Modelling the SARS-CoV-2 epidemic in Italy making use of the Istat seroprevalence survey

Marco Claudio Traini [1], Carla Caponi [2], Riccardo Ferrari [3], Giuseppe Vittorio De Socio [4]

Abstract

On 3rd August 2020, the Italian National Institute of Statistics – Istat published the preliminary results of the seroprevalence survey on the percentage of individuals affected by COVID-19.
This survey aims at defining (within the entire population of Italy) the portion of individuals who developed an antibody response against SARS-CoV-2. For the first time an estimate of the asymptomatic infected population is available so as to acknowledge its potential role in the infection spread in Italy, one of the most affected countries in Europe.
The information obtained allows a particularly sensitive validation of epidemiological models which include the asymptomatic class.
The present study is devoted to the construction of a model able to simulate, in a systematic way, the asymptomatic group whose relevance in the SARS-CoV-2 epidemic has been recently discussed. The investigation involves the description of the first epidemic outbreak in Italy as well as the predictive analysis of the ongoing second wave. In particular, the possible correction to the data of the serological tests because of their sensitivity and specificity.

Keywords: SARS-CoV-2, asymptomatic group, seroprevalence, infection spread, epidemiological models, sensitivity and specificity of the serological tests, vaccination strategies.

1 Full Professor of theoretical physics; Università degli Studi di Trento, Dipartimento di Fisica, Trento, Italy. Email: marcoclaudio.traini@unitn.it.

2 Physician, MD; Clinica Geriatrica, Azienda Ospedaliero-Universitaria, Perugia, Italy. Email: carla.caponi@ospedale.perugia.it.

3 Manager and Physicist; Bilubah LLC, Sheridan, WY 82801, U.S.A. Email: riccardo.ferrari@bilubah.com.

4 Physician, MD; Clinica Malattie Infettive, Azienda Ospedaliero-Universitaria, Perugia, Italy. Email: giuseppe.desocio@ospedale.perugia.it.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction[5]

Traditionally, epidemiological models have grouped people into two, three or four groups (compartments), usually denoted by Susceptible (*S*), Exposed (*E*), Infected (*I*), and Recovered (*R*). Contact between a member of the infected group (*I*) and another person belonging to the susceptible group (*S*) leads to the latter person becoming infected with a certain probability. Depending on the model, the susceptible person either becomes infected straightaway (SIR model), or enters an intermediate stage called Exposed (*E*) (SEIR model). In the latter scenario, it is assumed that contact between persons belonging to the (*E*) and (*S*) groups does not lead to fresh infections, because members of the (*E*) group do not carry a sufficient viral load to infect others through contact.

However, one of the characteristic features of the coronavirus pandemic is that many of the persons who contract the disease are "asymptomatic" (or group A). In fact the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) throughout the world has been extremely rapid suggesting the hypotheses of a crucial role played by these infected persons who remain asymptomatic even if contagious (Buitrago-Garcia, 2020). A recent paper (Oran, 2020) summarises the available evidences on asymptomatic SARS-CoV-2 infection concluding that "*asymptomatic persons seem to account for approximately 40% to 45% of the SARS-CoV-2 infections, and that they can transmit the virus to others for an extended period of time, perhaps longer than 14 days*". Oran and Topol (Oran, 2020) conclude with the need "t*hat testing programs include those without symptoms*". Other studies estimate the fraction of asymptomatic patients to be more than 50% (Mizumoto, 2020) or as high as 75% (Day, 2020). For this reason, asymptomatic patients remain "hidden" and cannot be identified except through tests of large portions of the population. An example is given in the study by Lavezzo, Franchin, Ciavarella, *et al.* (Lavezzo, 2020) reporting results of a detailed survey in the small town of Vo' (near Padua, Italy) where on 21st February 2020 a lockdown was imposed in the whole municipality as a first outbreak in Italy. After the lockdown they found a prevalence of 1.2% (95% CI:0.8-1.8%). Notably, 42.5% (95% CI:31.5-54.6%) of the confirmed SARS-CoV-2 infections detected across two surveys were asymptomatic (that is did not have symptoms

5 The authors thank Francesca Greselin of the Department of Statistics and Quantitative Methods - University of Milano Bicocca, Italy - for her interest and help in the elaboration of the model used.

at the time of swab testing and did not develop symptoms afterwards). After that a number of studies and projects have been developed to identify the role of asymptomatic in the pandemic infection in various countries (*e.g.* Buitrago-Garcia, 2020; Guerriero, 2020; Peterson, 2020; Pollán, 2020; Snoeck, 2020; Well, 2020; Yanes-Lane, 2020).

Table 1 - Preliminary results of the seroprevalence survey in Italy (a)

Region	IgG positive (95% CI)		Absolute value	Reported cases
	lower limit	upper limit		
Italy	2.3	2.6	1,482,377	244,708

Source: Istat, 2020

(a) The amount of IgG positive tests corresponds to an infected population of (almost) 1.5 million of individuals, when the official reported cases are 244,708 till 27th July 2020.

In August 2020 preliminary results of a seroprevalence survey have been presented for a set of 64,660 persons in Italy (Istat, 2020). The Istat survey aimed at defining (within the entire population) the portion of persons that developed an antibody response against SARS-CoV-2. The survey adopted a methodology allowing the evaluation of the seroprevalence in the population also estimating the fraction of asymptomatic (or subclinical) infections and the differences for age groups, gender, localisation etc. The results are still preliminary because involve a restricted number of tests, in particular those ones whose results have been reported before 27th July 2020. The post-stratified techniques adopted[6] (Little, 1993) allow the production of statistical estimates coherent with epidemic data (both at the international and local level). For the first time there is an estimate of the asymptomatic infected population and its influence in the infection spread in Italy, one of the most affected countries in Europe.

The results of the Istat survey, enlarged to the entire Italian country, are summarised in Table 1. The analysis concludes that 1,482,377 individuals developed IgG antibodies against SARS-CoV-2. A level of seroprevalence of 2.5% (95% CI: 2.3-2.6%) to be compared with 244,708 officially reported cases.

6 Due to intentional and/or unintentional processes the characteristics of a sample may not represent the characteristics of the population of interest. To mitigate this potential bias survey researchers post-stratify base probability of selection survey weights so that sample characteristics match population control totals.

Asymptomatic patients (*A*) differ from exposed patients (*E*) in one important respect. Unlike in traditional epidemiological models, contact between a person in the (*A*) group and another in the (*S*) group does lead to the latter getting infected, with a certain probability. In addition, as in other models, contact between a person in the (*I*) group and another in the (*S*) group also leads to the latter getting infected, with a similar probability. The seminal paper by Robinson and Stilianakis (Robinson, 2013) formulates and analyses a preliminary model that captures the asymptomatic phenomenon (*A*) and can be called SAIR, since the corresponding model including exposed (*E*) group is usually called SEIR. The two models differ in many aspects from a mathematical point of view; see Ansumali, *et al.* for a recent analysis and detailed discussion (Ansumali, 2020).

Moreover it is possible to develop more refined models of the pandemic by introducing additional categories such as Quarantined, Healed, Ailing, Recognised (or Detected), Threatened, etc. (*e.g.* Park, 2020; Giordano, 2020). By introducing more categories, the result is a more realistic model of the disease progression. On the other hand, the number of parameters to be estimated increases drastically. The ideal trade-off between these two conflicting considerations remains and has to be considered on a case by case basis (Jia, 2020*a*; Kinoshita, 2020; Yu, 2020).

The aim of the present paper is the study of asymptomatic compartment population during the SARS-CoV-2 epidemic event in Italy taking advantage of the Istat survey. The model used assumes *eight* compartments (groups) including symptomatic detected and undetected, quarantined, asymptomatic, threatened (hospitalised) and recovered, a model recently proposed to discuss the role of measures against the Italian outbreak (Traini, 2020).

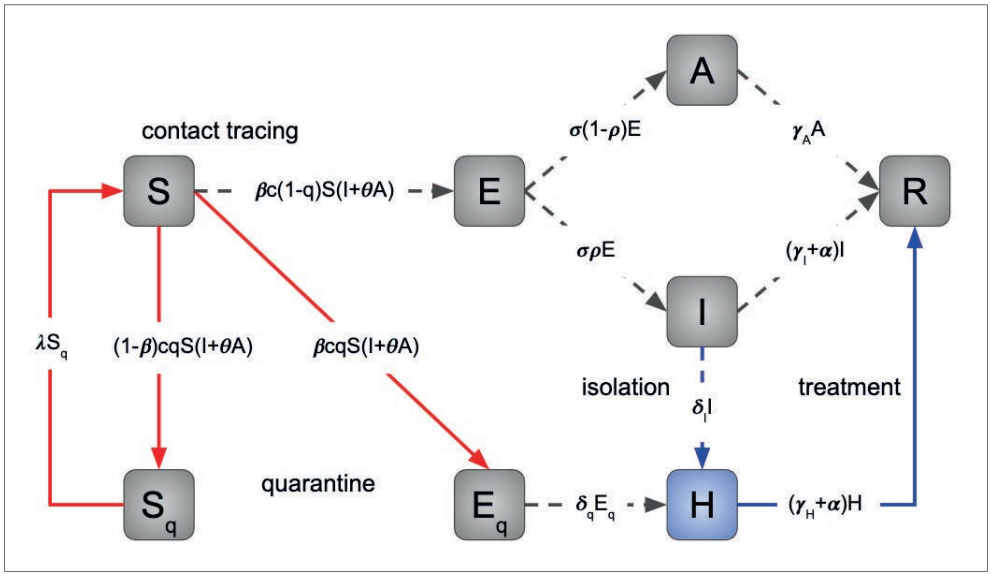
2. Methods and analysis

2.1 A time dependent quarantined model with isolation

The model we use in order to describe the time evolution of the Italian outbreak is an epidemiological model originally proposed by Tang *et al.* in order to study the Wuhan event (Tang, 2020. *J. Clin Med.*; Tang, 2020. *Infect. Dis. Model*). This model incorporates appropriate compartments relevant to intervention such as quarantine, isolation, and treatment. The population is stratified in Susceptible (S), exposed (E), asymptomatic infected individuals (A), infected with symptoms (I), hospitalised in a large sense (detected infected) (H), and recovered (R). Further stratification includes quarantined susceptible (S_q), and isolated exposed (E_q) compartments (see Figure 1) which describe the tracing procedures.

