A tentative tracking of the SARS-Cov2 pandemic in France, based on a corrected SIR model including vaccination effects.

Mathilde Varret[1], *François Xavier* Martin[2] and *François* Varret[3*]

[1] Université de Paris, INSERM, Laboratory for Vascular Translational Science (LVTS), F-75018 Paris, France
[2] Member of the Editorial Committee of the Ecole Polytechnique alumni review, 12 rue de Poitiers, 75007 Paris, France
[3] Emeritus Professor, Université de Versailles Saint-Quentin, 45 avenue des Etats-Unis, 78035 Versailles cedex, France

* corresponding author : varret_francois@yahoo.fr

Abstract. We developed successive extensions of the SIR model in order to track the dynamics of the SARS-Cov2 disease. The analysis of health system available data is published in a chronicle accessible on the net: https://corona-circule.github.io/lettres/. This chronicle was initiated on late march 2020 and up to now contains 50 issues. A constant concern was the reliability of the data: for instance, we very soon evidenced that the number of confirmed cases, because of the asymptomatic carriers and the erratic testing policy, was hugely underestimated. By the end of 2020 we made a basic change in the model which consisted in accounting for a constant contagiousness time (SIR-tcc) instead of the probabilistic evolution of the end of the infection assumed so far. Recently we completed this SIR-tcc model for the vaccination effects in order to properly track the evolution of the group immunity threshold. Calculations were performed using the Excel facility (Microsoft), allowing a manual fitting of the model parameters. The results have dealt with a large number of countries, but we focus here on the data regarding France. Further pieces of information are also presented, in order to help elucidating some the factors responsible for the complex history of the pandemic dynamics. (*submitted dec 14th 2021*)

1 The epidemiological context

1.1 The SARS-Cov2 virus

The 21st century saw the emergence of three coronaviruses responsible for deadly severe acute respiratory syndromes in humans (SARS): SARS-CoV [1,2], MERS-CoV [3], and SARS-CoV-2 [4,5] which is causing the COVID-19 disease. As for the first two, the reservoir host for SARS-CoV-2 is the bat [6]. Whereas palm civets [7] and dromedary camels [8] are recognized for the transmission of, respectively, SARS-CoV and MERS-CoV from bats to humans, the SARS-CoV-2 intermediate host is still not known. Like other respiratory pathogens, including influenza and rhinovirus, coronavirus transmission occurs through airborne droplets produced by sneezing, coughing, and breathing [9].

Coronaviruses are ribonucleic acid (RNA) viruses, their genetic material is a single-stranded RNA molecule, carrying four main structural proteins - Spike, envelope, membrane, nucleocapsid - and several non-structural proteins [10]. The spike surface protein of the coronaviruses mediates their entry into host cells through a variety of receptors and entry mechanisms [11]. Like SARS-CoV, SARS-CoV-2 uses the angiotensin-converting enzyme 2 (ACE2) as a receptor to enter cells [12,13]. Inside the host cell, viral RNA is translated into non-structural proteins that supress the host genes expression in favour of the virus ones [14,15]. Then, viral RNA is replicated, translated and new virus particles are formed that will strike other host cells [16]. Infected cell fuses to neighbouring cells expressing ACE2, giving rise to massive individual respiratory cells that prosper for long and produce more and more new virus particles [17]. Beside the rapid virus replication, the innate immune cells response takes place with the secretion of pro-inflammatory cytokines and proteases [18]. This excessive inflammation, called “cytokine storm”, leads to acute respiratory distress syndrome and pulmonary oedema [6]. However, since the virus supresses the host genes expression, including those encoding proteins that alert the immune system, it escapes the immune response of the host.

The consequences of Sars-CoV-2 infection vary greatly from one person to another. While 90% of the infected individuals are not very symptomatic or asymptomatic, 10% develop severe, or even critical, forms with pneumonia requiring an intensive care stay. The viral load is obviously a major risk factor. Age is also a major risk factor for the development of severe forms, with individuals over 65 years having the greatest risk of requiring intensive care [19], and men are more susceptible than women [20]. Nearly 25% of severe forms are due to a deficiency in the type 1 interferon (INF1) pathway, a cytokine usually produced rapidly by the immune system in response to a viral infection and whose main effect is to inhibit virus replication in infected cells). With 3-4% of *INF1* gene alterations [21] and 15-20% of presence of autoantibodies neutralizing IFN1 and blocking its antiviral action [22-23]. The presence of INF1 neutralizing autoantibodies is very rare before age 65 (0.2 to 0.5%) and then increases exponentially with aging to reach 4% between 70 and 79 years, and 7% between 80 and 85 years. The causes and mechanisms of this increase in the general population remain to be elucidated, but it partly explains why age is a major risk factor in the development of severe forms of Covid-19. Defects in the gene encoding the toll-like receptor 7 (TLR7), an RNA intracellular receptor playing a major role in INF1 production, are found in 1.8% of men with severe forms under age 65 [24]. The TLR7 gene being on the X chromosome explains why men are more frequently affected by severe Covid-19 than women.

1.2 SARS-CoV-2 variants

The SARS-CoV-2 virus is constantly mutating and numerous SARS-CoV-2 variants have emerged, naturally selected for their higher transmissibility, lethality or capacity to escape the host's immune response in comparison with the original virus (Table 1). Transmissibility is related to permeability and represents the ability of the virus to enter host cells. Lethality is the proportion of infected individuals who die from the infection in a given place, at a given time. Lethality is to be differentiated from virulence which corresponds to the propensity of the virus to harm its host. Virulence is quantified in the absence of specific care. The same variant will have a different lethality from one country to another depending, for example, on the quality of the hospital system, but its virulence will be unchanged.

Table 1. Effects of the main SARS-CoV-2 variants, in comparison with the original virus, on their transmissibility, lethality, and ability to escape the host immune response [25-28].

Variant	Transmis-sibility	Lethality	Immune escape
Alpha B.1.1.7	43-90% higher	No change	Slightly increased
Beta B.1.351	1.5x higher	Possibility increased fatality rate	Higher
Kappa & Delta B.1.617.1 &2	2.6 x higher	Elevated in Brazil (may be due only to healthcare system overload)	Higher
Gamma P.1	1.7-2.4 x higher	Higher	Higher
Epsilon B.1.427& 429	19-24% higher	Yet not known	Slightly increased

The most recent variant which is in the headlines at the end of 2021, Omicron (B.1.1.529), carries many mutations present in other variants. Of the 32 Omicron mutations in the spike protein, 16 are in the Delta variant. These known mutations seem worrying since they are linked to a greater infectivity and a better ability to escape the immune response. And numerous new mutations seem able to contribute to a further escape ability, and a more efficient virus replication [29-31]. Nevertheless, there is still a lot to understand about the Omicron variant at the biological, clinical, vaccination and epidemiological level.

1.3 Vaccines

In just one year, 322 candidate vaccines, based on the Wuhan virus, had been proposed and 9 were approved for use to adults, and in some cases to adolescents through various region-specific and regulatory agency approval procedures (Table 2).

Table 2. Characteristics of COVID-19 vaccines approved to date [32-33].

Vaccine manufacturer (vaccine name)	Platform	Efficacy (%)	
		From trials	For severity*
Sinovac Biotech (CoronaVac)	Inactivated virus	51	51 SC 100 SD, H
Sinopharm (BBIBP-CorV)	Inactivated virus	78	100 SD, H
AstraZeneca–University of Oxford (AZD1222)	Viral vector	81	100 H
Johnson & Johnson (Ad26.COV2-S)	Viral vector	66	85 SD
Gamaleya (Sputnik V)	Viral vector	92	-
Bharat Biotech (Covaxi)	Viral vector	78	100 H
Pfizer–BioNTech (BNT162b2)	mRNA	95	100 SD
Moderna (mRNA-1273)	mRNA	94	100 SD
VECTOR (EpiVacCorona, NCT04780035)	Protein subunit	-	-

* SD: severe disease, H: hospitalization, SC: symptomatic COVID-19 disease.

