Tiered social distancing policies and epidemic control

Pierre-Alexandre Bliman¹, Alessio Carrozzo-Magli², Alberto d’Onofrio³ and Piero Manfredi⁴

¹Inria, Sorbonne Université, Université Paris-Diderot SPC, CNRS, Laboratoire Jacques-Louis Lions, équipe Mamba, Paris, France
²Department of Economics, Università di Bologna, Bologna, Italy
³Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy
⁴Department of Economics and Management, University of Pisa, Pisa, Italy

Tiered social distancing policies have been adopted by many governments to mitigate the harmful consequences of COVID-19. Such policies have a number of well-established features, i.e. they are short-term, adaptive (to the changing epidemiological conditions), and based on a multiplicity of indicators of the prevailing epidemic activity. Here, we use ideas from Behavioural Epidemiology to represent tiered policies in an SEIRS model by using a composite information index including multiple indicators of current and past epidemic activity mimicking those used by governments during the COVID-19 pandemic, such as transmission intensity, infection incidence and hospitals’ occupancy. In its turn, the dynamics of the information index is assumed to endogenously inform the governmental social distancing interventions. The resulting model is described by a hereditary system showing a noteworthy property, i.e. a dependency of the endemic levels of epidemiological variables from initial conditions. This is a consequence of the need to normalize the different indicators to pool them into a single index. Simulations suggest a rich spectrum of possible results. These include policy suggestions and identify pitfalls and undesired outcomes, such as a worsening of epidemic control, that can arise following such types of approaches to epidemic responses.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare. c.6350106.
1. Introduction

During the last two decades, the field of mathematical epidemiology has observed the development of a new trans-disciplinary domain that was termed the behavioural epidemiology of infectious diseases (BEID) [1,2]. A key idea of BEID is that the success—and the design itself—of public health interventions aimed at controlling the spread of new, or re-emerging, infectious diseases, critically depends on humans’ spontaneous responses (at both the individual and collective levels) to the proposed policy programmes. The ongoing COVID-19 pandemic, with its pervasive impact, has represented an open-sky laboratory showing an endless list of behavioural phenomena related to the development of the epidemic and of the various actions enacted to control it. This regards both responses to non-pharmaceutical interventions [3] as well as pharmaceutical ones, first of all vaccination [4], as reviewed in [5].

Since the onset of the COVID-19 pandemic, the scientific community has mobilized to contribute to understand, predict and control it. This has led to the publication of an amazing number of high-quality modelling papers. Most such papers relied on the traditional ‘behaviour-free’ approach to epidemiological modelling [6–22]. However, several contributions also accounted for behavioural aspects. Some of the latter were mainly theoretical (i.e. using stylized models [23–27]), but some also attempted to incorporate behavioural responses (e.g. social distancing) into realistically parametrized COVID-19 models (see e.g. [28] and references therein).

It is worth noting that the main focus of BEID in the pre-COVID-19 era was on the role that individual-level risk perceptions and subsequent un-coordinated actions might play in undermining (or, more seldom, strengthening) the success of otherwise well-designed intervention programmes. That is, a prevailing idea was that, given a ‘responsible’ government (or public health system) acting to protect citizens’ health and welfare, the chief threat to citizens’ health was mainly represented by their own possible policy resistant behaviour [1,2,29]. The COVID-19 pandemic highlighted the boundedness of this perspective: behavioural effects have pervaded essentially all layers of our societies, from the bottom level (individual agents) to intermediate bodies and institutions, up to governments. What is remarkable are the cases of governments maintaining an attitude of denial towards COVID-19, thus dramatically delaying or weakening interventions such as e.g. the Brazilian case or the US case during 2020 [30,31]. In other cases, interventions were delayed—making them much less effective—due to engagement of governments in strategic interactions with other actors, such as political oppositions and lobbies [32]. These examples show the need to also deepen the implications of political decision making, and underlying determinants, during a pandemic (e.g. about start date, form and intensity of control measures) in BEID models.

In this article, we aim to provide a simple representation of the tiered social distancing policies adopted by many governments to mitigate the effects of COVID-19 after the epoch of generalized lockdowns adopted during the first pandemic wave. Indeed, awareness of the disruptive societal implications of generalized lockdowns—especially when unnecessary—emerged when the first wave was ongoing, in spring 2020 [33,34]. Tiered policies were systematically adopted in Europe during the second pandemic wave (since autumn 2020) and were maintained in various degrees even after the start of the COVID-19 vaccination campaigns (winter 2021). This was important for coping with the subsequent appearance of the various variants of concern—alpha [35], delta [36,37] and omicron [38,39]—for which the protection offered by the available vaccines against infection proved sharply lower compared to original strains. Since tiered systems represent so far the most agreed-upon approach to pandemic control in the absence of vaccination, it is important to fully understand the implications of this control approach e.g. towards the possibility of highly dangerous new COVID-19 variants capable of eluding vaccine protection against serious disease, but also towards future pandemic events.