A portion of susceptible, S, get in contact with infected individuals with rate $c\ (I + \theta A)$, where c is the contact rate, I the number of symptomatic infected individuals, A the asymptomatic infected individuals, and θ the contribution of asymptomatic infected to the infection spread. With contact tracing, a proportion, q, of susceptible, S, that get in contact with infected individuals is quarantined. Individuals receive the virus at rate β, which is the transmission probability, and become exposed. On the other side, the exposed individuals identified with contact tracing get quarantined at rate q. Therefore, we have three fluxes of individuals out of S: the quarantined with virus transmission going into E_q, $\beta cqS(I + \theta A)$; the quarantined without virus transmission going into S_q, $(1 - \beta)cqS(I + \theta A)$; the individuals with virus transmission but not identified and not quarantined going into E, $\beta c(1 - q)S(I + \theta A)$. Quarantined individuals without virus transmission are released at rate λ, generating an inbound flux of individuals to S given by λS_q. Exposed, infected, and quarantined individuals move to the hospitalised compartment at rate $\delta_q E_q$. Exposed, infected, and not quarantined individuals become infectious at rate σ. Some of them develop symptoms with a probability of ρ. Then, there are two outbound fluxes of individuals for the compartment E: the infected with symptoms, $\sigma\rho E$; the infected asymptomatic $\sigma(1 - \rho)E$. The infected with symptoms will eventually be detected and hospitalised with a rate of δ_I, which reflects the sanitary system's diagnostic capability. Finally, all the infected

Figure 1 - Diagram of the model simulating the novel Coronavirus (Sars-CoV-2) infection in Italy (a)



Source: Tang, 2020

(a) The population is stratified in Susceptible (S), exposed (E), asymptomatic infected individuals (A), infected with symptoms (I), detected infected (hospitalised in a large sense) (H) and recovered (R), quarantined susceptible (Sq), isolated exposed (Eq) compartments. Interventions like intensive contact tracing followed by quarantine and isolation are indicated.

will recover with rate: γ_A, for the asymptomatic, γ_I for the infected not hospitalised, γ_H for the infected hospitalised. The infected with symptoms, both hospitalised or not, have a mortality rate of α. For disease transmission, these individuals pass to the recovered compartment, as they are no more infectious. Summing the inbound and outbound fluxes at each compartment, we obtain the system of differential equations of the model:

$$\frac{dS}{dt} = -\left[\beta c(t) + c(t) q\,(1-\beta)\right]\, S\,(I + \theta\, A) + \lambda\, S_q; \quad (1)$$

$$\frac{dE}{dt} = +\beta c(t)(1-q)\, S\,(I + \theta\, A) - \sigma\, E; \quad (2)$$

$$\frac{dI}{dt} = +\sigma\, \varrho\, E - (\delta_I(t) + \alpha + \gamma_I)\, I \quad (3)$$

$$\frac{dA}{dt} = \sigma\,(1-\varrho)\, E - \gamma_A\, A; \quad (4)$$

$$\frac{dS_q}{dt} = +c(t)q\,(1-\beta)\,S\,(I+\theta\,A) - \lambda\,S_q; \quad (5)$$

$$\frac{dE_q}{dt} = +\beta c(t)q\,S\,(I+\theta\,A) - \delta_q\,E_q; \quad (6)$$

$$\frac{dH}{dt} = \delta_I(t)\,I + \delta_q E_q - (\alpha+\gamma_H)\,H; \quad (7)$$

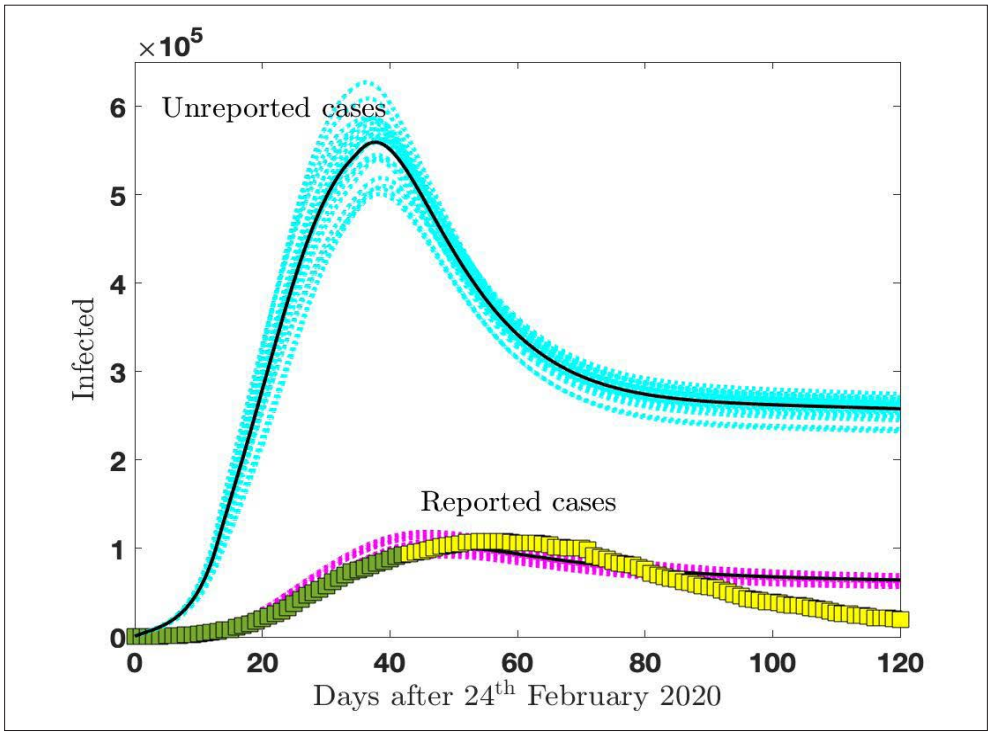
$$\frac{dR}{dt} = \gamma_I\,I + \gamma_A\,A + \gamma_H\,H; \quad (8)$$

and the values of the parameters are discussed in the next Section.

2.2 Fixing the model parameters: the outbreak

The method Markov Chain Monte Carlo (MCMC, Brooks, 2011; Hogg, 2018) is used to fit the model on the data of the outbreak in Italy in the period up to 6th April 2020. The procedure is implemented through an adaptive Metropolis-Hasting (M-H) algorithm used for four concatenated runs with 100,000 - 50,000 - 25,000 - 10,000 iterations within the MCMC toolbox for Matlab. Table 2 summarises the parameters. A peculiar aspect of the model is the time-dependence of the key parameters related to the contact rate *c* and the quarantined rate *q*. Following the control measures adopted in Italy, the flexibility of the model is such that it is possible to adapt the values of some parameters to the concrete social situation. In particular the values c_0, c_1, c_2 and q_0, q_1, q_2 are fixed on the period of the outbreak 24th February - 6th April 2020.

Figure 2 - A MCMC analysis of the infected individuals (reported and unreported cases) in Italy as a function of time in the period 24th February - 27th July 2020 (a)



Source: The data for the reported-infected individuals are by the Ministry of Health (*Ministero della Salute*, 2020), and are compared with our model predictions (magenta curves) which include statistical uncertainties

(a) The green data points refer to the data set used to fix the model parameters, while the yellow set spans the purely predicted period. The reported cases (magenta curves) are then compared with the predictions for the class of infected (asymptomatic and symptomatic) non reported cases (cyan curves) in the same period of time.

Table 2 - Parameters of the time-dependent model description

Param.	Definition [dimensions]	Estimated Value	St. Dev.	notes
$c = c_0$	Contact rate $[\mathrm{day}^{-1}]$	4.5248	0.2030	from 24th Feb. to 8th March
$c = c_1$		$c_0/2$	0.2030	from 8th March to 29th March
$c = c_2$		$1.4 \cdot c_1$	0.2030	from 29th March to 6th April
$c = c_3$		c_0	0.2030	from 6th April to 25th Dec.
$c = c_4$		$20 \cdot c_0$	0.2030	after 25th December
$c = c(t)$	see Section 2.3			to discuss the lockdown exit of 4th May
β	Probability of transmission per contact [u]	$1.5851 \cdot 10^{-8}$	$3.360 \cdot 10^{-10}$	
$q = q_0$	Quarantined rate of exposed	$1.155 \cdot 10^{-7}$	$7.720 \cdot 10^{-9}$	from 24th Feb. to 8th March
$q = q_1$	individuals [u]	$5.5 \cdot q_0$	$7.720 \cdot 10^{-9}$	from 8th March to 29th March
$q = q_2$		$2 \cdot q_1$	$7.720 \cdot 10^{-9}$	from 29th March to 6th April
$q = q_3$		$2 \cdot q_1$	$7.720 \cdot 10^{-9}$	from 6th April to 25th Sept.
$q = q_4$		$2 \cdot q_1$	$7.720 \cdot 10^{-9}$	from 25th Sept.
σ	Transition rate of exposed individuals to the infected class $[\mathrm{day}^{-1}]$	1/7		
λ	Rate at which the quarantined uninfected contacts were released into the wide community $[\mathrm{day}^{-1}]$	1/14		
ϱ	Probability of having symptoms among infected individuals [u]	0.2344	0.0060	
$\delta_I = \delta_{I_0}$	Transition rate of symptomatic infected individuals to the quarantined infected class $[\mathrm{day}^{-1}]$	0.1086	0.0010	
$\delta_I = \delta_I(t)$	see Section 2.3			to discuss the lockdown exit of 4th May 2020
δ_q	Transition rate of quarantined exposed individuals to the quarantined infected class $[\mathrm{day}^{-1}]$	0.1471	0.0010	
γ_I	Recovery rate of symptomatic infected individuals $[\mathrm{day}^{-1}]$	0.1704	0.0050	
γ_A	Recovery rate of asymptomatic infected individuals $[\mathrm{day}^{-1}]$	0.1177	0.0020	
γ_H	Recovery rate of quarantined infected individuals $[\mathrm{day}^{-1}]$	0.0515	0.0020	
α	Disease-induced death rate $[\mathrm{day}^{-1}]$	$1.2197 \cdot 10^{-5}$	$3.0 \cdot 10^{-7}$	
θ	Relative weight of asymptomatic infections [u]	0.0840	0.0010	
Initial values (24th February 2020)				
N		$6 \cdot 10^7$		
$E(0)$		21308	899	
$I(0)$		201	1.5	
$A(0)$		493	5.6	
$S_q(0)$		50000		
$E_q(0)$		522	1.3	
$H(0)$		221		
$R(0)$		8		

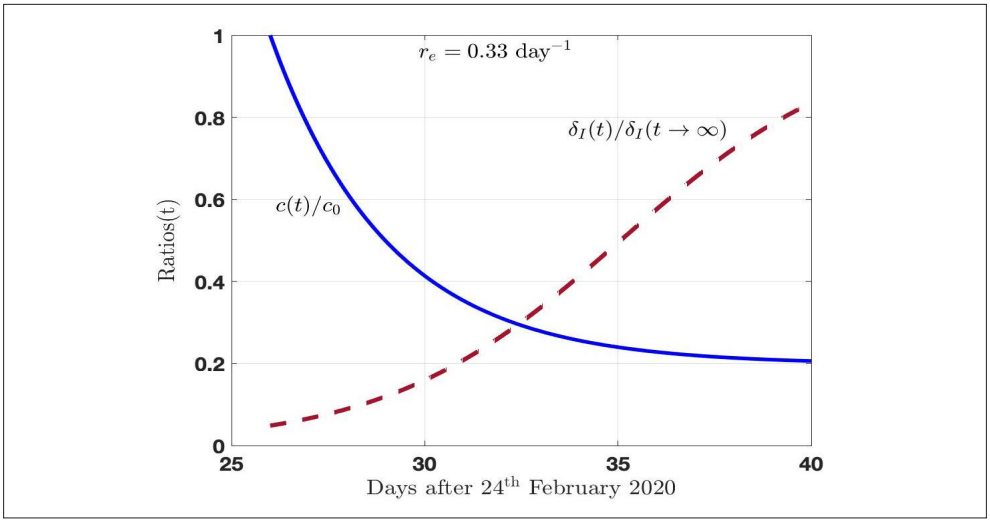
In Figure 2 some relevant predictions drawn from the model (together with some evident failures). The model is parametrised by means of the data on infected (reported cases) in the first time period (green data points in Figure 2) till 6th April 2020 (day 42). It remains consistent with data till end of May (day 90), then it deviates from data. In fact the model was designed to study possible secondary effects after 4th May 2020, when the lockdown in Italy

has been gradually mitigated (see next Section 2.3). Its longterm behaviour is discussed in detail (Traini, 2020). Here we want to emphasise an important result that can be derived from the model predictions without formulating a new parametrisation, namely the role of asymptomatic compartment as it emerges from the recent analysis by the Italian National Institute of Statistics – Istat (*Istituto Nazionale di Statistica*). As a matter of fact the model predicts also the global features of the distribution for the asymptomatic compartment illustrated in Figure 2 and discussed in Sections 3.1 and 4.2.