Three major vaccine technology platforms were exploited. In its most traditional form, the principle of a vaccine is to inoculate the virus against which we intend to prepare the immune system. To prevent the virus from causing disease, it will first be either inactivated (like the flu and polio vaccines) or attenuated (like the pertussis and tuberculosis vaccines). Viral vector and mRNA vaccines are of the so-called gene type. These technologies consist of inoculating the whole or parts of the virus genetic material, DNA or mRNA, so that the vaccinated organism himself will produce the immunogenic proteins. The carrier, or vector, is generally a benign virus, adenovirus, which will stimulate the immune response while remaining harmless. The protein subunit vaccine technology uses three peptides of the spike protein planted on a chimeric protein constituted of two parts (viral nucleocapsid protein and bacterial maltose-binding protein). The vaccinated organism will respond by making antibodies against the Spike protein. This technology has already been used, for the hepatitis-B vaccine by example.

Efficacy against symptomatic COVID-19 disease is calculated from end points defined in the clinical trial, severe disease, hospitalization, and/or symptomatic COVID-19 disease, which differ between trials, even for the symptom's definition (Table 2). It is an important value for the approval of vaccines, but does not reflect the real-world impact, especially when trials have been performed on specific population (younger, healthy adults, individuals at risk of severe disease), and in the absence of the more recently reported SARS-CoV-2 variants. It is thus essential to know the extent and duration of protection across all age groups and populations. Especially given the higher risk of severe form in the elderly (age > 70). The efficacy may be affected by various factors including the population, the vaccine, the vaccination schedules and the handling/administration of the vaccine.

Most of these vaccines were developed against the Wuhan virus and showed an efficacity between 50% and 95% (Table 3). Since numerous SARS-CoV-2 variants have emerged and continue to emerge, characterizing their impact on vaccines efficacy quickly became necessary. Their efficacy against the dominant variant today, Delta, remains very good (Table 3).

Table 3. COVID-19 vaccines efficacy according to SARS-CoV-2 variants [32-34].

Vaccine manufacturer (vaccine name)	Efficacy (%) against variant				
	Wuhan strain	Alpha	Beta	Gam-ma	Delta
Sinovac Biotech (CoronaVac)	50 70	-	-	-	59
Sinopharm (BBIBP-CorV)	78	-	-	-	-
AstraZeneca–University of Oxford (AZD1222)	55 81	75	10	78	92
Johnson & Johnson (Ad26.COV2-S)	66	70	57 72	68	-
Gamaleya (Sputnik V)	-	-	-	-	90
Bharat Biotech (Covaxi)	-	-	-	-	65
Pfizer–BioNTech (BNT162b2)	95	94	86	51	88
Moderna (mRNA-1273)	94	100	96	-	92

While vaccines are reported to reduce the number of symptomatic COVID-19 cases, direct evidence for their effect on virus transmission is limited. The BNT162b2 (Pfizer–BioNTech) vaccine has been shown to reduce the viral load [35]. And since a reduced viral load is associated with decreased transmission of the virus, these data together suggest that vaccination may reduce transmission.

1.4 Data and websites

Basic data for the number of confirmed cases, deceased persons and vaccine doses can be found on the various websites of National Health Systems [36]. Some sites update worldwide compilations [37, 38], while others built-up useful figures or maps which provide synthetic views of the evolution of the pandemic [39, 40]. Among them, the CovidTracker site [39] provides very easy-to-read figures of crucial effects such as the age effect which was the constant background of our investigation.

Prior to any analysis it may be noted that all reliable indicators (some are not!) of the pandemic are not necessarily sensible. For example, the number of removed people is not known in real time. Consequently, the number of infectious people at a given date is not known. Only the total number of infected persons, that is, which have been contaminated in the past, is provided by the health systems. Infected persons are usually termed "cases", and the recognition of a new case is based on the positive result of a PCR test. However, the number of cases is obviously underestimated due to asymptomatic carriers. This gives rise to the famous Trump theorem [40] "*more covid19 testing creates more cases*". On the contrary the number of reported deaths assigned to the COVID-19 appears to be (for most countries) reliable.

On the example of France, we could check a good compliance between the number of deaths quoted during the first pandemic and the seasonal excess mortality. [CC 4]

A model is used here for deriving correlations between the available sensible data, and eventually deriving the non-available ones, in order to elucidate some of the basic factors of the pandemic. The reverse way consists in using models for predicting the effects of possible factors. Such models may be extremely complex, they mostly deal with inhomogeneous effects and require the efforts of large groups or researchers [42-46]. Their conclusions were often used as guidelines for the present approach which modestly remained in the frame of the mean-field approximation.

The analysis of health system available data is published in a chronicle accessible on the net: https://corona-circule.github.io/lettres/. For convenience the chronicle issues are abbreviated here "CC n…" where n is the issue number

2 The SIR model

2.1 The uncorrected SIR model

The historical SIR model, where S, I, R respectively stand for susceptible, infectious and removed people, was created by mathematicians almost one century ago [47]. It was widely used till now through its mean-field approach, which actually was the "special case" for which an analytical solution could be given, see [48] for a recent description (restricted to the special case). For convenience we call this special case "the uncorrected SIR model". The population is considered as a 3-state system which irreversibly evolves from the unstable initial state S to the stable state R through the intermediate state I.

2.1.1 Description of the uncorrected SIR model

In the mean-field approach, S(t), I(t), R(t) are the time dependences of the population fractions associated with these 3 states. The probabilistic approach is based on transition rates β and γ for the successive transitions S $\rightarrow$ I and I $\rightarrow$ R respectively, see figure 1. The peculiarity of the pandemic mechanism is that the flux S(t) $\rightarrow$ I(t) is proportional to the infectious fraction. This gives rises a self-accelerated dynamics at the beginning of the pandemic, a well-known phenomenon in the field of population dynamics.

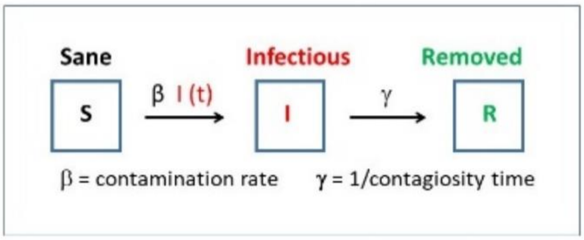

Fig. 1. A schematic view of the uncorrected SIR model.

β the contamination rate is defined as the probability per unit time for a sane person to be contaminated due to meeting a contagious one. This crucial parameter depends on the intensity/frequency of contacts between people as a well upon the intrinsic contagiousness of the virus. It is strongly impacted by the irruption of virus mutants (the so-called variants) and by the level of social constraints (lock-in, curfew, travel or activity restrictions, masks...). For these reasons it is essentially time dependent. The adjustment of β (t) to the sanitary data (deceases, hospitalization, contamination cases) is a must for any sensible tracking of the pandemic.

In the traditional approach **γ** is the probability per time unit for a given infectious person to reach the removed state, dead or recovered. It is the reverse of the average duration of the contagious state (or contagiousness time).