Tiered policies such as the one adopted by the Italian government [40–42] share a number of common features, namely (i) multi-criteria, i.e. they are informed by a multiplicity of indicators of epidemic activity, (ii) adaptivity, with a weekly recomputation and automatic upgrading of tiers.
of the different areas involved) based on the adopted indicators and (iii) short-term, i.e. including the values of the adopted indicators primarily from the recent past.

Therefore, in this article we aim at setting and responding two basic questions stemming from the aforementioned features, namely (i) the appropriate mathematical form of a tiered epidemiological control system and (ii) the properties and implications of tiered systems for epidemic control. To do so, we represent this policy system by a suitable extension of the behaviour implicit $M(t)$-index approach first proposed in [43] and widely applied since then (e.g. [44–46]). By this approach, we will summarize the government’s multiple indicators of epidemic trends and subsequent policy actions, such as (i) related evaluations of epidemic risk, (ii) related decisions in terms of strengthening/lifting measures and (iii) related communication and indications to citizens, by a single scalar index $M_C$. In its turn, index $M_C$ will determine the transmission rate at subsequent times. Formally, index $M_C(t)$ is represented as a composite indicator averaging a number of indicators of epidemic severity and of related societal costs. For simplicity, we considered three indicators of epidemic activity: one measuring the intensity of transmission, one related to incidence of new infections and one related to hospitals’ occupancy.

The resulting policy-informing index is then plugged within an SEIRS epidemic model with the purpose of investigating the dynamical implications and efficacy of such a fully endogenous tiered approach. Due to the inclusion of delayed indicators of epidemic activity, the final model is represented by a peculiar hereditary system showing interesting properties such as the dependence of endemic states on initial conditions. Simulation results suggest a rich spectrum of outcomes and highlight the possibility of pitfalls in this policy approach, potentially hindering epidemic control.

The rest of the manuscript reports the model (§2), its simulation results (§3) and concluding remarks (§4). Online electronic supplementary materials add some theoretical results on endemic state(s) of tiered systems and further simulations.

2. An epidemic model with tiered social distancing interventions

In this section, we present our transmission model with tiered social distancing interventions. We first introduce a baseline SEIRS framework (§2a) and then we expand it to include the modelling of the government response (§2b).

(a) A baseline epidemic framework

We keep the model as simple as possible by ruling out many COVID-19 complexities such as asymptomatic transmission or age-dependencies. Briefly, we assume that the transmission dynamics of the infection in the absence of governmental interventions is given by the following SEIRS ‘epidemic’ (i.e. not including vital dynamics) model without vaccination:

\[
\dot{S} = -\beta(t) \frac{I}{N} S + \theta R; \quad (2.1)
\]
\[
\dot{E} = \beta(t) \frac{I}{N} S - \alpha E; \quad (2.2)
\]
\[
\dot{I} = \alpha E - (\gamma + \delta) I \quad (2.3)
\]
\[
\dot{R} = \gamma I - \theta R, \quad (2.4)
\]

where $S, E, I, R$ denote the numbers of individuals who are susceptible, exposed to infection, infective and recovered, respectively. Moreover, $\beta(t)$ is the time-varying transmission rate, $\alpha$ is the rate of development of infectivity, $\gamma$ is the recovery rate, $\theta$ is the waning rate of natural immunity, which is known to be only temporary [47] and $\delta$ is the disease-specific mortality rate. Therefore, the population size $N = S + E + I + R$ is not constant as

\[
\dot{N} = -\delta I.
\]
By recalling that \( R = N - S - E - I \), the above model becomes:

\[
\begin{align*}
\dot{S} &= -\beta(t) \frac{I}{N} S + \theta(N - S - E - I); \\
\dot{E} &= \beta(t) \frac{I}{N} S - \alpha E; \\
\dot{I} &= \alpha E - (\gamma + \delta) I; \\
\dot{N} &= -\delta I. 
\end{align*}
\]

(2.5) - (2.8)

For the previous model well-known properties hold. If natural immunity is lifelong (\( \theta = 0 \)), the classical SEIR epidemic model is recovered. For \( \theta > 0 \), under a constant transmission rate \( \beta \) and in the absence of infection-related mortality (\( \delta = 0 \)), if the basic reproduction number (BRN) \( R_0 = \beta / \gamma \), representing the number of secondary infections caused by a single infective individual in a wholly susceptible population in the absence of interventions, is greater than one, the model has a unique endemic equilibrium which is also globally asymptotically stable (GAS). Endemicity is triggered by waning natural immunity that prevents the infection to die-out (as instead occurs in the SEIR model) by replenishing the susceptible compartment. Still for \( \theta > 0 \) and a constant transmission rate but positive infection-related mortality (i.e. \( \delta > 0 \)), the model has no endemic equilibria because any constant level of prevalence \( I \) would force the population to decline to zero according to (2.8). However, considering the epidemiological fractions \( s = S/N, e = E/N, i = I/N, r = R/N \), with \( s + e + i + r = 1 \), there is a unique (stable) endemic equilibrium \( s^*, e^*, i^*, r^* \), which arises when the corresponding BRN \( \beta / (\gamma + \delta) \) exceeds. This clarifies that population extinction will occur at constant rate \( -\delta i^* \), according to \( N'/N = -\delta i^* \) [48,49].