2.3 Predictive power: scenarios for secondary events

On 8th March 2020, the Italian Government announced the implementation of restrictions for controlling the infection. Strong control measures (like convincing all the residents to stay home and avoid contacts as much as possible) have been adopted (lockdown).

Figure 3 - Modelling the variation in time of the contact rate per person $c(t)$ $(c(t = 0) = c_0)$ **and of the rate from infected to quarantined classes** $\delta_I(t)$ **(a)**



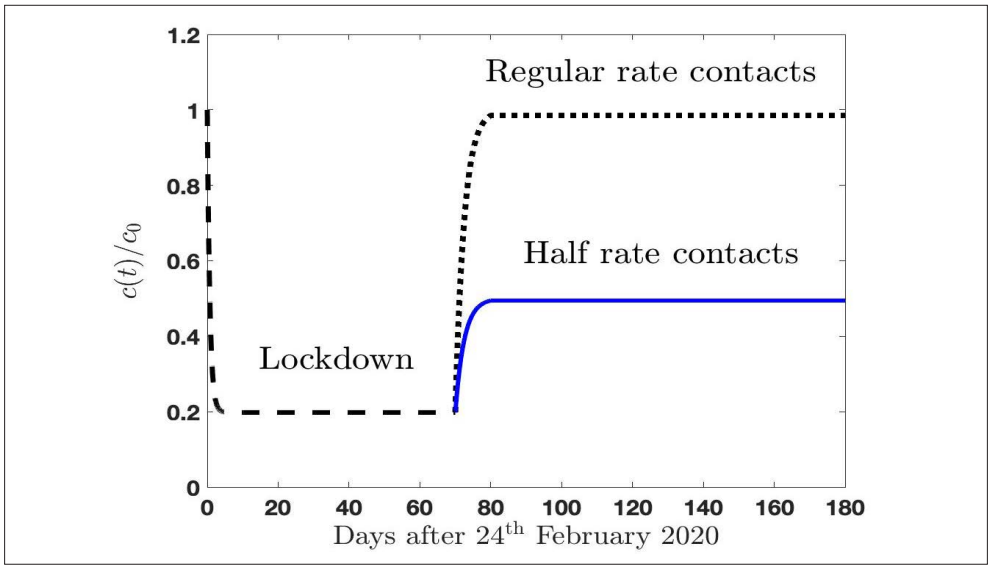
Source: Our processing
(a) See Equations (9) and (10).

From the model perspective, this can significantly contribute to decreasing the contact rate (c) among the persons (Remuzzi, 2020). Besides, the 2019-nCoV tests gradually shortened the time of diagnosis (*i.e.* the value of δ_I

increased gradually). Considering these control strategies, we could tune the model on the concrete Italian conditions.

The equations of the model, shown in Equations (1)-(8), contain parameters explicitly dependent on time. In particular the contact parameter c and the transition rate of declared infected individuals δ_I.

Figure 4 - Time behaviour of the contact rate simulating a possible secondary event in Italy at day=70, after 24th February, when the stringent measures of isolation are hypothetically relaxed (a)



Source: Our processing
(a) Two scenarios are introduced, the scenario of a rapid return to the old style of life (regular rate contacts), and a more realistic scenario where only half of the regular contacts are activated.

2.3.1 Time dependence

The initial Montecarlo analysis to fix the model parameters can be complemented with an accurate analysis of the effects of social measures or technological advances. An example is presented in Figure 3 where the time behaviour of the contact parameters and tracing rate is tuned following specific modelling related to concrete events changing the strength of the parameters involved. The time-dependence is parametrised as

$$c(t > t_0) = (c(t = t_0) - c_a)\, e^{-r_e\,(t-t_0)} + c_a, \qquad (9)$$

where $c_a/c_0 = 0.2$, $t = 0$ is fixed at 24th February and t_0 selects the initial day of a rapid implementation of the isolation measures for the entire population. At the same time the parameter δ_I, tuning the transition rate to quarantine of the infected individuals $(t_I = 1/\delta_I)$, increases because of a decreasing of the testing time:

$$\begin{aligned} & t_I(t) = [\delta_I(t > t_0)]^{-1} = \\ = \; & \left\{[\delta_I(t = t_0)]^{-1} - [\delta_I(t \gg t_0)]^{-1}\right\} e^{-r_e\,(t-t_0)} + \\ + \; & [\delta_I(t \gg t_0)]^{-1}. \end{aligned} \tag{10}$$

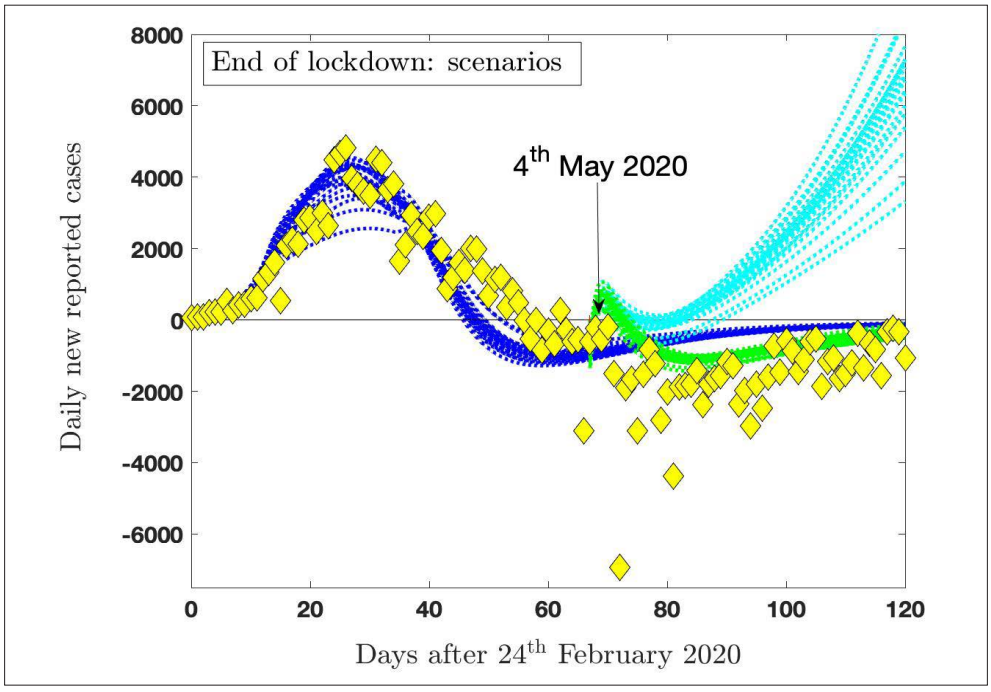
In Figure 3 the behaviour of the two parameters as a function of time. The rate of change $r_e = 0.33$ day^{-1} assumes the same values for the two gradually changing quantities. The parametrisation (10), in particular, is chosen to simulate the behaviour of the screening system to test infected (and asymptomatic) individuals.

2.3.2 Secondary events

Italy has been the first western country involved in the epidemic outbreak SARS-CoV-2. After 39 days (30th March 2020) the evolution was still in a first expanding phase. However, after almost 20 days of lockdown, the perspective of relaxing, at least partially, the social distancing measures appeared in many discussions at different levels and one can verify the predictive power of the model to simulate different scenarios of possible secondary effects. Let us assume that at a given day (4th May 2020 in the present study), the containment measures are relaxed totally or partially. The time evolution of the contact rate could be described as in Figure 4. The isolation value, as discussed in the previous Section, assumes the limiting value $c \approx 0.2$ (isolation), and at day = 70th from 24th February 2020, the isolation is interrupted and the "normality" activated. Two scenarios are introduced: i) a full return to the previous style of life (rather unlike, but it represents a reference point); ii) a more realistic scenario with half isolation.

Assuming the parametrisation of Table 2, one can try to reproduce the data on the reported cases. The good comparison between data and model can be valid for a short time, but we are not interested in reproducing the exact numbers, but the relative effects of a possible secondary event since the first

Figure 5 - The model predictions for the daily new reported cases after 24th February 2020 in Italy, compared with the official data (a)



Source: Our processing using the Italian Ministry of Health data, 2020 (*Ministero della Salute*, 2020)
(a) In addition to the fully locked scenario (blue lines), two other scenarios are included for a mitigation of the lockdown period in Italy starting from May 4th. Scenario A (cyan lines): a social contact rate somewhat similar to the regular rate (only a reduction of a factor of two is included) would result in a violent secondary outbreak. Scenario B (green lines): A stronger reduction of the contact rate which excludes a large part of social events (school, sports, large events, etc.) with the further inclusion of tracing and quarantine.

outbreak did not exhaust its virulence. In addition, guided by the analysis of the effects of the technological developments in the screening phase done in the previous sections, one can try to see their effects in the secondary events and the results of such an investigation are shown in Figure 5 where the comparison between the model predictions and the data is stringent. The data of the daily *new* (reported) cases, from 24th February on, show a large dispersion due to the influence of strategic decisions after the lockdown period (*i.e.* after 4th May 2020). The behaviour of the data follows rather consistently the model predictions of the reported cases of Equation (7) once scenarios for the mitigation of the lockdown are introduced. In particular, the data seem to follow rather consistently a scenario (green lines) where

the contact rate excludes a large part of social events (school, sports, large events) and it includes the introduction of tracing and quarantine. The fact that the longterm data (after 4th May 2020) are well reproduced by the model means that distancing and tracing measures proposed in Italy had a positive impact captured by the model. More dramatic scenarios (as described, for instance, by the cyan curves) have been avoided. Particularly relevant the value of $t_I = 1/\delta_I$ reached during the transition time to normality. Lowering the diagnostic time to $t_I \approx 7$ hours, or $t_I \approx 5$ hours has the power to substantially mitigate the secondary events, in particular, if one keeps "half rate contacts" (as schematically designed in Figure 4).

3. The Asymptomatic prevalence: results and discussion

Epidemiological surveillance of COVID-19 cases captures only a portion of all infections because the clinical manifestation of infections with SARS-CoV-2 ranges from severe diseases, which can lead to death, to asymptomatic infection. The official sources of data in Italy (Ministry of Health - *Ministero della Salute*, 2020) do not provide information about the asymptomatic patients thus limiting their usefulness in view of calculating interesting epidemic parameters, not last the lethality rate. Attempts to exploit the existing available data in order to estimate the prevalence and the lethality of the virus in the total Italian population has been proposed by Bassi, Arbia and Falorsi (Bassi, 2020). They used a post-stratification (Little, 1993) of the official data in order to derive the weights necessary for re-weighting the sample results. The re-weighting procedure artificially modify the sample composition so as to obtain a distribution which is more similar to the population. They obtain a prevalence of 9%, a rather large number.