The master equation of the SIR system follows:

$$dS/dt = -\beta S I\;;\;\; dI/dt = \beta S I - \gamma I\;;\;\; dR/dt = \gamma I \qquad (1)$$

such as $dS/dt + dI/dt + dR/dt = 0$

because $S(t) + I(t) + R(t) = 1$

The set of coupled differential equations (1) is easily solved by using an Excel spreadsheet (Microsoft Office), where time t is increased by steps of one day. The adjustment is performed manually thanks to the real-time display of the calculated curves. Our first calculation, dated 2020/03/31, see Figure 2, led to claim that the top of the first wave was nearly reached. The peak of daily cases actually occurred 1 – 2 weeks later.

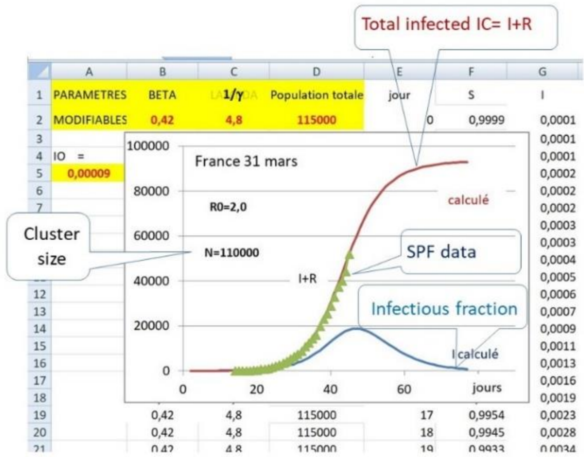

Fig. 2. An early adjustment using the uncorrected SIR model based on the total case data for France (adapted from [CC 2]). IC is the fraction associated with the total number of cases

In this preliminary calculation, all parameters were kept constant. The cluster size, β and ϒ were both adjusted. Later on, β was taken time-dependent, and γ was given a better value 1/14 days suggested by health system websites.

A crucial problem was to properly assign the date of the beginning of the pandemics. In principle the use of continuous functions is not consistent with the existence of a "patient zero". In addition, the date of the first contamination was unknown – and till now remains so. In practice, this problem is solved by arbitrarily fixing the first day of the pandemics, and adjusting the initial (non-null) value of the infectious fraction. <u>Here, day 1 was arbitrarily fixed March 1, 2020.</u>

2.1.2 Initial dynamics

The initial dynamics of the pandemic is exponential. Indeed for small t:

$$S(t) \sim 1 \text{ and } I(t) \sim 0 \quad \Rightarrow \quad dI/dt \sim (\beta - \gamma)\, I$$
$$\Rightarrow \quad I(t) \sim I(0) \exp\{\, t / (\beta - \gamma)\,\}$$

The exponential is growing for $\beta > \gamma$ and decreasing in the opposite case.

2.1.3 Reproduction rates

<u>The initial reproduction rate R_0</u> is the average number of persons infected by each infectious person at the beginning of the pandemic. $\mathbf{R_0 = \beta/\gamma}$, that is, the contamination rate x the contagiousness time. For $R_0 > 1$ the exponential speeds up, for $R_0 < 1$ it slows down, in agreement with sec. 2.1.2.

The value of R_0 depends on both the virus contagiousness (which depends on the variant) and on the frequency of contacts between persons. Social restrictions tend to decrease the reproduction rate, while the irruption of each more contagious variant tends to increase it. The balance between these opposite effects is crucial and has to be accurately followed.

However, the most popular rate is the effective reproduction rate $R_{eff}(t) = R_0(t) \times S(t)$, which through eqn. (1) governs the dynamics of the pandemic at time t. For $R_{eff}(t) > 1$ the pandemic grows, for $R_{eff}(t) < 1$ it shrinks. The tracking of $R_{eff}(t)$ is a central problem (see [49] for a review of the methods). Santé Publique France quotes the Pasteur Institute results obtained through the Cori method [50]; a very simple - approximate - method was recently introduced by the CovidTracker site [39].

In our approach $R_{eff}(t)$ is an output of the data analysis, it is derived from the fitted sequence of the $\beta(t)$ values.

2.1.4 Group immunity

Group immunity is a state expected when the effective rate $R_{eff}(t) = R_0(t) \times S(t)$ becomes smaller than the threshold value 1 due to the irreversible decrease of S(t). For a constant R_0 the behavior is clear and once the threshold value $S_{thres} = 1/R_0$ is reached, the pandemic irreversibly slows down and the infected fraction tends to saturate significantly below the ultimate limit 1. For a time-dependent $R_0(t)$ the behavior is far from clear and group immunity should be better understood as a condition which is (temporarily!) fulfilled when S(t) is smaller than the threshold value $1/R_0(t)$. In absence of vaccination effects this is strictly equivalent to $R_{eff}(t) < 1$. The additional immunity created by large scale vaccination will make it different (see Sec. **2**.4).

2.2 Adaptation of the uncorrected SIR model

2.2.1 Hidden part and visibility ratio

A first adaptation was required by to the large number of asymptomatic carriers, which was first recognized by the Pasteur Institute in April 2020 [51], and led us an analysis exclusively based on the decease data.

We therefore completed the master equation (1), with the deceased fraction, $D(t) = \ell R(t)$, (2)
where ℓ is the lethality ratio, that is, the probability for a contaminated person - non-vaccinated - to decease. This lethality ratio is strongly age-dependent, it drastically increases as a function of age.

Meanwhile, we observed that the adjusted size of the cluster was continuously increased at each new analysis; we were also puzzled by the incredible diversity of the lethality rates reported for various countries. We thought that this diversity was due to the existence of asymptomatic carriers and introduced the idea of the "**hidden part**" of the pandemics [CC 4],

Due to the presence of this hidden part, we changed the basic assumptions of the analysis, as follows: (i) the case data were corrected by an adjustable constant factor which is the **visibility ratio**" of the pandemics quoted in the previous section; (ii) the lethality ratio was initially fixed: ℓ =0.53 % in agreement with a detailed report of the Pasteur Institute [52], and later on 1.06% [CC 23] according to a more extended study [53] ; (iii) the total population (France = 67 million inhabitants) was accounted for; (iv) last, but not least, the contamination rate $\beta(t)$ was no longer a constant. To avoid handling too many free parameters, we used a stepped $\beta(t)$ function with 10-day steps (5-day, later on).

We show in Figure 3 an example of such an analysis when the first wave was just over. The value of the visibility ratio was adjusted so as to match both the case and death data.

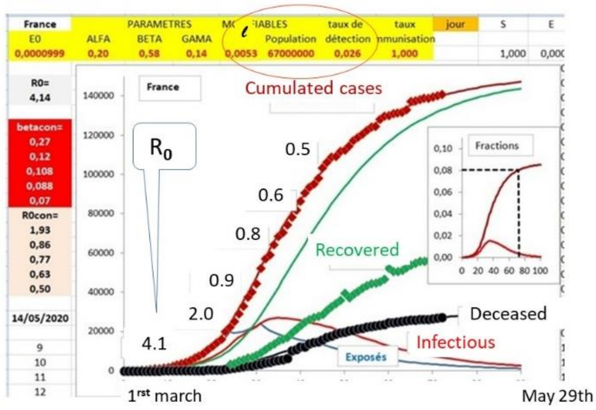

Fig. 3. The first wave in France, dated late May 2020, adapted from [CC 4]. Complete lock-in was effective since mid-march.

The main result shown in Figure 3 is the decrease of the reproduction rate, associated with the decay of the first wave. This decay is due to the complete lock-in from mid-March to mid-May. The visibility ratio, taken here as a constant, is extremely small, a few % only, and was confirmed by the detailed report of the Pasteur Institute in May 2020 [52].