Let us now define the current effective reproduction number as

\[ \mathcal{R}_E(t) = \frac{\beta(t)}{\gamma} s(t). \] (2.9)

Quantity (2.9) represents the number of secondary infections caused by a single infective individual in a partly immune population in the presence of control measures targeting \( \beta \). In the absence of control measures (i.e. for constant \( \beta \)), this reduces to the well-known formula \( \mathcal{R}_E = R_0 s(t) \).

The previous SEIRS model will be used as a benchmark to ground the results provided by the tiered-control models developed in the subsequent sub-sections. In particular, in our analyses we will assume that governmental interventions start after an initial phase of free epidemic where no governmental control actions take place (though some spontaneous behavioural responses could occur), i.e. the epidemics will evolve under constant \( \beta \). Therefore, we make the following:

**Assumption 2.1.** Government interventions start at time \( t = t_{st} > 0 \). For \( t \in [0, t_{st}) \) it holds that \( \beta(t) = \beta_H \).

The level \( \beta_H \) will represent the transmission upper bound in subsequent analyses (\( H \) stands for ‘High transmission level’).

(b) Modelling transmission under a tiered social distancing policy accounting for epidemic severity only

To represent the characteristics of tiered social distancing policies, namely multidimensionality, adaptiveness and short-term, we will use a suitable extension of the behaviour implicit \( M(t) \)-index approach first proposed in [43]. By this approach we will summarize the government evaluation of epidemic trends by an appropriate composite information index \( M_{Gov}, M_G \) for brevity. In its turn, the information index \( M_G(t) \) will determine the government actions on current transmission rate \( \beta(t) \). In what follows we first consider a basic case where \( M_G \) only accounts for epidemic severity, which we later extend to also include societal costs of the epidemic.
(i) The composite index $M_G$ and its effects on transmission

Let us first describe the time-varying transmission rate $\beta(t)$ as a function of index $M_G$. From assumption (2.1):

$$\beta(t) = \begin{cases} 
\beta_H & \text{if } t \in (0, t_{st}), \\
F(M_G(t)) & \text{if } t \geq t_{st}, 
\end{cases} \quad (2.10)$$

where $F(z)$ is a strictly decreasing function such that

$$F(+\infty) = \beta_L \in [0, \beta_H)$$

where $\beta_H = F(0)$ is the aforementioned transmission baseline and $\beta_L$ represents the minimal transmission level achievable by the adopted interventions. This includes the possibility that interventions bring transmission (in a susceptible population) below threshold, which occurs for $R_L = \beta_L/\gamma < 1$.

Letting $R(t) = \beta(t)/\gamma$ to represent the current basic reproduction number when current transmission is represented by $\beta(t)$, it holds

$$R_L \leq R(t) \leq R_H.$$ 

This allows us to also represent situations where, even in the absence of official governmental measures, some spontaneous behavioural responses (e.g. wearing masks) might occur, lowering background transmission, as observed in the earlier phases of the COVID-19 pandemic.

As for $M_G(t)$, we take it as a scalar yardstick of the intensity of indicators of epidemic severity. We refer to the Italian case where, after the first wave, the government adopted a set of 21 indicators to monitor epidemic activity. For the sake of simplicity and for consistency with our SEIRS framework, we consider here the following three indicators of epidemic activity:

(i) An indicator of the state of transmission, as represented by the current effective reproduction number $R_E(t) = R(t)S(t)$, denoted as $M_1(t)$.

(ii) An indicator of the level of infection incidence (whose current value at time $t$ is $\alpha E(t)$), denoted as $M_2(t)$.

(iii) An indicator of the level of hospitalization incidence, denoted as $M_3(t)$. As the current level of hospitalization occupancy $H(t)$ is not represented in our SEIRS framework, we assume that $H(t)$ is given by a convolution of past infection incidence through an appropriate kernel.

Given the different indicators $M_1(t), M_2(t), M_3(t)$, the composite index $M_G(t)$ will be obtained by suitably aggregating the chosen indicators after a normalization [50,51]. Normalization of indicators is a necessary step prior to their aggregation into the final index, given that different indicators will typically have different scales. By taking the most common normalization criterion, namely Min–Max normalization [50,51], indicator $M_i$ will take the normalized form

$$Z_i = \frac{M_i - \min_{s \in \Sigma} M_i}{\max_{s \in \Sigma} M_i - \min_{s \in \Sigma} M_i}, \quad (2.11)$$

where $\min_{s \in \Sigma} M_i$ and $\max_{s \in \Sigma} M_i$ are chosen over an appropriate reference space $\Sigma$.