Conversely, a (population-based) seroepidemiological survey can quantify the portion of population which developed antibodies agains SARS-CoV-2 providing information on the exposed individuals and on the remaining susceptible subjects (assuming that antibodies are marker of total or partial immunity). A certain number of surveys of SARS-CoV-2 have been realised or planned (in addition to the already quoted references: Sood, 2020; Valenti, 2020; Stringhini, 2020; Bryan, 2020; Shakiba, 2020; Doi, 2020; Eriks 2020; Wu, 2020; N. Bobrovitz, 2020).

3.1 Model results and Istat survey

The data, elaborated by Istat in their report, have been collected in a period of time (25th May -15th July 2020) running over (almost) two months. They are summarised in the upper and lower limits and the absolute value of Table 1 or (approximately) in the following ratio

$$\text{ratio}|_{\text{survey}} = \frac{\text{absolute value}}{\text{reported cases}} \approx 6. \qquad (11)$$

The ratio (11) is a clear sign of the large impact of the asymptomatic prevalence on the total infected population with positive IgG in Italy even

if no time dependence is assumed for those data, an assumption that needs some more attention. In fact, the model description we are proposing can be of some help to answer questions on the time dependence of the data. One can presume a roughly constant behaviour of the ratio (11) as a function of time under the following assumptions:

i) The antibodies have a lifetime significantly longer than the time needed for collecting the data.

ii) The social contact rate regime is rather stable during the collection of data. In this way the fraction of infected population is (on average) connected with the transmission probability of the infection which is assumed to be constant. In the present model $\beta = (1.5851 \pm 0.0336) \cdot 10^{-8}$ day^{-1}, (see Table 2).

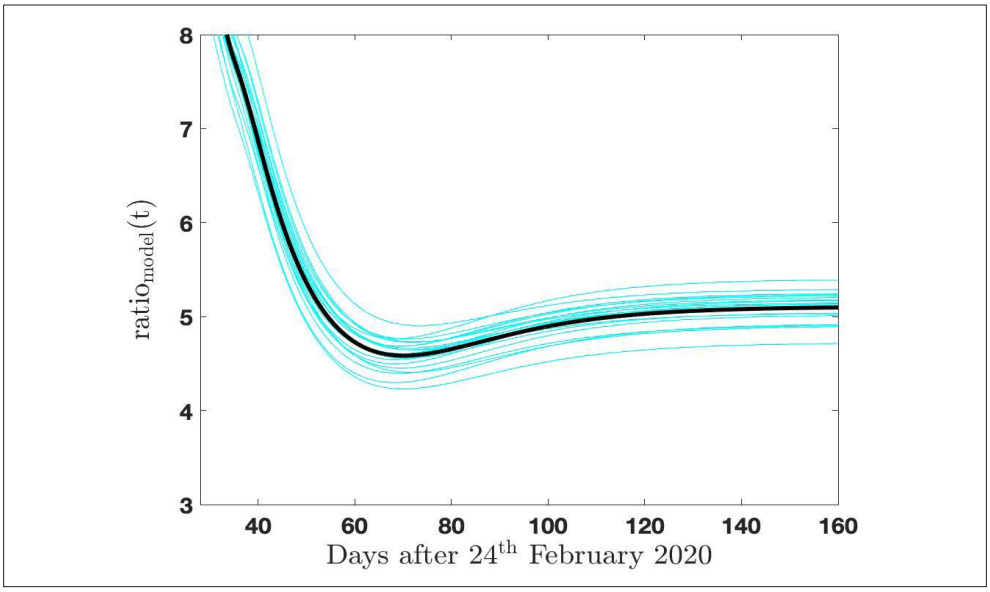
iii) In a (large and representative) sample, the portion of individuals reached by the infection in the previous months is fixed (next discussion will be largely devoted to clarify this point).

The assumption i) sounds reasonable and the assumption ii) can be explicitly checked in our model by defining the following (time dependent) ratio

$$\text{ratio}|_{\text{model}}(t) = \frac{A(t) + I(t) + H(t)}{H(t)}. \qquad (12)$$

The numerator, in Equation (12), counts, for each day, the total number of infected (reported and unreported with symptoms plus unreported asymptomatic infected individuals) and the denominator counts the number of reported cases. Both numerator and denominator are largely time dependent functions as already discussed in the previous sections. The model ratio (12) is shown in Figure 6. Despite the rapid decrease of the ratio before 25th May 2020, its value remains, after that date and till the end of July, approximately constant ($\text{ratio}_{\text{model}}(t) \approx 5$). In fact, the contact rate between 24th February -25th May 2020 had large variation due to lockdown and social restriction measures adopted in Italy, while in the period 25th May -31st July the social situation was rather stable. The model captures such variations validating the assumption ii) for that period.

Figure 6 - The model ratio of Equation (12) as a function of time



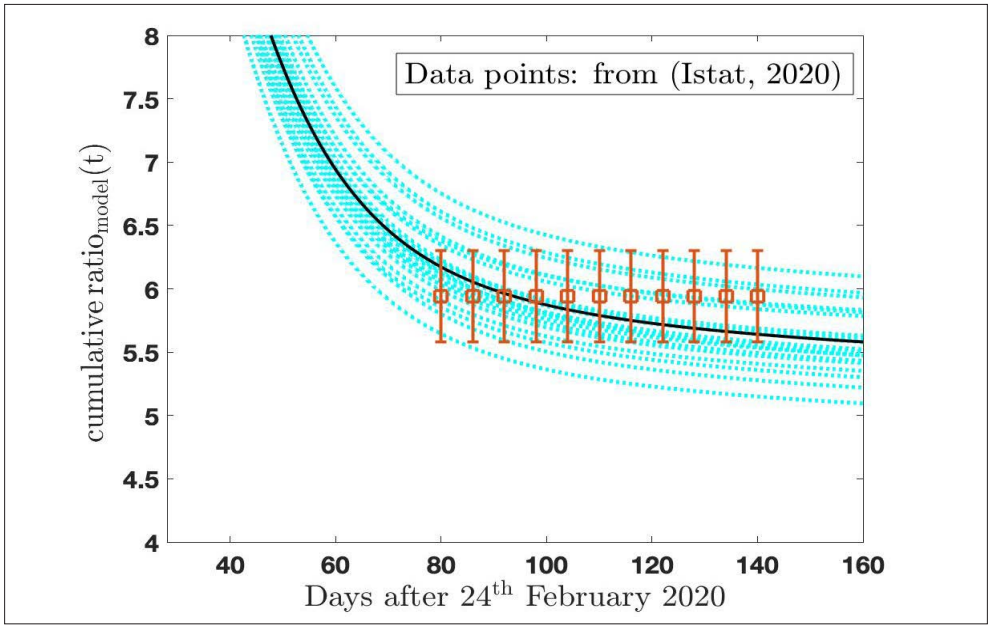
Source: Our processing

However the ratio (12) cannot be compared directly with the results of the Istat survey. As a matter of fact the survey measured the number of individuals who developed antibodies without selecting the period of infection. The serological test cannot answer to time dependent questions only measuring the presence of antibodies which simply count the total amount of infections in the period preceding the test. In order to reproduce, as close as possible, the analysis from Istat, we can calculate the cumulative values of the numerator and denominator of Equation (12) summing (up to a given day (t)) the components. The numerator of Equation (11) (*i.e.* the absolute value) is given by the sum of the Asymptomatic population (A), the (unreported) Infected (I) and the official reported cases (H); the denominator the cumulative value of the official reported cases. One gets:

$$\text{cumulative ratio}|_{\text{model}}(t) = \frac{\sum_{t'=0}^{t}\left[A(t') + I(t') + H(t')\right]}{\sum_{t'=0}^{t} H(t')}. \qquad (13)$$

The cyan curves in Figure 7 summarise the model predictions for the ratio (13). The black curve representing the mean value and the dotted lines the normal distributed variations.

Figure 7 - The fixed ratio of Equation (14), elaborated from the results of the seroprevalence survey (data-points) is compared with the model ratio of Equation (13) (cyan curves)



Source: Our processing

The results of the Istat survey can be transformed in an analogous ratio in a simple way: the absolute value corresponds to an average value of the prevalence (2.3 + 2.6)/2 ≈ 2.5 (see Table 1). Therefore the upper and lower limit of the infected population divided by the reported cases is summarised by the ratio

$$\text{ratio}|_{\text{survey}} = 5.94 \pm 0.36, \qquad (14)$$

which now includes the uncertainties estimated in the Istat survey and updates the approximation (11). The data points in Figure 7 give a graphical representation of the ratio (14) assumed, in a first step, to be time-independent on the basis of the assumption i) and ii) already discussed.

From this first analysis one can conclude that:

1. the global comparison in Figure 7 between the estimated values and the evidences of reference (Istat, 2020) is surprisingly successful, in particular when statistical uncertainties are included in the analysis

of the survey and in the model predictions. For a long period of time during the lockdown months, speculations on the Asymptotic population gave strongly divergent estimates. Our results of Figure 7 are obtained within a model fixed in 6th April 2020 and published before the survey (Traini, 2020); data confirm that the epidemiological models can offer predictions rather stable and realistic also for the Asymptomatic compartment.

2. The model results for the cumulative ratio of Figure 7 start with a value ≈ 8 at the beginning of April 2020, reaching an asymptotic *average* value ≈ 5.5 at the end of July 2020. For the whole period of the preliminary survey (25th May -15th July) the ratio of Equation (14), (data-points in Figure 7), is reproduced by the model calculation of Equation (13) within the statistical uncertainties.
3. The conclusions of the Istat survey (*i.e.* "*the seroprevalence data at regional level, to be integrated with the epidemiological surveillance data, are particularly relevant to identify, on one side, the portion of individuals reached by the infection in the previous months, and for programming measures to prevent future possible second waves, on the other side*"[7]) can easily be applied to the relevance of modelling the behaviour of the asymptomatic population, a key ingredient to manage the future of the pandemic event (*e.g.* Gandhi, 2020).

3.2 Sensitivity and specificity of the serological tests

A first correction to the seroprevalence analysis presented in the previous Section is due to the sensitivity and specificity of the serological tests adopted in the survey screenings (for a recent note on the false positive and false negative in diagnosis of COVID-19, see Jia, 2020*b*).