When studying the following phase of the pandemic, we found that such a simultaneous fit of the death and case data was no longer possible. Indeed, the value of the visibility ratio irreversibly increased for each new wave. We then decided to fit a single set of data; to start with, the decease data, because they were consistent with the seasonal mortality data (Sec. 1.4). Then the visibility ratio became another output of the analysis, time dependent, defined as the total number of confirmed cases divided by the calculated total number of infected persons. This ratio presently reaches around 66 % (detailed data in [CC 43]).

2.3 The corrected SIR-tcc model

By the end of the year 2020 our attention was drawn to a contribution to the French Ecole Polytechnique alumni review [54] which pointed out the lack of relevance of the uncorrected SIR model, where the recovery from (or death after) the contamination of a given person is a random process governed by a transition rate. The main drawback of this assumption is that the recovering probability of a given person is maximum on the very day of contamination and smoothly decreases as a function of time. This is totally against the idea of an average recovery time, which is commonly admitted in the medicine field (for instance, around 14 days for the covid-19 disease).

It is very surprising that the obvious deterministic character of I → R process was clearly accounted for so lately (the model is 90-years old). For our part, we turned to the duly corrected version and termed it SIR-tcc (for constant contagiousness time).

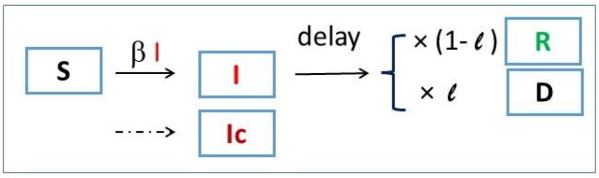

Fig. 4. The corrected model (SIR-tcc). Ic stands for the total infected fraction

The change in the spreadsheet organization was straightforward [CC 23] by defining delayed output fluxes from I(t) to R(t) and D(t), which merely are the input flux at time (t - delay), multiplied by the respective (1 - ℓ) and ℓ factors. The delay value was taken in continuity of the previous model, that is, 1/ϒ, same for all and constant in time (15 days).

In practice, the output flux was taken as a 7-day average value in order to smooth eventual discontinuities generated by the model [CC 24].

2.3.1 Comparison to the previous model

As evidenced by one of us (FXM) [55], the uncorrected SIR model provides a too weak and delayed response to the variations of the β-parameter.

If β increases, the uncorrected SIR model equations lead to overestimate the number of recoveries and deaths, as they do not take in account the delay between the S → I and I → R processes for an given individual. As a consequence, the SIR computed number of infectious is too small.

For the same reason, if β decreases, the uncorrected SIR model equations lead to underestimate the number of recoveries and deaths: the computed number of infectious is too high.

The uncorrected SIR model is not reactive enough to β variations. No doubt the SIR-tcc model definitely has to be preferred.

Accordingly, the analysis of the data using the SIR-tcc model will obviously provide results different from those of the uncorrected model. This is illustrated in Figure 5.

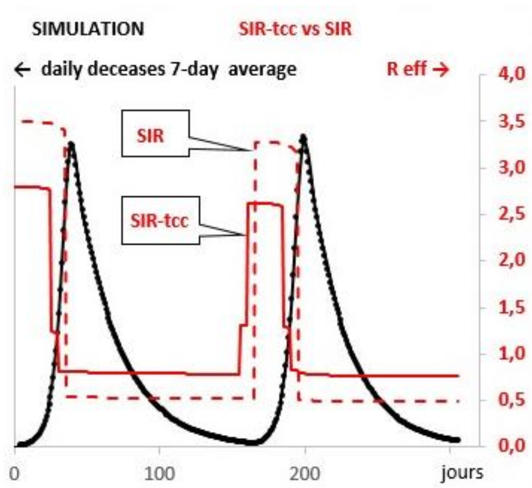

Fig. 5. Comparison of the $R_{eff}(t)$ curves derived from obtained with the uncorrected SIR and SIR-tcc models, for the same set of (simulated) data.

For convenience the data shown in Figure 5 were generated using the uncorrected model, with a schematic squared-shaped β(t) involving two contamination waves. The asymmetric shape of the resulting peaks is worth noting and contrasts with the symmetrical shape obtained with the corrected model (not shown here).

The responses of the SIR and SIR-tcc analyses of these simulated data, as shown in Figure 5, are sizably different: the uncorrected model analysis overestimates the fitted β(t) curve – with respect to the corrected model analysis - so as to compensate its weaker response to β-variations. The uncorrected model analysis delays the fitted β(t) curve, a far from intuitive consequence of its delayed response to β(t) variations. The latter effect is larger on the decreasing side of the peaks.

In addition, we found that the manual adjustment procedure was much easier with the SIR-tcc model. This might be a consequence of the endless character of the recovery process, which induces strong mutual influences between the values of the adjustable parameters during the fitting procedure.

2.3.2 Results for France

The results for France, recently updated, are shown below:

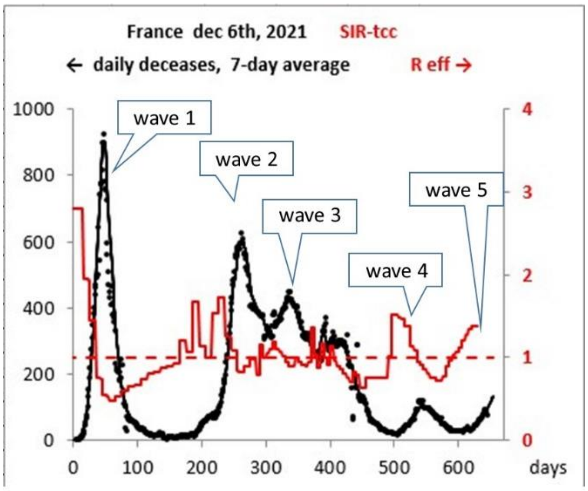

Fig. 6. The successive 5 waves of the pandemic in France, analysed through the SIR-tcc model, adapted from [CC 43].

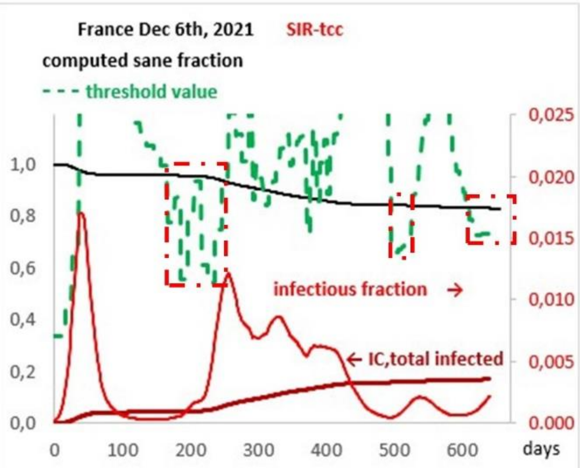

Fig. 7. The compared evolutions of the computed fraction S(t) (solid black line) and of its group immunity threshold (broken green line), for France. The situation of group immunity has been reached several times, when the solid black line was below the slashed green one; the situations where group immunity was not reached are been pointed out by red dashed frames. Adapted from [CC 43].