Therefore, in general it is

$$M_G(t) = \Phi(Z_1(t), Z_2(t), Z_3(t)), \quad (2.12)$$

where $Z_i(t)$ for $i = 1, 2, 3$ are the normalized forms of the $M_i$'s, and $\partial_{Z_i} \Phi > 0$, $i = 1, 2, 3$. For the sake of simplicity and realism, we assume that $M_G(t)$ is a weighted arithmetic average of the three
normalized indicators considered, that is

\[ M_C(t) = c_1 Z_1(t) + c_2 Z_2(t) + c_3 Z_3(t), \]  

(2.13)

where \( c_i \geq 0, \sum c_i = 1 \), and the adopted normalization is

\[ Z_i(t) = \frac{M_i(t)}{\max_{s \leq t} M_i(s)} \]  

(2.14)

Form (2.14) represents what we term a dynamic normalization, i.e. one where the reference space \( \Sigma \) is represented by the set of positions visited by the dynamic trajectory of the indicator up to time \( t \), and where \( \min_{s \leq t} M_i(s) \) is set to zero, corresponding to the natural baseline where the infection is absent.

(c) Remarks on the proposed approach

In the proposed approach, the government (i) collects and collates data on epidemic severity by suitable indicators \( M_i(t) \), (ii) summarizes the adopted indicators by a composite index \( M_C \) and (iii) uses \( M_C \) to inform the current social distancing policy by tuning the transmission rate \( \beta(t) \) according to (2.10).

As for the choice of indicators, in principle our approach allows us to consider any number of additional indicators of epidemic severity such as e.g. disease-related mortality. However, we feel that the three indicators considered here represent an appropriate compromise between parsimony and exhaustiveness.

As for the use of a weighted arithmetic mean to aggregate different indicators into index \( M_C \), this represents only a reasonable choice because we are not aware of any ‘official’ indications on how the Italian government actually combined available indicators in taking its decisions to change tiers. However, we feel that any associative mean, i.e. any mean representable according to the classical Kolmogorov–Nagumo–De Finetti (KNDF) theorem [52,53]), could be chosen. A mean is said to enjoy the KNDF associative property if it is invariant to aggregation of the original data, i.e. when it can be computed as the ‘mean of the means’: (i) the original data are partitioned into groups (for example: all indicators related to hospitalization are grouped together, all those related to incidence are grouped, etc.), (ii) each group’s information is summarized by the desired mean and (iii) the overall mean is computed simply using the groups’ means rather than the original data. This is a critical advantage, because it allows us to meaningfully summarize large sets of indicators (as the 21-set used by the Italian government) by sharply reducing complexity while, at the same time, preserving the original richness. A critically important sub-class of associative means is that of power means. Given a positively supported random variable \( Y \) with cumulative distribution function \( F_Y(y) \), the corresponding \( P \)-order power mean is defined as

\[ \bar{y}_P = \left[ \int_{-\infty}^{+\infty} y^P dF_Y(y) \right]^{1/P}. \]  

(2.15)

By tuning parameter \( P \), the previous formula encompasses all most common means having a physical meaning e.g. the minimum of data (for \( P \) approaching \(-\infty \)), harmonic (\( P = -1 \)), geometric (\( P \to 0 \)), arithmetic (\( P = 1 \)), up to the maximum of data (\( P \to +\infty \)). This suggests the following conceptual approach to the analysis of complex policies management based on multiple indicators, as pandemic control by tiered systems, in order to avoid complexity to blow-up: (i) identifying the desired indicators, (ii) grouping indicators into homogeneous classes, (iii) identifying an appropriate mean to summarize each group of indicators and (iv) providing an overall summary (the composite index \( M_C \)) by taking the (selected) mean over the means of each group.

As for the normalization, our dynamic normalization strategy relates each indicator to its dynamic maximum reached during the entire previous epidemic history, according to Min–Max normalization criterion which, besides being intuitive, has the advantage of mapping the indicator values onto the \([0, 1] \) set. In particular, our dynamic normalization has the advantage of
being fully *endogenous* to the analysed system, i.e. it employs the ‘natural’ \( \Sigma \) space, represented by the evolution of the dynamic system until time \( t \). This avoids the super-imposition of arbitrary external constraints typical of fixed or exogenous normalizations. By a fixed normalization we mean e.g. the case where hospital occupancy is normalized by the maximal number of hospital beds available in the system or similar figures. The latter strategy would require us to include in the model rules to represent the system behaviour when the relevant public health resources have been saturated (e.g. increasing mortality due to the saturation of intensive care resources). For the incidence indicator, things become even more arbitrary (which is the appropriate maximal incidence?) because there is no obvious bound, unless one makes some coarse hypothesis e.g. taking the predicted incidence of a free epidemic. Clearly, further dynamic normalization options can be considered e.g. by restricting the choice of the maximum over time intervals of the form \([t - T, t]\). Besides the ensuing need to choose length \( T \), this option would turn the model into a full delay-differential equations system, requiring more caution. For all these reasons we felt that, given the nature of COVID-19 as an emerging, unknown, disease, for which no straightforward information were available for setting the quantities \( \min_{s \in \Sigma} M_i \) and \( \max_{s \in \Sigma} M_i \), the chosen dynamic normalization was the the natural approach.