The report by Istat (Istat, 2020: page 9) indicates in *not less than 95%* the specificity of the tests and in *not less than 90%* their sensitivity. The consequent false negative and false positive classified individuals can be

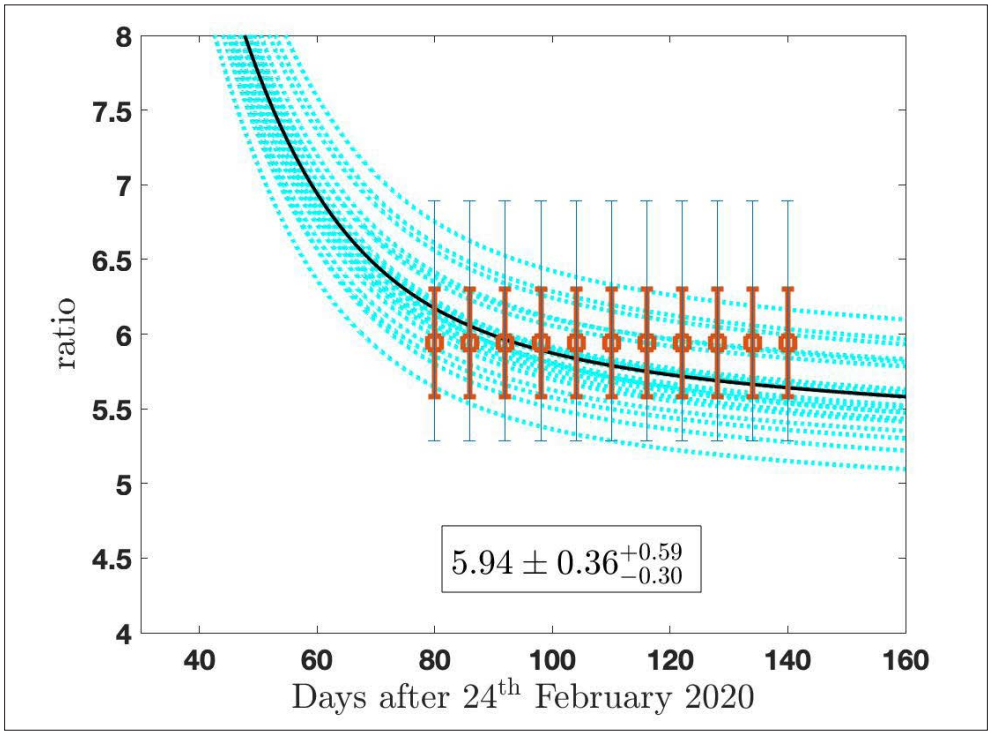
7 Our translation. "*I dati di siero-prevalenza a livello regionale, da integrare con quelli di sorveglianza epidemiologica, sono particolarmente preziosi sia per conoscere la quota di popolazione che è stata infettata nei mesi precedenti, sia per la messa a punto di programmi sanitari al fine di prevenire future ondate dell'epidemia e orientare adeguatamente le politiche sanitarie*" (in Italian). (Istat, 2020: page 1).

taken into account increasing the uncertainties of the ratio (14) of an amount of 10% and 5% in the two directions, assuming the maximum error in the data. One gets:

$$\text{ratio}|_{\text{survey}} = 5.94 \pm 0.36^{+0.59}_{-0.30}, \tag{15}$$

which replaces the value (14) to take into account the (asymmetric) corrections due to false responses. Figure 7 is, consequently, replaced by Figure 8 where the model results are compared with the more elaborated analysis. The asymmetric increase of the error-bars is clearly shown, however, at the same time, the conclusions of Section 3.1 remain basically valid.

Figure 8 - The ratios of Equations (14) and (15) compared with the cumulative model ratio of Equation (13) (cyan curves) (a)



Source: Our processing on Istat 2020 data

(a) The smallest error bars take into account the statistical variation suggested by the Istat report (Istat, 2020). The largest error bars include the modifications induced by the sensitivity and specificity of the IgG tests as indicated by the same Istat report.

4. Critical aspect and limits

The present Section is devoted to the critical aspects of the results and methods used in the present investigation from two points of view: the data and the model. From the point of view of the data an important element is still not well fixed: the time dependence of the data during their collection, and it will be discussed in the next Section 4.1. From the point of view of the model study one cannot remain only with its analysis of the past and in Section 4.2 a study of the present (and future) distribution of the reported cases and of the asymptomatic cases will be discussed. A stringent test for the model.

4.1 Time dependence of antibody tests

A recent meta-analysis by Deeks *et al.* (Deeks, 2020) observed substantial heterogeneity in sensitivities of IgA, IgM and IgG antibodies, or combinations, for results aggregated across different time periods post-symptom onset. They based the main results of the review on the 38 studies that stratified results by time since symptom onset[8].

In particular IgG/IgM all showed low sensitivity during the first week since onset of symptoms (all less than 30.1%), rising in the second week and reaching their highest values in the third week. The combination of IgG/IgM had a sensitivity of 30.1% (95% CI 21.4 to 40.7) for 1 to 7 days, 72.2% (95% CI 63.5 to 79.5) for 8 to 14 days, 91.4% (95% CI 87.0 to 94.4) for 15 to 21 days. Estimates of accuracy beyond three weeks are based on smaller sample sizes and fewer studies. For 21 to 35 days, pooled sensitivities for IgG/IgM were 96.0% (95% CI 90.6 to 98.3). There are insufficient studies to estimate sensitivity of tests beyond 35 days post-symptom onset. Summary specificities (provided in 35 studies) exceeded 98% for all target antibodies with confidence intervals no more than 2 percentage points wide. Assuming as a reference point the results of the meta-analysis by Deeks *et al.*, one must correct further the uncertainties of Figure 8. The questionnaire filled by the people involved in the screening for the Istat survey (Istat, 2020) included questions on the exact period of the symptom onset, a relevant piece of information in order to correct the data and to analyse the effects of the time-dependent sensitivity of

8 The numbers of individuals contributing data within each study each week are small and are usually not based on tracking the same groups of patients over time.

the tests. However information is not available at the moment, consequently it is necessary to introduce more drastic assumptions and corrections. One remains with the simple assumption that sample has no privilege with respect to the time dependence of the sensitivity. As a consequence the single test has to be considered (in average) affected by the average value of the sensitivity among 30.1% (days 1-7), 72.2% (days 8-14), 91.4% (days 15-21), 96% (days 22-35), and 90% for days > 35 (and till the end of the screening period, *i.e.* the remaining 17 days). 90% is indeed the minimum value of the sensitivity proposed by Istat and discussed in Section 3.2:

$$\frac{7 \times 30.1 + 7 \times 72.2 + 7 \times 91.4 + 14 \times 96.0 + 17 \times 90.0}{7 + 7 + 7 + 14 + 17} = \frac{4229.9}{52} \approx 81.3. \qquad (16)$$

The sensitivity of the test decreases from *not less than 90%* to 81.3%, while the specificity remains not less than 95%.

Taking into account the new estimated sensitivity (16), the ratio (15) is replaced by

$$\text{ratio}|_{\text{survey}} = 5.94 \pm 0.36^{+1.11}_{-0.30}, \qquad (17)$$

and the Figure 8 by the Figure 9. The asymmetric increase of the error-bars is again clearly shown, and the conclusions drawn in Section 3.1 become more weak.

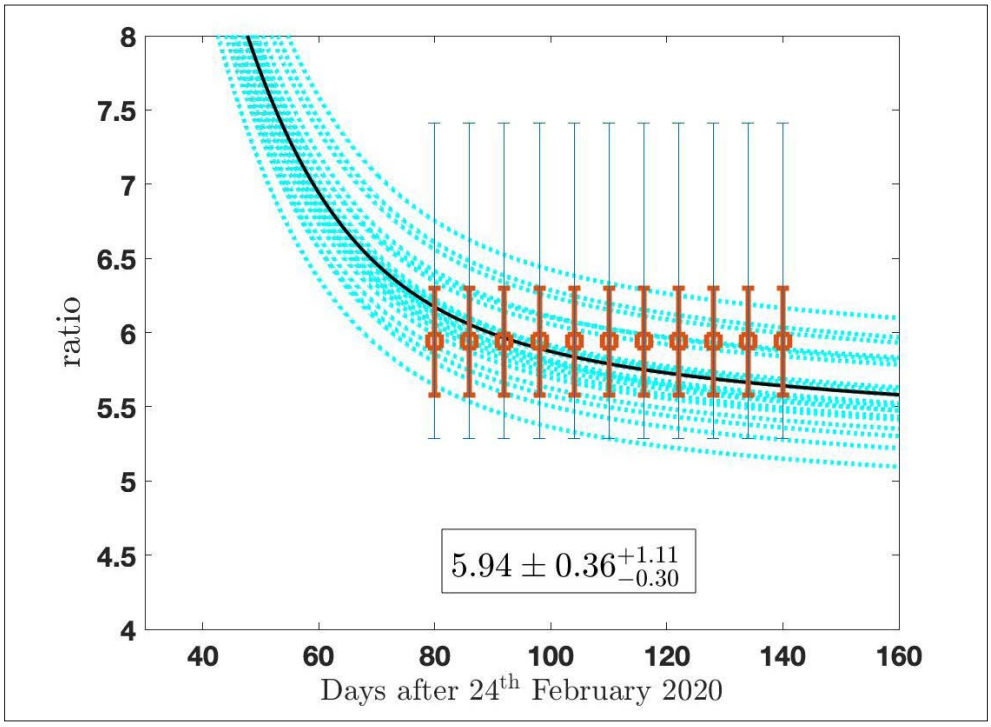
1. The comparison between the model estimates and the data corrected for the time dependence of the sensitivity of the tests as in Figure 9 is less accurate and the data cannot be considered a stringent test for models. The estimate of the asymptomatic prevalence remains valid, but with a larger interval of confidence.
2. Model and corrected data are still consistent for the whole time period.
3. The validity of the model description remains a guide in the interpretation of the unknown asymptomatic distribution within a larger interval of values. A more careful analysis of the Istat data, in particular the reference to the mentioned questionnaire (Istat, 2020) can help in making again the comparison more stringent.

4.2 Longer term study: last months of the year 2020

The present Section is devoted to the validation of the model (and predictions) with the most recent data on the second epidemic wave in Italy (after 25th September 2020), in particular the predictions for the asymptomatic compartment. The parametrisation is again based on values indicated in Table 2, no *ad hoc* modifications are introduced. The model is simply normalised to the value of the reported cases of 25th September (47,718), as indicated in the official site of the Italian Ministry of Health (*Ministero della salute*, 2020).

25th September 2020 is assumed as the beginning of the second wave of infection. The new results are summarised in Figure 10 (analogous to the Figure 5 describing the previous period) and Figure 11 (analogous to Figure 2).

Figure 9 - The ratios of Equations (14) and (17), are compared with the cumulative model ratio of Equation (13) (cyan curves) (a)



Source: Our processing

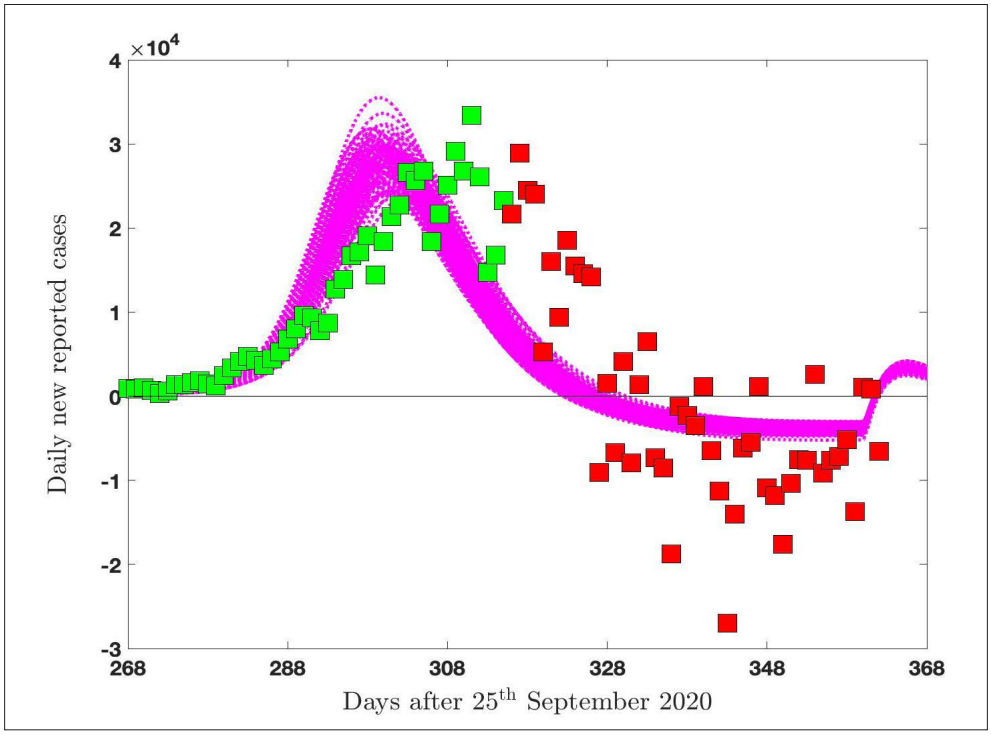
(a) The smallest error bars take into account the statistical variation suggested by the Istat report (Istat, 2020). The largest error bars include the modifications induced by the sensitivity and specificity of the IgG and their time-dependence as elaborated by Deeks *et al.*, 2020.