2.3.3 Alternative indicators

Alternative indicators such as the daily number of confirmed cases, hospitalisations or admissions to the intense care sections may be used. These data presumably are almost as reliable as the decease ones. We have treated the hospitalization data in France, in the same way we did for the decease data. The lethality ratio merely is substituted by a **severity ratio** assumed to be time-independent and the same for all. The value given to this severity ratio (4 %) was determined in order to yield visibility ratio values close to those obtained by the decease-based analysis. [CC 30]

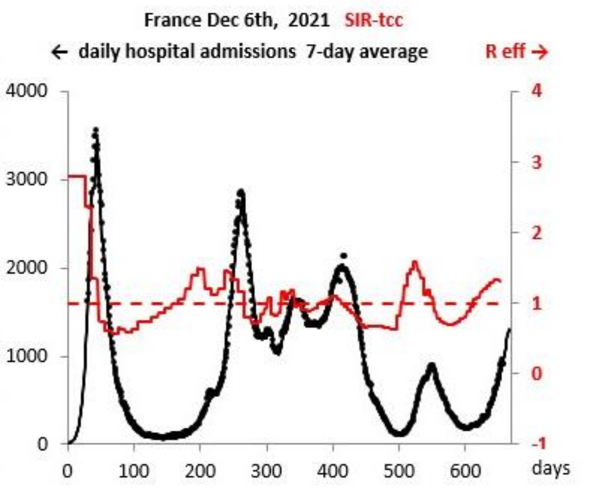

Fig. 8. An independent analysis based on the daily hospitalization data of France. Note that the $R_{eff}(t)$ curve is similar to that obtained from the decease data (figure 6). Updated from [CC 30].

The R_{eff} (t) curves displayed on figures 6 and 8 are similar. A complete comparison of the curves associated with the possible 3 indicators (cases, hospitalizations, deaths) will be presented in Sec. 3, Figure 16.

2.3.4 Apparent lethality and severity ratios

Daily hospitalization and death data may be directly compared to the case data, provided that a convenient delay is introduced. We denoted these ratios "apparent" because they refer to an indicator – the number of cases – which is subject to the absence of asymptomatic cases. The apparent lethality ratio is the number of deaths on day d, divided by the number of cases confirmed on day d – 14 (elsewhere denoted *the case fatality rate*). The apparent severity ratio is the number of persons hospitalized on day d, divided by the number of cases confirmed on day d – 2 (to start with: confirmed on day d-14, but this excessive delay value was corrected later on [CC 43]). Both ratios are impacted by the intensity of the testing campaign.

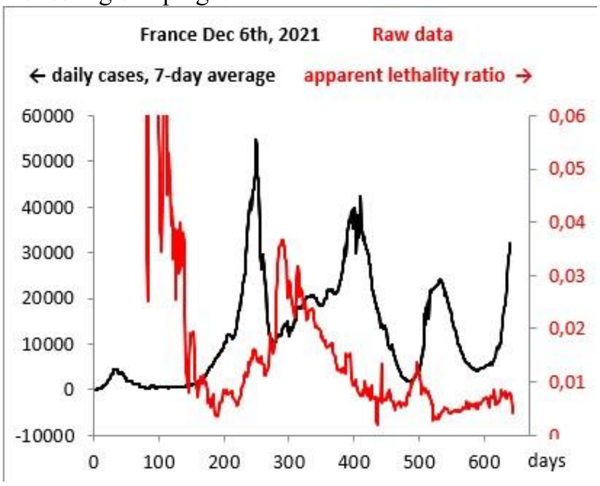

Fig. 9. The evolution of the apparent lethality ratio in France. The large values prior to day 200 are attributed to the sufficient number of tests. The increase in the ratio after day 200 is assigned to the onset of variants (Beta, and later on Delta). The sizeable decrease after day 300 reflects the efficiency of a massive vaccination campaign targeted to the most fragile or eldest persons. Updated from [CC 30].

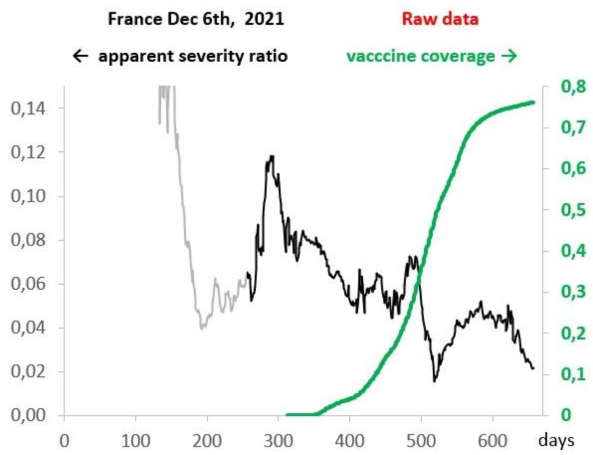

Fig. 10. The evolution of the apparent severity ratio in France. The global decrease after day 300 confirms the efficiency of vaccination reported in the previous figure. Additional features around days 300 and 480 might be assigned to concomitant weaknesses of the testing activity. Adapted and updated from [CC 30].

2.4 The SIR-tcc-vaccination extension

Worldwide vaccination campaigns started on the very end of 2020. The campaign was rather massive in France and reached a high coverage value (number of vaccinated persons / total population) close to 75 %, see Figure 10 (in previous section).

It appeared that this campaign has significantly hindered the larger contagiousness (by 50 %) of the so-called British (Alpha) variant with respect to the initial virus. In terms of protection against severe forms of the disease leading to intensive care sections and ultimately, to death, efficiency ratios up to 90% or more were reported, see Sec. 1.3. But, for a long time, little was known about the protection against contamination. Thanks to a detailed study of the correlation between the vaccinal coverage as a function of age and the relative intensity of the fourth wave [CC 42] we reached the conclusion that the protection against contamination was lesser, say 60-70 % smaller. It seems that the reduction of the viral load due to vaccination is the mechanism which leads to the large decrease in the probability of severe forms and death.

The model extension basically consisted in creating a second sane fraction for the vaccinated persons, and splitting the infectious fraction according to their protection or not against severe forms and death, see figure 11.

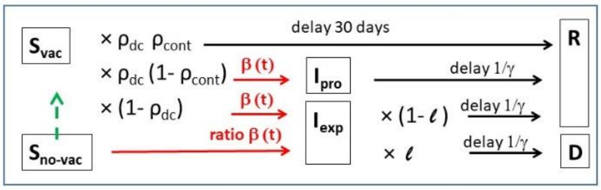

Fig. 11. Scheme of SIR-tcc-vaccination model. S_{vac}, S_{novac}, respectively stand for the vaccinated or non-vaccinated sane fractions, I_{pro}, I_{exp} for the protected or exposed infectious fractions, and ρ_{dc}, ρ_{cont} for the protection ratios against decease and against contamination.

The flux $S_{no\text{-}vac} \rightarrow S_{vac} = S_{no\_vac}(t) \times d\, Couv(t)/dt$ (3)

where Couv (t) = is the vaccinal coverage (2-doses here). We assume for simplicity the same contamination rate β(t) for both I_{pro}, I_{exp}. The equations follow straightforwardly and can be found in [CC 43].

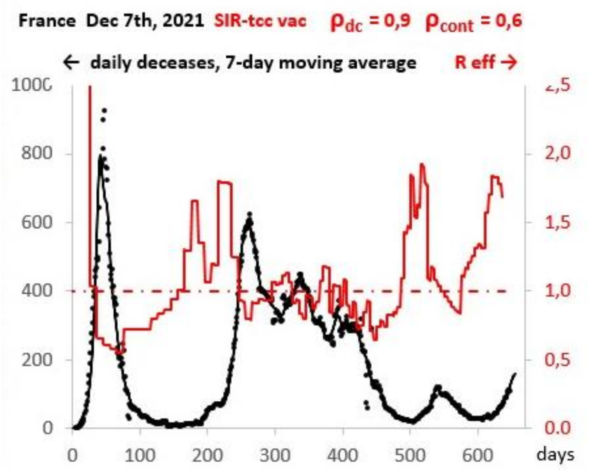

Fig. 12. The France deceased data, analysed through the SIR-tcc-vaccination model, adapted from [CC 43]. To be compared to Figure 6.