Further, (2.10) postulates instantaneous implementation of governmental responses in terms of new social distancing rules modifying current transmission \( \beta(t) \) based on epidemic status as summarized by \( M_E(t) \). This choice reflected the Italian realm of the COVID-19 pandemic, when the Italian government systematically resorted to nation-wide decrees to suddenly implement changes in the tiers of Italian regions based on their local evidence.

Last, the model does not include spatial dimensions. This might seem odd, in that the aim of tiered policies \([33,34]\) was to tailor the intensity of restrictions over different geographical settings based on their local levels of epidemic activity, thereby avoiding further generalized lockdowns. However, our idea is that most conceptual aspects of tiered policies can be understood without explicitly including space.

### (d) Modelling the dynamics of information indicators

The three aforementioned indicators \( M_i \) are here formally defined as follows:

\[
M_1(t) = \int_0^{+\infty} R_E(t-x)K_1(x)\,dx = \int_0^{+\infty} S(t-x)\frac{\beta(t-x)}{\gamma}K_1(x)\,dx; \tag{2.16}
\]

\[
M_2(t) = \int_0^{+\infty} \alpha E(t-x)K_2(x)\,dx \tag{2.17}
\]

and

\[
M_3(t) = \int_0^{+\infty} H(t-x)K_3(x)\,dx, \tag{2.18}
\]

where the \( K_j(x) \) \( (j = 1, 2, 3) \) are suitable nonnegative delaying kernels

\[
\int_0^{+\infty} K_j(x)\,dx = 1.
\]

By (2.16)–(2.18), we assumed that each indicator \( M_i \) is a weighted mean of the past history of some underlying epidemic dimension, namely reproduction \((R_E(t))\), incidence of new cases \((\alpha E(t))\), and hospital occupancy \((H(t))\). In the case of the COVID-19 epidemic, the inclusion of delaying processes reflects a number of substantive issues.

As for the transmission indicator \( M_1 \), the delaying kernel is necessary in view of the standard approach used to estimate the current reproduction number \( R_t \) \([54,55]\). This approach relies on the knowledge of the past incidence of confirmed cases jointly with an appropriate (external) estimate of the generation time distribution. This implies that data from two to three weeks ago are actually used \([54,55]\). Note that the full inclusion in the model of this approach would imply the need to also relate transmission to past incident cases according to an even more complicated delaying kernel.
As for the incidence indicator $M_2(t)$, one would like to consider the actually observed (or estimated) incidence of new infective cases. For COVID-19 this is typically not the case for a range of reasons. In the worst case, i.e. in the absence of any contact tracing, the system will observe new infection cases with a substantial delay given by the sum of the average latency time, the time of onset of apparent symptoms, and the time needed for case testing and confirmation. In Western countries during the first wave, the average delay between infection and case confirmation—in the absence of contact tracing—was about 10–15 days even in the absence of public health resource saturation [56]. Such delays could be substantially reduced in the presence of contact tracing. The kernel $K_2(x)$ primarily reflects these effects. Still regarding $M_2$, for the sake of precision, a further 1-day reporting delay should be included, so that the correct model would be as follows:

$$M_2(t) = \int_{t-w}^{+\infty} \text{Inc}(t-x)K_2(x)\,dx,$$

where $\text{Inc}(t)$ actually represents the infection incidence in the day before time $t$, i.e.:

$$\text{Inc}(t) = \int_{t-w}^{t} \alpha E(y)\,dy,$$

and $w$ has the size of 1 day. Therefore, we will use the following approximation:

$$\text{Inc}(t) \approx w\alpha E(t).$$

As for the hospitalization indicator $M_3$, given that the hospitalization dimension is not included in our framework, we assume that $H(t)$ is given by a convolution of past incident ($aE$) or prevalent ($I$) cases. Opting for the latter route, we take

$$H(t) \approx \phi \int_{-\infty}^{+\infty} K_H(x)I(t-x)\,dx. \quad (2.19)$$

Following the literature, we assume that kernel $K_H(y)$ obeys a Gamma distribution [57] and $\phi \in (0, 1)$ is an appropriate scaling proportion representing the fraction of past infective cases requiring hospitalization. As our focus is on qualitative mechanisms, we assume that $K_H(x)$ is an Erlang density of order 2, i.e. $K_H(x) = q^2 x \exp(-qx), q > 0$. This allows the following finite dimensional reduction of (2.19) [58]:

$$\dot{H}_1 = q(\phi I - H_1) \quad (2.20)$$

and

$$\dot{H} = q(H_1 - H), \quad (2.21)$$

where $H_1$ is an auxiliary variable. Notably, this implies that indicator $M_3$ depends on past incident cases through the convolution of kernels $K_3$ and $K_H$.

**Remark 2.2.** The previous discussion suggests that indicators $M_2$ and $M_3$ both reflect past (confirmed) cases through appropriate kernels. As also $M_1$ accounts for past incidence through the estimates of $R_E(t)$, this recalls that most indicators adopted in a tiered approach reflect the same epidemic facts through the lenses of different delaying kernels.

