabitants of Vo' in Italy, the passengers of a cruise ship in Japan, and the attendees of a wedding in Jordan, respectively) where the Covid-19 infection had spread. On the contrary, the study in China (Li et al., 2020b) was based on a mathematical model comparing mobility data and infection diffusion in mainland China after the start of the Covid-19 epidemic.

A serological investigation in the Italian population conducted by the Italian National Institute of Statistics (2020) on 64,660 subjects revealed that the percentage of paucisymptomatic and asymptomatic infectives up to mid-July 2020 was 52.0%.

A Review by Oran and Topol (2020) of the available evidence on asymptomatic SARS-CoV-2 infectives found that asymptomatic subjects accounted for approximately 40%–45% of the total number of infections and could transmit the virus to others. The authors of the Review also pointed out that the high frequency of asymptomatic infections could at least partly explain the rapid spread of the virus, since infected subjects who feel and look well are likely to have more social interaction compared to symptomatic infectives.

The results obtained in the previous investigations (Lavezzo et al., 2020; Mizumoto et al., 2020; Yusef et al., 2020; Oran and Topol, 2020) concerned asymptomatic infected subjects, while the results found in our study included all the undetected infectives, both asymptomatic subjects, and paucisymptomatic or presymptomatic individuals. This can explain why the percentages of asymptomatic infected subjects in those studies turned out to be a bit lower than the percentage we found for all the undetected infectives.

The 95% confidence intervals of the epidemiological parameters reported in Table 1 were only associated with the error intrinsic to the mathematical model used in this study, while the uncertainty of the data concerning removed infectives was not included, although the recorded positive cases were probably underestimated as a consequence of the low efficiency in administering swabs and serological tests to the population in most Italian regions.

Conclusions

Our derivation of the percentage of undetected infectives only relied on the SIR model, a cornerstone in the study of infectious disease dynamics. Despite its simplicity, the SIR model describes the global dynamics of an epidemic and allows for the evaluation of

several epidemiological parameters. However, more complex and realistic generalizations of the SIR model could be introduced to further refine and improve the true picture of an epidemic.

The general expression of the percentage of undetected infectives found in this study only requires knowledge of the basic reproduction number R_0 . Other methods involve numerous variables in order to achieve a more accurate description of an epidemic. However, these methods require specific assumptions about unknown parameters of the underlying mathematical framework.

The main conclusion which can be drawn from the results obtained in this study is that undetected infections play a crucial role in the transmission of SARS-CoV-2. The high percentage of undetected infections poses a major challenge for the control of Covid-19 and highlights the necessity to carefully monitor and adjust social distancing and other preventive measures until a sufficiently high vaccination rate is reached.

Authors' contributions

The authors contributed equally to the article.

Conflicts of interest

None declare.

Funding

None declare.

Ethical approval

Not applicable.

Acknowledgments

The authors are grateful to Anna Maria Koopmans for translations, professional writing assistance, and manuscript preparation.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ijid.2021.01.010>.

References

- Italian Ministry of Health and Civil Protection Department. Data about the Covid-19 Epidemic in Italian Regions. 2020. <https://github.com/pcm-dpc/COVID-19/tree/master/schede-riepilogative/regioni>.
- Italian National Institute of Statistics. Preliminary Results of the Investigation on Sars-CoV-2 Seroprevalence. 2020. <https://www.istat.it/it/files/2020/08/ReportPrimiRisultatiIndagineSiero.pdf>.
- Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics. Proc R Soc Lond A 1933;141:94–122.
- Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Luisa Barzon L, Del Vecchio C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. Nature 2020;584:425–9. doi:<http://dx.doi.org/10.1038/s41586-020-2488-1>.
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020a;382:1199–207. doi:<http://dx.doi.org/10.1056/NEJMoa2001316>.
- Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020b;368(6490):489–93. doi:<http://dx.doi.org/10.1126/science.abb3221> Epub 2020 Mar 16.
- Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill 2020;25(10). doi:<http://dx.doi.org/10.2807/1560-7917.pii=2000180>.
- Murray JD. Mathematical Biology. I: An Introduction. Third edition New York: Springer-Verlag; 2002.

- Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In: Chowell G, Hyman JM, Bettencourt LMA, Castillo-Chavez C, editors. *Mathematical and Statistical Estimation Approaches in Epidemiology*. Dordrecht: Springer; 2009, doi:http://dx.doi.org/10.1007/978-90-481-2313-1_5.
- Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection. A narrative review. *Ann Intern Med* 2020;(June), doi:<http://dx.doi.org/10.7326/M20-3012>.
- Walsh KA, Jordan K, Clyne B, Rohde D, Drummond L, Byrne P, et al. SARS-CoV-2 detection, viral load and infectivity over the course of an infection. *J Infect* 2020;8:357–71, doi:<http://dx.doi.org/10.1016/j.jinf.2020.06.067>.
- Wolfram Research, Inc. *Mathematica 12.1 (Trial Version)* Champaign, Illinois, US. 2020.
- Yusef D, Hayajneh W, Awad S, Momany S, Khassawneh B, Samrah S, et al. Large outbreak of coronavirus disease among wedding attendees, Jordan. *Emerg Infect Dis* 2020;(September), doi:<http://dx.doi.org/10.3201/eid2609.201469> [Online Publication Date: 20 May 2020].
- Zhanga S, Diaob MY, Yuc W, Peic L, Lind Z, Chena D. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: a data-driven analysis. *Int J Infect Dis* 2020;93:201–4, doi:<http://dx.doi.org/10.1016/j.ijid.2020.02.033>.