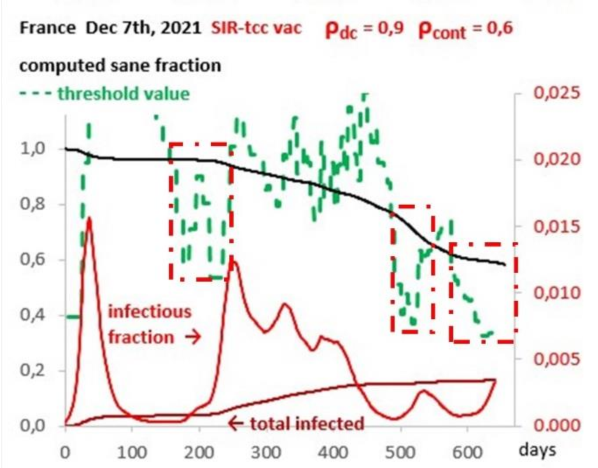

Fig. 13. The sane fraction S(t) (solid black line) and of its group immunity threshold (broken green line), obtained through the SIR-tcc-vaccination model. Adapted from [CC 43]. To be compared to Figure 7. It is worth noting that the infected fraction I(t) presently is in large excess of the peak of the previous wave.

An obvious change provided by the SIR-tcc-vac model is the lowering of both the S(t) and threshold value curves (compare Figures 7, 13). The combination of both effects finally leads to hinder the condition of group immunity. In other words, the analysis discarding vaccination effect sizably **overestimates** the possibility of reaching a future group immunity.

The $R_{eff}(t)$ curve reported in Figure 12 only shows minor differences with respect to Figure 6. These differences presumably are due to the imperfect character of the manual adjustment procedure. An automatic fitting procedure might be helpful here.

On the contrary, the $R_0(t)$ values are visibly increased by the account of the vaccination effect, see Figure 14.

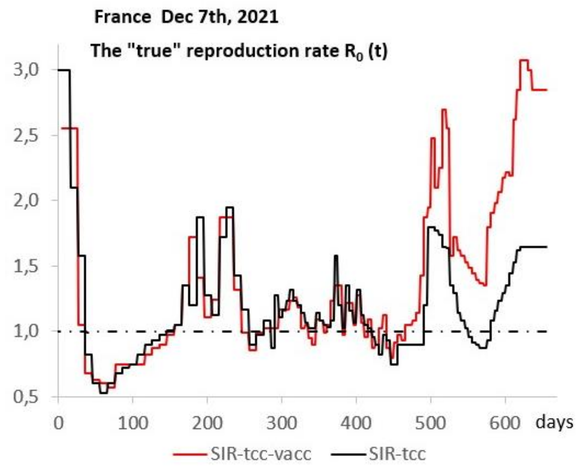

Fig. 14. The "true" reproduction rate, $R_0(t)$ obtained through the SIR-tcc and SIR-tcc-vaccination models, for France, adapted from [CC 43].

<u>$R_0(t)$, corrected for the vaccination effect, is presently larger than it was at the beginning of the pandemics</u>**.** <u>This statement means that the present social restrictions, however quite severe, hardly balance the contagiousness excess of the present variant (δ).</u>

3 Discussion

Numerous research works are presently carried out by Universities or Institutes. They are mainly based on sophisticated models which aim to reveal all various factors which may govern the dynamics of the pandemic. Understanding these factors is needed to make sensible predictions of the long-term evolution of the pandemic.

Our approach is slightly less ambitious, it only allows short-term extrapolations. However, the analysis of the past evolution, like as any historic research work, provides *a posteriori* measurements of the effect of social decisions such as lock-in, curfew, vaccination, sanitary pass… but this is a sociologic matter which is not in the very scope of the present work.

3.1.1 The CovidTracker website approach

Meanwhile, the huge amount of data was processed in several websites [39, 40], so as to provide an easier access to the pandemic evolution through maps, graphs and figures. The CovidTracker site [39] reports an alternative determination of the effective reproduction rate, see Figure 15:

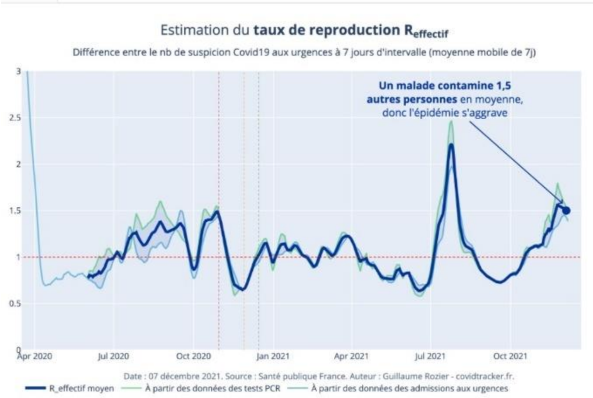

Fig. 15. The effective reproduction rate, derived from the relative variations of indicators (case, hospitalisations), after [39] dated 2021/12/08.

Here, $R_{eff}(t)$ is assumed to be the relative variation of any indicator over 7 days (of course using 7-day moving averages). The relevance of this model-free method was a posteriori justified by the close compliance with the values quoted by Santé Publique France [36], that is, determined according to the Cori method [50] by the Pasteur Institute simulation group, see for example [56]. These results are also in good compliance with those of the present analysis. This simple approach required, of course, the tuning of the time interval over which the relative variation is calculated. We noticed that this 7-day interval exactly matches half the contagiousness time of the SIR-tcc model (14 days). A sensible explanation is that a given contagious person may have been contaminated by a previous one all along its contagiousness time, that is, after 7 days in average [CC 40].

Such a model-free determination of $R_{eff}(t)$ avoids the hazardous step of manual fitting. We used it to illustrate the various time shifts associated with the possible indicators, see Figure 16.

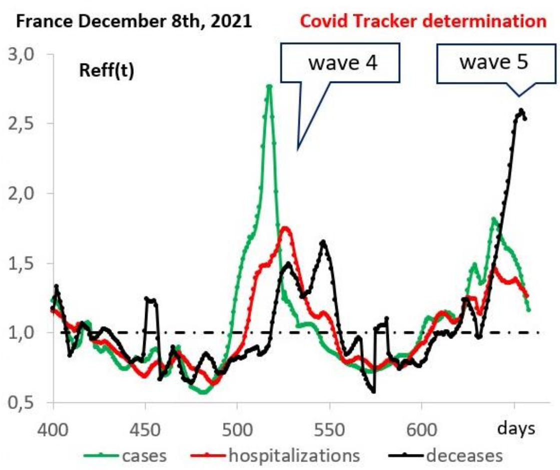

Fig. 16. An example of $R_{eff}(t)$ variations determined from the CovidTracker method. The hospitalization peak is clearly delayed with respect to the case peak, by roughly 6-7 days. The decease peak - split - occurs obviously later on. Adapted from [CC 41]

3.1.2 The impact of age

Since the very beginning, it was observed that most of deceases were those of aged persons. The younger persons were thought to be almost free from severe forms, and even from contamination. This is no longer true, and presently the number of cases drastically increases in the young population. We have followed the evolution of the average age for the various steps of the disease, see figures 17, 18, adapted from [CC 39].

.