The social measures adopted in the second period are rather different from the strict lockdown of the first events, in particular the measures assumed are locally differentiate and following the local virulence of the infection. The new approach of the governmental institutions assumes a strategically flexible response to the virus allowing for a possible "coexistence". The response is therefore more rigid and without rapid variations. A behaviour compatible with the results shown in Figure 10 where the data for the daily variations exhibit a broader aspect in comparison with the model predictions. Despite such a differences, the model remain rather consistent with data of Figure 10 which represent a stringent test since they are related to the derivative of the distributions. In addition the data are submitted to large fluctuations due to data collection variation and local inhomogeneities. The data are presented in two different colors: green (till 13th November 2020) and red afterwards (last update 28th December 2020). The simple raison is that the model has been applied for the first time to the new data in 13th November 2020. After 13th November the additional official data have been added to this Figure, day by day, and no modification or renormalisation is introduced, *i.e.* the model predictions are fixed.

Much more stable are the results of Figure 11 where data points (by Ministero della salute, 2020) for the daily reported cases are shown as a function of time (days), from 25th September on (last update 28th December). Once again the model predictions are compatible with data to a large extent. Such a consistency allows again a risky prediction for the unreported (asymptomatic and symptomatic) cases. The prevalence is rather large (approximately a factor of 4) quite similar to the prevalence emerging from the analogous Figure 2 (factor of 5 roughly, see also Figure 6). Obviously one needs detailed serological tests to have confirmation, however the stability of the virus is assumed by the parametrisation of Table 2 and seems to be consistent with data.

A last piece of information emerges from Figure 11: an indication of the possible effect due to the Christmas period. In Figures 11 and 10 one can see the effect of a foreseeable relaxation of the distancing measures in the behaviour of families and friends group. An average increase of the contact rate (as indicated in Table 2) produces a small wave emerging just after 25th December 2020. A graphical indication of a possible increase of the number of infected.

Figure 10 - The model predictions for the daily new reported cases 25th September - 31st December 2020 in Italy, compared with the official data of the Ministry of Health

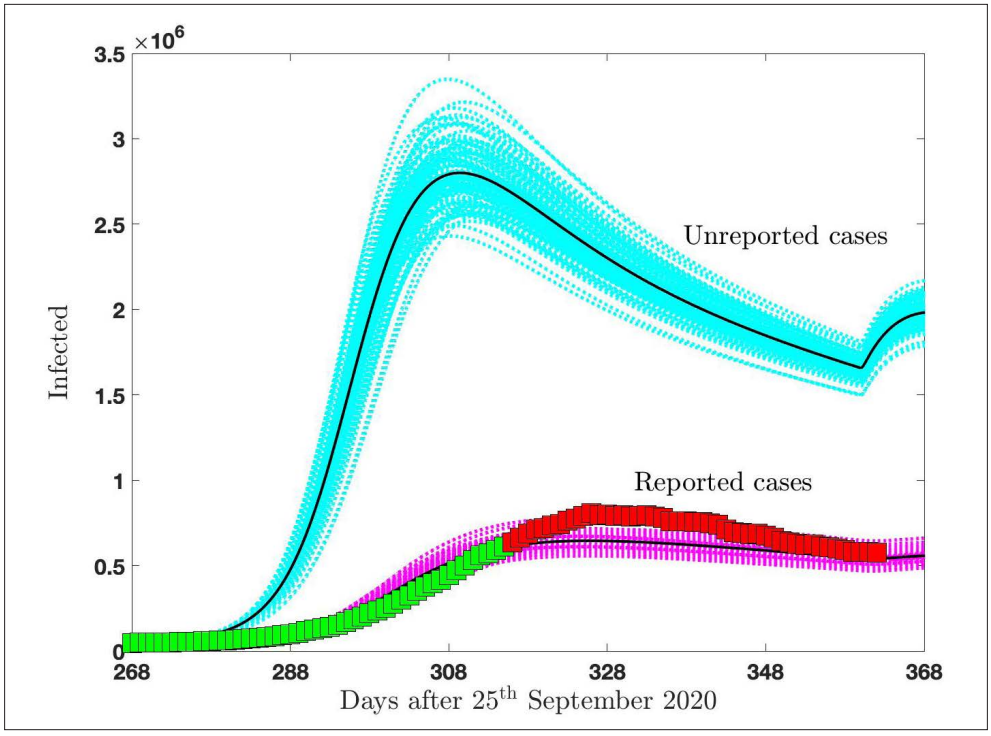


Source: Our processing using the Italian Ministry of Health data, 2020 (*Ministero della Salute*, 2020)

5. A look at vaccination

The present Section is devoted to a burdensome aspect of the nowadays discussion[9] on SARS-CoV-2: vaccination. The United Kingdom started the vaccination campaign earlier, while all the European countries started their own campaign on 27th December 2020. The organisation of a massive historical event is rather heavy and it will need the effort of many institutions.

Figure 11 - The MCMC analysis of the infected individuals (reported and unreported cases) in Italy as a function of time in the period 25th September - 31st December 2020 (a)



Source: The data for the reported-infected individuals are by the Italian Ministry of Health (*Ministero della Salute*, 2020) and are compared with our model predictions (magenta curves) which include statistical uncertainties.
(a) The reported cases are compared with the predictions for the class of infected non reported cases (asymptomatic and symptomatic) in the same period of time (cyan curves).

9 21st December 2020.

5.1 Modelling vaccination: scenarios

A preliminary aspect is the assessment of a possible vaccination scenario. We do not discuss specific strategies, but we want to establish general scenarios to illustrate main advantages and disadvantages within simple assumptions. Basically we will assume that the year 2021 will be devoted to vaccination in Italy and that the order of magnitude of the immunised persons per day is between 40,000 and 80,000 (to be selected with specific social and homogeneity criteria) (*e.g.* Makoul, 2020).

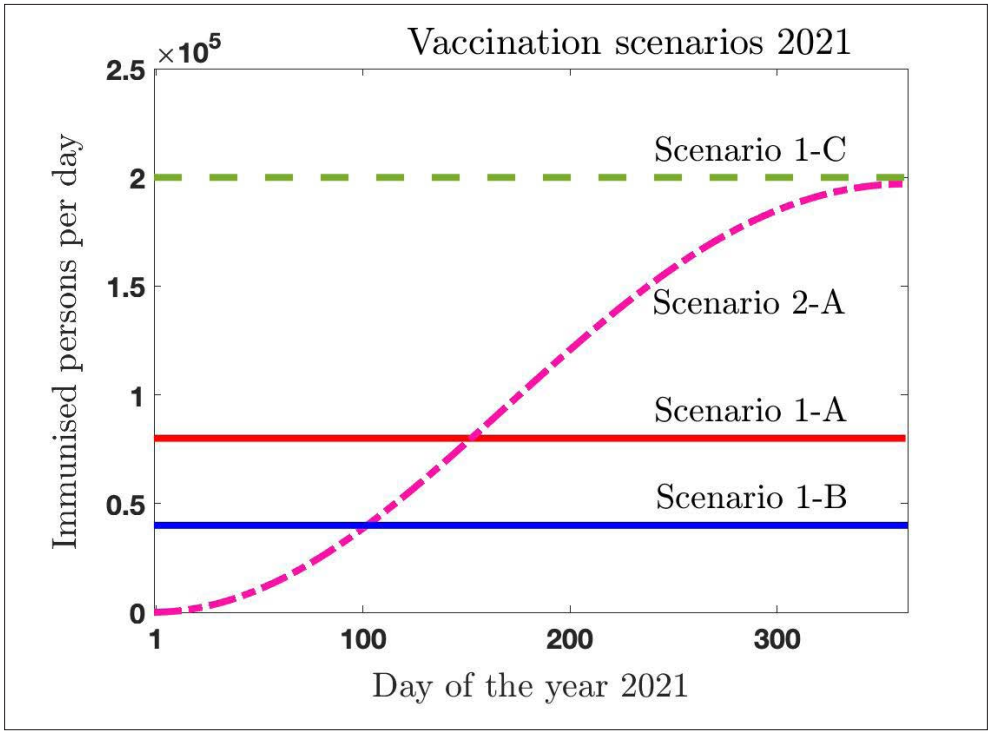
In the following two basic scenarios will be introduced, as illustrated in Figure 12.

i) The first scenario assumes an homogenous distribution, during the 2021, of a fixed number of immunised persons per day: 80,000 (scenario 1-A), (full line in Figure 12), and also 40,000 (scenario 1-B) will be considered. The effort to start with a large portion of population is high and a second scenario is investigated;

ii) The second scenario assumes that the amount of immunised persons during the first period is rather low and the process will accelerate during the year (dotted line in Figure 12, scenario 2-A).

iii) Almost 30 million people are immunised at the end of the year in both scenarios.

Figure 12 - A constant number of person are immunised per day during the year 2021; 80,000, full curve scenario (1-A); 40,000, scenario (1-B); 200,000, scenario (1-C) (a)



Source: Our processing

(a) Alternatively, in scenario (2-A), an increasing number of people are immunised during 2021 allowing for a less strong starting effort. The total number of immunised people results (almost) 30 million for both scenarios, at the end of the year.