Note also that, considering a specific epidemic stage where the current reproduction number is $R < R_H$ (the discrepancy being due to spontaneous behaviour responses), it holds that

$$M_1(t) \leq R. \quad (2.22)$$

This follows from

$$M_1(t) = \int_{0}^{+\infty} R_E(t-x)K(x)\,dx \leq \int_{0}^{+\infty} RK(x)\,dx = R. \quad (2.23)$$

In our formulation, the delaying kernels $K_j(x), j = 1, 2, 3$ reflect precise epidemiological facts and data collection practices. Moreover, they in turn combine with further delaying patterns, as is the case of the hospitalization indicator $M_3$. To set up a general theoretical analysis would require us to consider wide families of kernels starting from the simplest case where the indicators...
\( M_i \) only account for the current levels of the underlying epidemiological variables up to cases where the indicators account for the entire past epidemic trends. In this work, we will focus on the mostly relevant case where all \( K_i, i = 1, 2, 3 \) are exponentially fading kernels
\[
K_i(x) = a_i e^{-a_i x}. \tag{2.24}
\]
This assumption, besides being realistic, allows us to considerably simplify the model. Indeed, in this case the integro-differential dynamics of the indicators \( M_i(t), i = 1, 2, 3 \) are the integral forms of first-order low-pass linear filters [58–61], allowing reduction to the following ordinary differential equations [58].
\[
\dot{M}_1(t) = a_1 \left( S(t) \frac{\beta(t)}{\gamma} \text{Heav}(t) - M_1 \right), \tag{2.25}
\]
\[
\dot{M}_2(t) = a_2 (w\alpha E - M_2) \tag{2.26}
\]
and
\[
\dot{M}_3(t) = a_3 (H - M_3). \tag{2.27}
\]
To properly set up initial conditions for the previous equations, note that we are interested in the ‘natural’ scenario where an epidemic outbreak debuts at an unknown time and we start observing it at a certain subsequent time \( t_{st} \), arbitrarily fixed at 0. Clearly, for times \( t \leq t_{st} = 0 \) we have no information on infection spread, apart the possible availability of estimates of its basic reproduction number \( R_0 \) from e.g. outbreaks occurring in the past or in other geographical sites where the epidemic started earlier. It follows that variables \( M_i(t), i = 1, 2, 3 \) solve the system
\[
\dot{M}_1(t) = a_1 \left( S(t) \frac{\beta(t)}{\gamma} \text{Heav}(t) - M_1 \right), \tag{2.28}
\]
\[
M_1(0) \geq 0 \tag{2.29}
\]
\[
\dot{M}_2(t) = a_2 (w\alpha E(t) \text{Heav}(t) - M_2), \tag{2.30}
\]
\[
M_2(0) = 0 \tag{2.31}
\]
\[
\dot{M}_3(t) = a_3 (H(t) \text{Heav}(t) - M_3) \tag{2.32}
\]
and
\[
M_3(0) = 0, \tag{2.33}
\]
where \( \text{Heav}(t) \) is the Heaviside function. Therefore, it follows that
\[
M_i^{\text{max}}(0) = 0
\]
and that at \( t = t_{st} > t_0 \) the values \( M_i^{\text{max}}(t_{st}) \) depend on both \( t_{st} \) and the initial values of the three epidemiological state variables: \( S(0), E(0), I(0) \).

(e) Including societal losses as a further indicator

The previous case, where interventions are based only on indicators of epidemic severity, i.e. of direct impact on health only, can hold only in special circumstances. For COVID-19 this had been partly true during the first pandemic wave but it will no longer hold under an enduring attack. The need to incorporate the indirect injuries of the pandemic on economic and societal life was considered in early COVID-19 research by economists (see e.g. [62,63]) and later promoted the adoption of tiered systems [33,34].

A simple way to include societal costs into our framework is by considering a fourth indicator \( M_4(t) \), proportional to the socio-economic loss, such that the overall index \( M_G \) has the form
\[
M_G(t) = \Phi(Z_1(t), Z_2(t), Z_3(t), Z_4(t)), \tag{2.34}
\]
where \( Z_4(t) \) is the normalization of \( M_4(t) \) and function \( \Phi \) is nondecreasing with respect to \( Z_1, Z_2, Z_3 \) and non increasing with respect to \( Z_4 \).
A possible functional form, which we will employ in our simulations, is

$$M_G(t) = \frac{c_1 Z_1(t) + c_2 Z_2(t) + c_3 Z_3(t)}{1 + c_4 Z_4(t)}. \quad (2.35)$$

As a simple model for $M_4$, we will assume that societal costs essentially reflect the perpetuation of social distancing measures over time, e.g.

$$M_4(t) = \int_0^{+\infty} \left( R_H - \frac{\beta(t - x)}{\gamma} \right) K_4(x) \, dx. \quad (2.36)$$

The idea underlying (2.36) is that enduring social distancing measures aimed at reducing transmission at certain levels $\beta(t)$ below the normal level ($\beta_H$) will cumulatively hit those social contacts that are essential to sustain economic, social and relational life, and this will—in its turn—bring societal costs. Though seemingly simplistic compared to the formulations adopted e.g. by economists ([62,64] and references therein), the present approach is ‘holistic’ because any altered level of social contacts due to enforced restrictions will impact on virtually every aspect of societal and relational life.