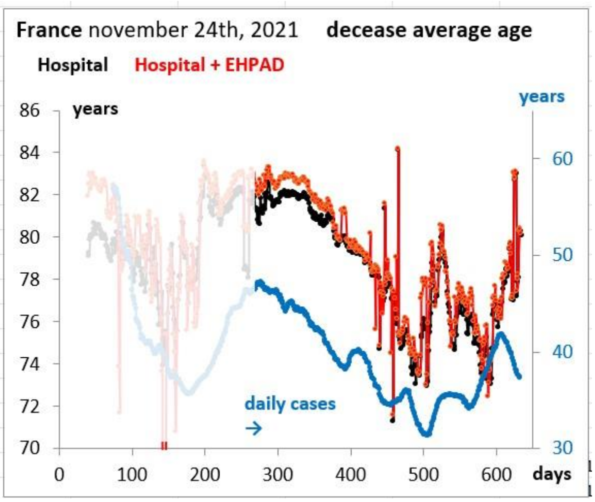

Fig. 17. Evolution of the average ages for contamination and decease. EHPAD stands for Hospitals for Aged or Dependent Person.

The decrease in the average decease age, between days 300-500 is assigned to the impact of the vaccination campaign targeted to the eldest population. The final increase may reveal a sizable weakening of the vaccinal protection, which the reason why a third vaccine dose was decided in most countries. Note that that the case curve also exhibited a strong increase (weakening of the vaccinal protection against contamination), before a final decrease which may reveal a recent change in the virus circulation, which now is extremely active in the younger population.

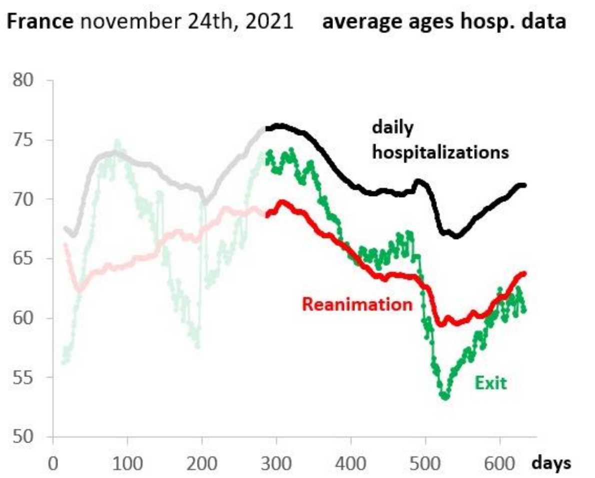

Fig. 18. Evolution of the average ages for hospitalization, admission in reanimation care service, and exit.

The recent increase, over the last two months, shows that the present (5th) wave again induces severe forms of the covid disease in the eldest population.

The different behaviours of the various age ranges are conveniently illustrated in the CovidTracker site, see Figure 19.

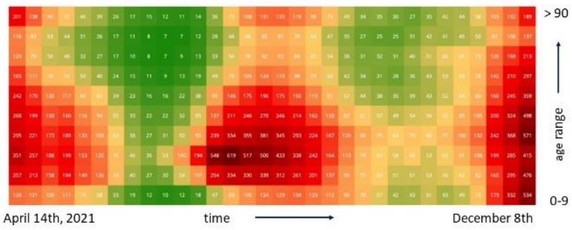

Fig. 19. Evolution of the incidence rate (number of new cases over the last 7 days per 100 thousand persons), as a function of time (horizontal scale), for the various age ranges (vertical scale). The colour scale spans from green to dark red, on increasing values of the incidence rate. Adapted from [39]

Figure 9 covers waves 3, 4 and 5, which are featured by dark areas. Remarkably wave 4 is peaked on the 20-29 range, while the other waves have a wider extension. The peculiar character of the 4th wave is interpreted as a consequence of the unachieved character of the vaccination campaign when wave 4 began. Details are given below. On the contrary wave 5 is widespread over all ages. The intense contamination of the 0–9-range is a true concern because this range is not yet allowed for vaccination in our country. It is worth mentioning that several countries already allowed it.

We have more precisely documented [CC 42] the effect of the vaccination coverage at the beginning of wave 4 by comparing the data of cases, hospitalizations and deceases, cumulated over two periods: period 1, before the beginning of the vaccination campaign, that is, before day 300, and period 2 associated with wave 4, that is, from day 488 to day 579. The ratios of the data period 2 / period 1 are plotted in figure 20 for each age range, as a function of the vaccine coverage on day 300. As the testing activity was extremely different in the various age ranges, we corrected the case ratio, by dividing it by the ratio of the total test numbers in each age range.

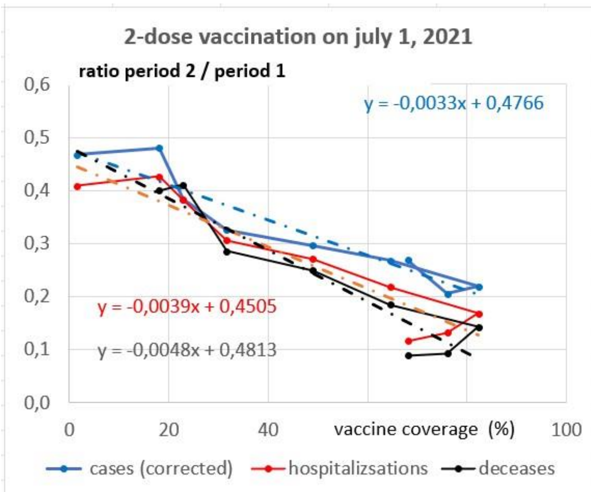

Fig. 20. The ratio of cumulated data over periods 1 and 2 (see Figure 19), plotted versus the vaccinal coverage of the various

age ranges on the beginning of period 2, with the fit of their linear correlations. Adapted from [CC 42]

A clear dependence upon the age is obtained for all indicators, which suggests a linear correlation, except for the oldest two age ranges. A possible reason for the peculiar behaviour of the upper age tranches might be a reinforcement of the sanitary protection, with respect to wave 1; indeed, on the beginning of the pandemic, mortality was extremely high in the medical nursing homes (EHPAD), presumably because of an obvious lack of knowledge and experience in front of an unknown disease.

Whatsoever, the linear fits of the curves of Figure 20 leads to correlation slopes which shows that vaccination provides different degrees of protection against contamination (case data), severe forms (hospitalization data) and decease. In particular, the vaccination efficiency against contamination reveals to be approximately 25 % smaller than against severe forms and decease. Since efficacy against decease and severe forms is close to 90 % (sec. 1.3); against contamination it should be in the range 60-70 %. This value remains to be confirmed by real life data.

3.1.3 Summary of the main results

Important features were revealed by the present analysis which combined model improvements and a critical analysis of "experimental" data. The SIR model was corrected for a more sensible approach of the I → R process. We recently completed it in order to account for vaccinal immunity against severe forms and decease, and independently against contagiousness. The application to the French data showed the reality of this immunity, at different protection degrees. We also showed that presently - on the manuscript submission date - the group immunity condition is far from being fulfilled.

We stress on the need for a critical analysis of the data (this holds any kind of experimental data). The most useful result of our investigation may be the recognition of the hidden part of the pandemic, leading to bias any analysis exclusively based on the case data (in other words, based on the incidence ratio). We preferred a decease-based analysis, casually completed by analyses based on hospitalisation data.

The main outputs of our calculations were the reproduction rates: the effective rate $R_{eff}(t)$ characterizes the evolution of the dynamics while the "true" $R_0(t)$ reveals the combined evolutions of the (average) individual behaviour and of the virulence of the prominent variant.

We pointed out the interest of additional information derived from raw data, for instance in the terms of "apparent severity" and "apparent lethality ratios", we introduced.

A less intuitive comparison of different periods of the pandemic allowed drawing correlations between the impact of the pandemic and the degree of vaccinal coverage, on the occasion of wave 4 which occurred at a time when the coverage degree was strongly age-dependent.

We have provided figures which illustrate the successive waves of the pandemic. We did not pretend to explain their detailed features, obviously impacted by the various regulations (lock-in, curfew, travel restrictions, homework, closure of non-essential activities, sanitary pass …). Such explanations require a detailed modelling which is only tractable by large institutions.