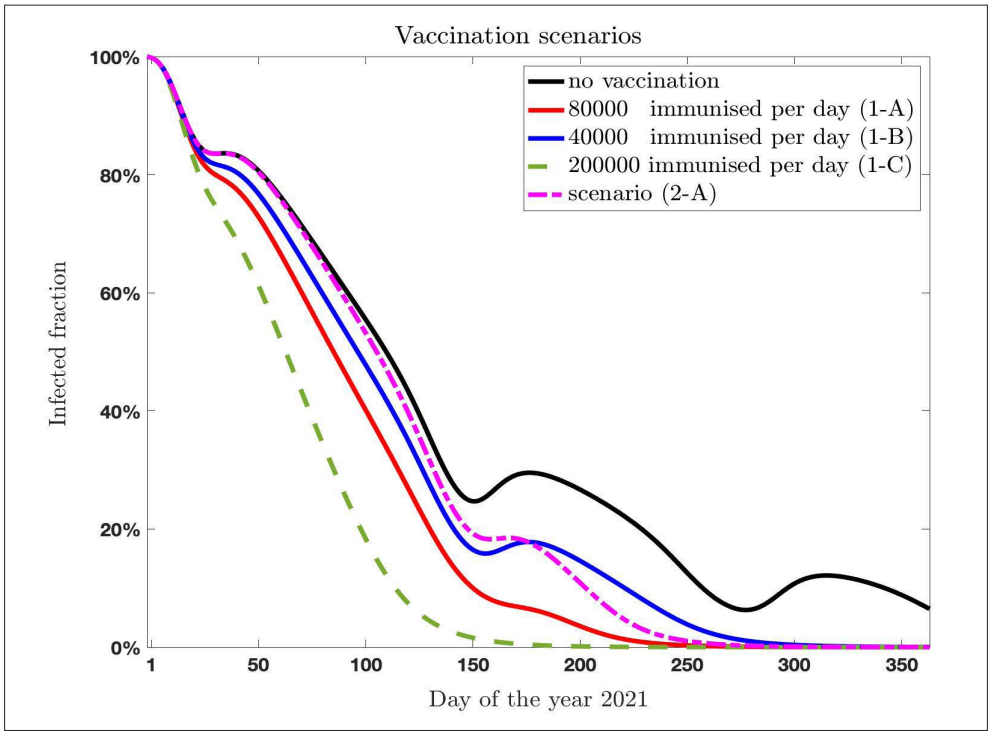
5.2 Year 2021: a perspective

The Section is devoted to a general perspective of the year 2021 offering a macroscopic view of the pandemic event in Italy and the advantages of the vaccination campaign. In Figure 13 one can have an approximated idea of the time-evolution of the SARS-CoV-2 infection in Italy during the year[10] in the case

10 The qualitative intent of the present discussion is put in further evidence by the fact that no-Covid deaths and newborns are not included in the investigation. One is simply assuming that they compensate approximately during the 2021. In addition we will not discuss vaccine efficacy assuming $VE = 95\%$ for two reasons: i) we do not know the real efficacy of the different vaccines; ii) it is rather simple to rescale results for an accepted efficacy (Moghadas, 2020).

of *no vaccination*, an assumption which is ruled out, but it can offer a reference point to appreciate the advantages of the vaccines. Such a scenario is described by the full black line (Figure 13). One can immediately see that the number of the daily reported cases is decreasing during the year since an increasingly large number of "susceptible" individuals get immunised (or dead) after the infected period. The year 2021 will not be sufficient to obtain a vanishing number of infected: in summer they will be around 150,000 and during the next Christmas period roughly 50,000. A rough estimate which takes into account

Figure 13 - The fraction of the reported cases as a function of the day of the year 2021 (a)



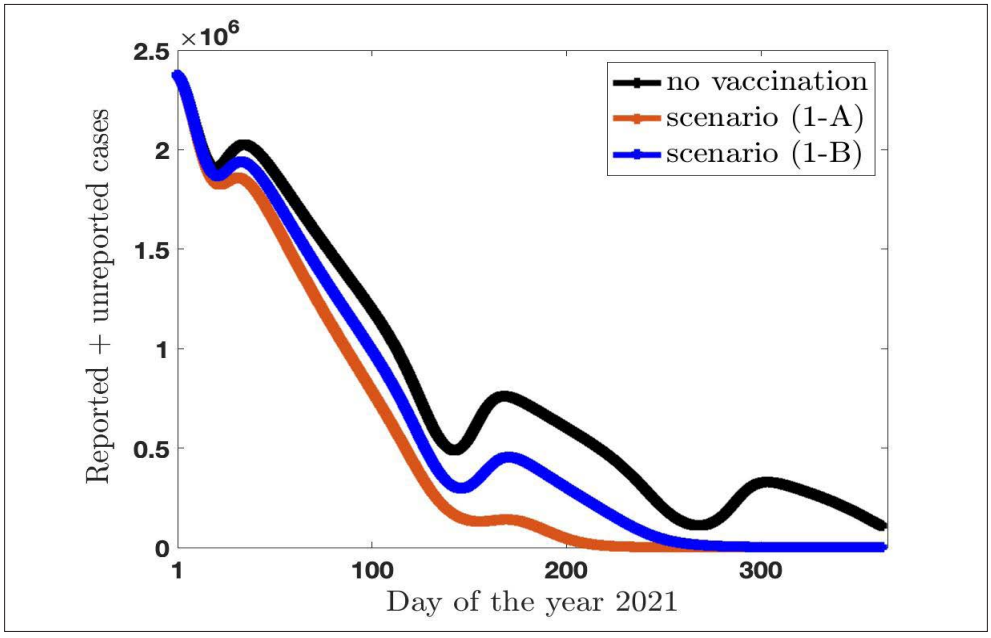
Source: Our processing

(a) On 1st January they were 574,767, equivalent to 100%. The scenario assuming no vaccination (continuous black line) is compared with the scenario (1-A) which assumes a constant rate of vaccines of 80,000 immunised each day (continuous red line). The blu line, scenario (1-B), shows results for 40,000 immunised per day at constant rate. The scenario (2-A), where the rate is not constant and a low starting period is compensated by an accelerated second immunisation period (see Figure 12), is illustrated by the dotdashed (magenta) curve.

oscillations due to possible waves of infections[11] as it emerges from the wave behaviour of the curve. The situation changes largely with the introduction of the vaccination at a rate of 80,000 persons per day (red curve, scenario (1-A)). A reduction of roughly 100,000 cases is already evident in April, while in summer (June) one reaches the 20,000 cases instead of 150,000, in August the virus will give a definitive "*ciao*". As a matter of fact one is gaining several months (from six to ten months) because of vaccination without counting the number of deaths.

The scenario (1-B) (blu line) assumes half immunised persons per day (40,000) and the disadvantages are evident with respect the previous hypothesis of 80,000. One has to wait end of October to eliminate the virus infection and in June one has still 80,000 cases.

Figure 14 - Both reported and unreported cases as a function of the day of the year 2021 (a)



Source: Our processing

(a) The scenario assuming no vaccination (full black line) is compared with the scenario (1-A) which assumes a constant rate of vaccines of 80,000 immunisations each day (full red line). The blu line, scenario (1-B), shows results for 40,000 immunised per day at constant rate.

11 The oscillating behaviour is simulated by an oscillating contact rate $c(t) = c_0 + 10\ c_0 \cdot \cos^2(\Omega t)$, with $0 \leq t \leq 365$ and $\Omega = 4\pi/365$, see Table 2.

Also the scenario (2-A) has been implemented in the model and one can see the effects of a slow rate at the beginning of the year in the lower panel of the same Figure 13. The results of the vaccination are almost invisible till the end of May. They appear rapidly in the second part of the year producing the final result at the end of October. The summer is still a hard period and scenario (2-A) is rather similar to scenario (1-B).

Looking at Figure 13 is not intuitive to accept that the reduction of the cases from 500,000 to zero in one single year can produce the end of the infection in Italy with 60 million of residents.

To partially restore intuition, one has to realise that, after the discussion of the asymptomatic cases of Section 3, a large part of the job is done by the unreported cases. In Figure 14 the behaviour of the total (unreported + reported cases) during the year as indicated by the model. They complement Figure 13 showing that the decreasing behaviour of the reported cases has a specular behaviour for the asymptomatic. The vaccine helps in a strong way the reduction of both.

6. Concluding remarks

The traditional compartmental classes of Susceptible, Exposed, Infected and Recovered which characterise a large fraction of epidemical models, are not sufficient to simulate the coronavirus (SARS-CoV-2) pandemic infection. Many of the people who contract the diseases are Asymptomatic, they are infected and contagious and are often invoked as one of the causes for the rapid spread of the infection. It is hard to estimate the amount of asymptomatic, estimates ranges from 40% to 75% (Buitrago-Garcia, 2020; Oran, 2020; Day, 2020; Muzimoto, 2020) of the total infected population. In a small scale the particular conditions of the Vo' village (near Padua, Italy) allowed for two detailed surveys after a localised lockdown: the analysis found a prevalence of 1.2% (95% CI:0.8-1.8%). Notably, 42.5% (95% CI:31.5-54.6%) of the confirmed SARS-CoV-2 infections were asymptomatic (without any symptoms at the time of positive swab testing, and did not develop symptoms afterwards (Lavezzo, 2020)).

In August 2020, the Italian National Institute of Statistics – Istat presented the preliminary results of seroprevalence survey on the percentage of individuals reached by the SARS-CoV-2 infection in the previous months. The survey aims at estimating, in a methodological precise way, also the asymptomatic infected population and its role in the infection spread in Italy, one of the most affected countries in Europe.

The present study investigates a model description of the entire infected population in Italy considering a eightfold compartmental model which includes Infected, Asymptomatic and Quarantined population in addition to the more classic Susceptible, Recovered, Exposed. The model is illustrated in few examples and used to investigate in detail the asymptomatic prevalence including also sensitivity and specificity of the antibody tests used for the surveys. The results validate the model description and encourage other studies to detail, in a more quantitative way, the role of time-dependence of the sensitivity of the test used for the antibody screening.

Also predictions for the period of the second wave in Italy are presented and discussed, including asymptomatic predictions. Despite the fact that the epidemiological surveillance of the second wave of the epidemic event in Italy is characterised by a strong use of swabs, the model predictions for

the amount of unreported cases is of the same order of magnitude of the percentage already seen in the first outbreak. Future antibody screening will verify the present prediction.

To complete modelling, Section 5 has been devoted to the vaccination scenarios. The role played by the introduction of vaccination is clearly shown in reducing in a significative way the time to reach the end of the infection. The strategies of the administration can be simulated and they favour an intense administration from the beginning. Starting with a low rate of administration the risk is to loose a large part of the advantage.

References

Ansumali, S., S. Kaushal, A. Kumar, M.K. Prakash, and M. Vidyasagar. 2020. "Modelling a pandemic with asymptomatic patients, impact of lockdown and herd Immunity, with applications to SARS-CoV-2". *Annual Reviews in Control*, Volume 50: 432-447.

Bassi, F., G. Arbia, and P.D. Falorsi. 2020. "Observed and estimated prevalence of COVID-19 in Italy: How to estimate the total cases from medical swabs data". *Science of the Total Environment*, Volume 764: 142799.

Bobrovitz, N., R.K. Arora, T. Yan, H. Rahim, N. Duarte, E. Boucher, J. Van Wyk, and T. Grant Evans. 2020. "Lessons from a rapid systematic review of early SARS-CoV-2 serosurveys". *medRxiv – The Preprint Server for Health Sciences*.

Brooks, S., A. Gelman, G. Jones, X.-L. Meng. 2011. *Handbook of Markov Chain Monte Carlo*. Milton Park, U.K.: Chapman & Hall/CRC.

Bryan, A., G. Pepper, M.H. Wener, S.L. Fink, C. Morishima, A. Chaudhary, K.R. Jerome, P.C. Mathias, AL. Greninger, and A.J. McAdam (*ed.*). 2020. "Performance Characteristics of the Abbott Architect SARS-CoV-2 IgG Assay and Seroprevalence in Boise, Idaho". *Journal of Clinical Microbiology*, 58: e00941-20.