Also for kernel $K_4(x)$, we assume an exponentially fading memory: $K_4(x) = a_4 e^{-a_4 x}$, yielding

$$\dot{M}_4(t) = a_4 \left( \frac{\beta_H - \beta(t)}{\gamma} \right) H_{\text{heavy}}(t) - M_4. \quad (2.37)$$

The definition of $M_4(t)$ and equations (2.36) and (2.37) imply that for $t \in [0, t_{st}]$ it holds $M_4(t) = 0$, thus for times equal or ‘larger but very close’ to $t_{st}$, it is not possible to normalize $M_4(t)$. It would also not be reasonable to take into account $Z_4(t)$ because it is unlikely that the effects of social distancing are appreciable over a very short interval. Consequently, we define a further time point $t^* > t_{st}$ such that for $t > t^*$, $Z_4$ is taken into account in the computation of $M_G$, as follows:

$$M_G(t) = \frac{c_1 Z_1(t) + c_2 Z_2(t) + c_3 Z_3(t)}{1 + c_4 Z_4(t)}, \quad (2.38)$$

where

$$c_4(t) = d_4 H_{\text{heavy}}(t - t_s). \quad (2.39)$$

(f) Final form of the model for tiered epidemic dynamics

The above-described model defines a functional–differential system, also termed a hereditary system [65], i.e. a dynamical system that depends on the current and past values of its state variables. However, the proposed model can be rewritten in the form of a usual dynamical system augmented by a constraint on initial values. Namely, by setting

$$U = (M_1, M_2, M_3, M_4, S, E, I, R),$$

the model proposed in the previous sections can be rewritten in the following general form:

$$\dot{U} = f(U, W); \quad (2.40)$$

$$\dot{W}_j = H_{\text{heavy}}(U_j - W_j)(f_j(U, W))_+ \quad (2.41)$$

and

$$W_j(0) = U_j(0) \quad (2.42)$$

where $(z)_+ = z H_{\text{heavy}}(z)$, $j \in \{1, \ldots, n_1\}$, $n_1 = 4$. It is easy to verify that (2.41) and (2.42) imply $W_j(t) = \max_{s \in [0, t]} M_j(s)$. Note that for all $t \geq 0$ it is $\dot{W}(t) \geq 0$.

Due to constraint (2.42) and to the special type of equations for variables $W$, systems like (2.40)–(2.42) are such that their steady states will depend on initial conditions. In the electronic supplementary materials, we have provided a pedagogic illustration, based on an elementary system, illustrating this dependency of steady states on initial conditions.
By the above formalism, we can summarize the proposed model in the following extended form:

\[
\dot{S} = -\Psi(t, M_1, M_2, W_1, M_3, W_2, M_4, W_3, W_4) \frac{I}{N} S + \theta (N - S - E - I),
\]

\[
\dot{E} = \Psi(t, M_1, M_2, W_1, M_3, W_2, M_4, W_3, W_4) \frac{I}{N} S - \alpha E,
\]

\[
\dot{I} = \alpha E - (\gamma + \delta) I,
\]

\[
\dot{N} = -\delta I,
\]

\[
\dot{H}_1 = q(\phi I - H_1),
\]

\[
\dot{H} = q(H_1 - H),
\]

\[
\dot{M}_1(t) = a_1 \left( S(t) \frac{\beta(t)}{\gamma} \text{Heav}(t) - M_1 \right),
\]

\[
M_1(0) \geq 0,
\]

\[
\dot{W}_1 = \text{Heav}(M_1 - W_1) \left| a_1 \left( S(t) \frac{\beta(t)}{\gamma} \text{Heav}(t) - M_1 \right) \right|_+,
\]

\[
W_1(0) = M_1(0)
\]

\[
\dot{M}_2(t) = a_2 (\omega \alpha E(t) \text{Heav}(t) - M_2),
\]

\[
M_2(0) = 0,
\]

\[
\dot{W}_2 = \text{Heav}(M_2 - W_2) \left| a_2 (\omega \alpha E(t) \text{Heav}(t) - M_2) \right|_+,
\]

\[
W_2(0) = M_2(0)
\]

\[
\dot{M}_3(t) = a_3 (H(t) \text{Heav}(t) - M_3)
\]

\[
M_3(0) = 0,
\]

\[
\dot{W}_3 = \text{Heav}(M_3 - W_3) \left| a_3 (H(t) \text{Heav}(t) - M_3) \right|_+,
\]

\[
W_3(0) = M_3(0)
\]

\[
\dot{M}_4(t) = a_4 \left( \frac{\beta H - \beta(t)}{\gamma} \text{Heav}(t) - M_4 \right)
\]