4 Conclusion

As reported in the previous section (**3**.1.3) the simple SIR mean-field model, corrected and completed for vaccinal immunity, has provided interesting results upon the average dynamics of the pandemic at the national level. However, we have used an extremely small part of the available data, which are worldwide and declined according multiple criteria. We have often analysed the data of foreign countries and twice those of the age ranges in France [CC 22-23, CC 33]. In all cases each country or age range was treated independently from the others.

A generalized approach involving several countries or age ranges would require introducing the possible interactions between them. Obviously, the resolution of such a problem with input parameters in the form of a β_{ij} (t) matrix is not tractable by the means of Excel spreadsheets. *This consequently is another story.*

As for us, modesty is required. We intend to continue the current chronicle, accessible as usual, as long as the COVID-19 disease is not mastered. Presently, the growth rhythm of wave 5 seems to slow down. However, the Omicron variant transmissibility, lethality, and vaccines response are not well characterised.

Acknowledgements are due to our friends: Henri Lorain and Professor Didier Astruc, for their stimulating discussions and advices.

References

1. C Drosten , S Günther, W Preiser, et al. *N Engl J Med*. **348**, 1967–1976 (2003).
2. TG Ksiazek, D Erdman, CS Goldsmith, et al. *N Engl J Med*. **348**, 1953–1966 (2003).
3. AM Zaki, S van Boheemen, TM Bestebroer, et al. *N Engl J Med*. **367**, 1814–1820 (2012).
4. N Zhu, D Zhang, W Wang, Li X., et al. *N Engl J Med*. **382**, 727–733 (2020).
5. C Huang, Y Wang, MD, X Li, et al. *Lancet* **395** (10223), 497–506 (2020).
6. P Zhou, XL Yang, XG Wang, et al. *Nature* **579** (7798), 270–273 (2020).
7. B Kan, M Wang, H Jing, et al. *J. Virol*. **79**, 11892–11900 (2005).
8. BL Haagmans, SH Al Dhahiry, CB Reusken, et al. *Lancet Infect. Dis*. **14**, 140–145 (2014).

9. CC Wang, KA Prather, J Sznitman, et al. *Science* **373** (6558), eabd9149 (2021).
10. JFW Chan, KH Kok, Z Zhu, et al. *Emerg Microbes Infect*. **9**, 221–236 (2020).
11. MA Tortorici & D Veesler. *Adv. Virus Res*. **105**, 93–116 (2019).
12. R Yan, Y Zhang, Y Li, et al. *Science* **367**, 1444–1448 (2020).
13. AC Walls, YJ Park, MA Tortorici, et al. C*ell* **180**, 181–292 (2020).
14. Y Finkel, A Gluck, A Nachshon, et al. *Nature* **594** (7862), 240-245 (2021).
15. K Zang, L Miorin, T Makio, et al. *Sci Adv*. **7** (6), eabe7386 (2021).
16. M Scudellari. *Nature* **595**(7869):640–644. (2021).
17. L Braga, H Ali, I Secco, et al. *Nature* **594** (7861), 88-93 (2021).
18. MZ Tay, CM Poh, L Rénia, et al. *Nat Rev Immunol*. **20**, 363–374 (2020).
19. M O'Driscoll, GRD Santos, L Wang, et al. *Nature* **590**, 140–145 (2021).
20. P. Brodin. *Nat. Med*. **27**, 28–33 (2021).
21. Q Zhang, P Bastard, Z Liu, et al. *Science* **370** (6515), eabd4570. (2020)
22. P Bastard, LB Rosent, Q Zhang, et al. *Science* **370** (6515), eabd4585. (2020)
23. P Bastard, A Gervais, T Le Voyer, et al. *Sci Immunol*. **6** (62), eabl4340. (2021)
24. T Asano, B Boisson, F Onodi, et al. *Sci Immunol*. **6**(62): eabl4348. (2021)
25. E Boehm, I Kronig, RA Neher, et al. *Clin Microbiol Infect*. **27** (8), 1109–1117 (2021).
26. SM Soh, Y Kim, C Kim, et al. *J Microbiol*. **59**, (9), 807–818 (2021).
27. E Janik, M Niemcewicz, M Podogrocki, et al. *Pathogens* **210** (6), 633 (2021).
28. A Kumar, R Parashar, S Kumar, et al. *J Med Virol*. Epub ahead of print (2021). doi: 10.1002/jmv.27467.
29. J Chen, R Wang, NB Gilby, et al. *ArXiv*. 2112.01318v1. Preprint (2021).
30. SJ Gao, H Guo, and G Luo. *J Med Virol*. Online ahead of print (2021). doi: 10.1002/jmv.27491
31. E Callaway. Nature. **600** (7887), 21 (2021).
32. M Levine-Tiefenbrun, I Yelin, R Katz. et al. *Nat. Med*. **27**, 198–199 (2021).
33. JS Tregoning, KE Flight, SL Higham, et al. *Nat Rev Immunol*. **21** (10), 626-636 (2021).
34. A Puranik, PJ Lenehan, E Silvert et al. medRxiv. Preprint (2021). doi:10.1101/2021.08.06.21261707
35. HH Ibkel. *Rev Med Virol* e2313. Online ahead of print (2021). doi: 10.1002/rmv.2313.
36. *https://www.data.gouv.fr/fr/organizations/sante-publique-france/*
37. *https://www.eficiens.com/coronavirus-statistiques/*
38. *https://ourworldindata.org/coronavirus*
39. *https://covidtracker.fr/covidtracker–france*
40. *Visualisation Covid-19 (germain-forestier.info)*
41. *https://www.statnews.com/2020/07/20/trump-said-more-covid19-testing-creates-more-cases-we-did-the-math/*
42. French platform *MODCOV19 (cnrs.fr)*
43. Montpellier University *« Avec le Covid-19, la modélisation des épidémies s'ouvre aux données massives » | Pour la Science*; *Modélisation de l'épidémie de COVID-19 (ouvaton.org)*
44. Pasteur Institute : *Modélisation COVID-19 - Modélisation COVID-19 (pasteur.fr)*
45. M. U. G. Kraemer, V. Hill, C. Ruis, et al, *Science* **373**, 889–895 (2021)
46. N. Hozé, J. Paireau, N. Lapidus, et al, *Lancet Public Health* **6** e408-15 (2021)
47. W.O. Kermack, A. McKendrick, G.T. Walker, *Proc. R. Soc. A* **115**, 700–721 (1927)
48. *https://images.math.cnrs.fr/Modelisation-d-une-epidemie.html*
49. K. M. Gostic, L. McGough, E. B. Baskerville, et al, *PLOS Comput. Biol*. 16 (12), e1008409. (2020)
50. A. Cori*, N. M. Ferguson, C. Fraser, et al, *Am. J. Epidemiol* **178** 1505–1512 (2013)
51. Press Review, Pasteur Institute, April 2020
52. H. Salje, C. Tran Kiem, N. Lefrancq, et al, *Science* **369**, 6500 (2021)
53. M. O'Driscoll, G. R. Dos Santos, Lin Wang, et al, *Nature* **590**, 140–145 (2021)
54. *Covid-19 l'erreur de 1927 | CORONA-CIRCULE*
55. *https://www.lajauneetlarouge.com/covid-19-interrogations-sur-le-modele-epidemiologique-prise-en-compte-de-la-vaccination-et-du-variant-anglais/*
56. P. Bosetti, C. Tran Kiem, A. Andronico, et al, preprint (2021) https://hal-pasteur.archives-ouvertes.fr/pasteur-03272638/document