Buitrago-Garcia, D., D. Egli-Gany, MJ. Counotte, S. Hossmann, H. Imeri, A.M. Ipekci, G. Salanti, and N. Low. 2020. "Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis". *PLoS Medicine*, 17(9): e1003346.

Day, M. 2020. "COVID-19: Identifying and isolating asymptomatic people helped eliminate virus in Italian village". *The BMJ*, 368: m1165.

Deeks, J.J., J. Dinnes, Y. Takwoingi, C. Davenport, *et al.* 2020.

Deeks, J.J., J. Dinnes, Y. Takwoingi, C. Davenport, R. Spijker, S. Taylor-Phillips, A. Adriano, S. Beese, J. Dretzke, L. Ferrante di Ruffano, I.M. Harris, M.J. Price, S. Dittrich, D. Emperador, L. Hooft, M.Mg. Leeflang, A.Van den Bruel, and Cochrane COVID-19 Diagnostic Test Accuracy Group. 2020. "Antibody tests for identification of current and past infection

with SARSCoV-2". *The Cochrane Database of Systematic Reviews*, 6(6): CD013652.

Doi, A., K. Iwata, H. Kuroda, T. Hasuike, S. Nasu, A. Kanda, T. Nagao, H. Nishioka, K. Tomii, T. Morimoto, and Y. Kihara. 2020. "Estimation of seroprevalence of novel coronavirus disease (COVID-19) using preserved serum at an outpatient setting in Kobe, Japan". *medRxiv – The Preprint Server for Health Sciences*.

Erikstrup, C., C.E. Hother, O.B.V. Pedersen, K. Mølbak, R.L. Skov, D.K. Holm, S.G. Sækmose, A.C. Nilsson, P.T. Brooks, J.K. Boldsen, C. Mikkelsen, M. Gybel-Brask, E. Sørensen, K.M. Dinh, S. Mikkelsen, B.K. Møller, T. Haunstrup, L. Harritshøj, B.A. Jensen, H. Hjalgrim, S.T. Lillevang, and H. Ullum. 2020. "Estimation of SARS-CoV-2 infection fatality rate by real-time antibody screening of blood donors". *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, Volume 72(2): 249-253.

Gandhi, M., D.S. Yokoe, and D.V. Havir. 2020. "Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control COVID-19". *The New England Journal of Medicine*, 382: 2158-2160.

Garcia-Basteiro, A.L., G. Moncunill, M. Tortajada, et al. 2020. "Seroprevalence of antibodies against SARS-CoV-2 among health care workers in a large Spanish reference hospital". *Nature Communication*, 11: Article n. 3500 (2020).

Giordano, G., F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri. 2020. "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy". *Nature Medicine*, 26. 855-860.

Guerriero, M, Z. Bisoffi, A. Poli, C. Micheletto, and C. Pomari. 2020. "Prevalence of asymptomatic SARS-CoV- 2- positive individuals in the general population of northern Italy and evaluation of a diagnostic serological ELISA test: a cross-sectional study protocol". BMJ Open, 10(10): e040036.

Hogg, D.W., and D. Foreman-Mackey. 2018. "Data Analysis Recipes: Using Markov Chain Monte Carlo". The Astrophysical Journal, Supplement Series, Volume 236: 11.

Istituto Nazionale di Statistica – Istat (Italian National Institute of Statistics), and Ministero della Salute (Ministry of Health). 2020. "Primi risultati dell'indagine di seroprevalenza sul SARS-CoV-2". Comunicato Stampa. Roma, Italia: Istat. https://www.istat.it/it/archivio/246156.

Jia, X., J. Chen, L. Li, N. Jia, B. Jiangtulu, T. Xue, L. Zhang, Z. Li, R. Ye, and B. Wang. 2020*a*. "Modeling the Prevalence of Asymptomatic COVID-19 Infections in the Chinese Mainland". The Innovation, 1: 100026.

Jia, X., L. Xiao, Y. Liu, 2020*b*. "False negative RT-PCR and false positive antibody tests-Concern and solutions in the diagnosis of COVID-19". *The Journal of Infection*, Volume 82(3): 414-451.

Kinoshita, R., A. Anzai, S.-M. Jung, N.M. Linton, T. Miyama, T. Kobayashi, K. Hayashi, A. Suzuki, Y. Yang, A.R. Akhmetzhano, and H. Nishiura. 2020. "Containment, Contact Tracing and Asymptomatic Transmission of Novel Coronavirus Disease (COVID-19): A Modelling Study". *The Journal of Clinical Medicine*, 9 (10). 3125.

Lavezzo, E., E. Franchin, C. Ciavarella, et al. 2020. "Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'". *Nature*, 584: 425-429.

Little, R.J.A. 1993. "Post-Stratification: A Modeler's Perspective". J*ournal of the American Statistical Association*, Volume 88, N. 423: 1001-1012.

Makoul, M., H.A. Houssein, H. Chemaitelly, S. Seedat, G.R. Mumtaz, S. Al-Omari, and L.J. Abu-Raddad. 2020. "Epidemiological Impact of SARS-CoV-2 Vaccination: Mathematical Modeling Analyses". *Vaccines*, 8 (4): 668.

Ministero della Salute – Italian Ministry of Health. 2020. *Portale: Nuovo Coronavirus*. Roma, Italia: Ministero della Salute. http://www.salute.gov.it/portale/nuovocoronavirus/homeNuovoCoronavirus.jsp

Mizumoto, K., K. Kagaya, A. Zarebski, and G. Chowell. 2020. "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020". *Eurosurveillance Journal*. Volume 25, Issue 10: 2000180.

Moghadas, S.M., T. N. Vilches, K. Zhang, C.R. Wells, A. Shoukat, B.H. Singer, L.A. Meyers, K.M. Neuzil, J.M. Langley, M.C. Fitzpatrick, and A.P. Galvani. 2020. "The impact of vaccination on COVID-19 outbreaks in the United States". *Clinical Infectious Diseases*: ciab079.

Oran, D.P., and E.J. Topol. 2020. “Prevalence of Asymptomatic SARS-CoV-2 Infection”. *Annals of Internal Medicine*, 174(2): 286-287.

Park, S.W., D.M. Cornforth, J. Dushoff, and J.S. Weitz. 2020. “The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak”. *Epidemics*, N. 31: 100392.

Peterson, I., and A. Phillips 2020. “Three Quarters of People with SARS-CoV-2 Infection are Asymptomatic: Analysis of English Household Survey Data”. *Clinical Epidemiology*, Volume 12: 1039-1043.

Pollán, M., 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.L. Paniagua, J.F. Muñoz-Montalvo, F. Blanco, and R. Yotti. 2020. “Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study”. *The Lancet*. Volume 396, Issue 10250: 535-544.

Remuzzi, A., and G. Remuzzi. 2020. “COVID-19 and Italy: what next?”. *The Lancet*, Volume 395, Issue 10231: 1225-1228.

Robinson, R., and N.I. Stilianakis. 2013. “A model for the emergence of drug resistance in the presence of asymptomatic infections”. *Mathematical Biosciences*, 243(2): 163-177.

Salje, H., C.T. Kiem, N. Lefrancq, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. Le Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle, and S. Cauchemez. 2020. “Estimating the burden of SARS-CoV-2 in France”. *Science*, Volume 369, Issue 6500: 208-211.

Shakiba, M., M. Nazemipour, A. Salari, F. Mehrabian, S.S.H. Nazari, S.M. Rezvani, Z. Ghasempour, A. Heidarzadeh, and M.A. Mansournia. 2020. “Seroprevalence of COVID-19 virus infection in Guilan province, Iran”. *Emerging Infectious Diseases*, 27(2): 636-638.

Snoeck, C.J., M. Vaillant, T. Abdelrahman, V.P. Satagopam, et al. 2020. “Prevalence of SARS-CoV-2 infection in the Luxembourgish population - the CON-VINCE study”. *medRxiv – The Preprint Server for Health Sciences*.

Sood, N., P. Simon, P. Ebner, D. Eichner, J. Reynolds, E. Bendavid, and J. Bhattacharya. 2020. "Seroprevalence of SARS-CoV-2-specific antibodies among adults in Los Angeles County, California, on April 10-11, 2020". *JAMA*, 323(23): 2425-2427.

Stringhini, S., A. Wisniak, G. Piumatti, A.S. Azman, S.A. Lauer, H. Baysson, D. De Ridder, D. Petrovic, S. Schrempft, K. Marcus, S. Yerly, I.A. Vernez, O. Keiser, S. Hurst, K.M. Posfay-Barbe, D. Trono, D. Pittet, L. Gétaz, F. Chappuis, I. Eckerle, N. Vuilleumier, B. Meyer, A. Flahault, L. Kaiser, and I. Guessous. 2020. "Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study". *The Lancet*, Volume 396, Issue 10247: 313-319.

Tang, B., X. Wang, Q. Li, N.L. Bragazzi, S. Tang, Y. Xiao, and J. Wu. 2020. "Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions". *Journal of Clinical Medicine*, Volume 9(2): 462.

Tang, B., N.L. Bragazzi, Q. Li, S. Tang, Y. Xiao, and J. Wu. 2020. "An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov)". *Infectious Disease Modelling*, Volume 5: 248-255.

Traini, M.C., C. Caponi, R. Ferrari, and G.V. De Socio. 2020. "A study of SARSCoV-2 epidemiology in Italy: from early days to secondary effects after social distancing". *Infectious Diseases*, Volume 52, Issue 12: 866-876.

Valenti, L., A. Bergna, S. Pelusi, F. Facciotti, A. Lai, M. Tarkowski, A. Berzuini, F. Caprioli, L. Santoro, G. Baselli, C. della Ventura, E. Erba, S. Bosari, M. Galli, G. Zehender, and D. Prati. 2020. "SARS-CoV-2 seroprevalence trends in healthy blood donors during the COVID-19 Milan outbreak". *medRxiv – The Preprint Server for Health Sciences*.

Wells, P.M., J.K. Doores, S. Couvreur, R.M. Nunez, J. Seow, C. Graham, S. Acors, N. Kouphou, S.J.D. Neil, R.S. Tedder, P.M. Matos, K. Poulton, M.J. Lista, R.E. Dickenson, H. Sertkaya, T.J.A. Maguire, E.J. Scourfield, R.C.E Bowyer, D. Hart, A. O'Byrne, K.J.A. Steel, O. Hemmings, C. Rosadas, M.O. McClure, J. Capedevilla-Pujol, J. Wolf, S. Ourselin, M.A. Brown, M.H. Malim, T. Spector, and C.J. Steves. 2020. "Estimates of the rate of infection and asymptomatic COVID-19 disease in a population sample from SE England". *The Journal of Infection*, Volume 81, Issue 6: 931-936.

Wu, X., B. Fu, L. Chen, and Y. Feng. 2020. "Serological tests facilitate identification of asymptomatic SARS-CoV-2 infection in Wuhan, China". *Journal of Medical Virology*, Volume 92, Issue 10: 1795-1796.

Yanes-Lane, M., N. Winters, F. Fregonese, M. Bastos, S. Perlman-Arrow, J.R. Campbell, and D. Menzies. 2020. "Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: A systematic review and meta-analysis". *PLoS ONE*, Volume 15(11): e0241536.

Yu, X. 2020. "Modeling return of the epidemic: Impact of population structure, asymptomatic infection, case importation and personal contacts". *Travel Medicine and Infectious Disease*, Volume 37: 101858.