\[
M_4(0) = 0,
\]

\[
\dot{W}_4 = \text{Heav}(M_4 - W_4) \left| a_4 \left( \frac{\beta H - \beta(t)}{\gamma} \text{Heav}(t) - M_4 \right) \right|_+
\]

\[
W_4(0) = M_4(0)
\]

where, using formula (2.10),

\[
\Psi(t, M_1, M_2, W_1, M_3, W_2, M_4, W_3, W_4) = \beta(t) = \begin{cases} 
\beta_H & \text{if } t \in (0, t_{st}), \\
F(M_C(t)) & \text{if } t \geq t_{st},
\end{cases}
\]

and \(M_C(t)\) is given by (from formula (2.34))

\[
M_C(t) = \Phi \left( \frac{M_1}{W_1}, \frac{M_2}{W_2}, \frac{M_3}{W_3}, \frac{M_4}{W_4} \right).
\]

As the previous model is a fairly complicated one, in this work we mainly resort to simulation. Nonetheless, we provided an approximated analytical characterization of the model endemic states in the electronic supplementary materials.
3. Simulation results

The following simulation analysis aims to broadly assess the effects of the key parameters tuning the proposed model for tiered control based on the composite index $M_G$. A clue for interpreting the subsequent results lies in previous findings of the behavioural epidemiology literature based on the $M(t)$ index approach. In particular, the effects of behaviour-based risk-avoidance or vaccine-hesitant responses on the transmission and control of SIR endemic infections in a behaviour-implicit setting were analysed in \cite{43,46,66–68}. These works highlighted that complicated dynamics can emerge when behavioural responses are intense and delayed. In the pure epidemic case, it was shown in \cite{66} that—for both deterministic and stochastic epidemics—an intense delayed risk-avoidance response (by individuals or the government) can trigger multiple epidemic waves during a single outbreak.

The key parameters include the weights ($c_i$) of indicators $M_i$ and the rates $a_i$ tuning the delay kernels $K_i$, given the response function $F(M_G)$. In particular, we will consider the two main cases previously discussed, namely the case where index $M_G$ only includes indicators of epidemic severity ($c_4 = 0$, §3a) and the case where $M_G$ also includes societal losses ($c_4 > 0$, §3b).

Basic epidemiological parameters were assigned as follows (we adopted 1 day as time unit): $R_H$ was set to 3.6, as estimated from the first COVID-19 wave in Italy \cite{15}; the average duration of the latent period was set to 5 days ($\alpha = 1/5 \text{d}^{-1}$); the average duration of the infective period was set to one week ($\gamma = 1/7 \text{d}^{-1}$); the transmission upper bound $\beta_H$ was computed as $\beta_H = \gamma R_H$; the disease-specific mortality rate was set to $\delta = 0.025 \gamma$ i.e. mimicking COVID-19 mortality in the overall population (rather in specific high-risk groups); the average duration of natural immunity was set to 1 year ($\theta = 1/365.25 \text{d}^{-1}$) \cite{47}; the fraction of infective cases requiring hospitalization was set to $\phi = 0.025$ \cite{15,56}. The rate $q$ tuning the hospitalization kernel $K_H$ was set to $q = 1/5 \text{d}^{-1}$, implying an average delay between infection and hospitalization of 10 days, broadly consistent with \cite{57}.

As for the population size, we take $N(0) = S(0) + E(0) + I(0) + R(0) = 1$ in all simulations. Baseline initial conditions for epidemiological variables were set to $I(0) = 0.0002$, $E(0) = 0.0008$, $R(0) = 0$ and $S(0) = 1 - E(0) - I(0) - R(0)$. As for the initial condition $M_1(0)$ of $M_1$, summarizing initial knowledge on transmission, we set $M_1(0) = 1.8$. This reflects the idea that some initial knowledge on transmission of a pandemic attack exists, e.g. from pandemic preparedness plans. In particular, we borrowed the level 1.8 from the estimates of the BRN of the 1918 Spanish flu \cite{69}, which was taken—prior to COVID-19—as a baseline for pandemic BRNs. However, in some illustrations we also considered $M_1(0) = 0$. For the other indicators we set $M_2(0) = M_3(0) = M_4(0) = 0$.

As for function $F(M_G)$, we chose

$$F(M_G) = \frac{\beta_H}{1 + \omega M_G}.$$  

In particular, we set $\omega = 3$, corresponding to a 75% decrease of the force of infection if $M_G = 1$. This yields a decline of the BRN to 0.9 (therefore below threshold) when $M_G$ is set to its maximum ($M_G = 1$).

As regards the rates $a_i$ tuning the exponentially fading kernels $K_i$ of indicators $M_i$, $i = 1, 2, 3, 4$, we proceeded as follows. For the transmission indicator $M_1$, we set the baseline $a_1 = 1/3 \text{d}^{-1}$, corresponding to an average delay of 3 days. This short delay corresponds to the case where the government is able to use very recent data on transmission. However, we also considered the values $a_1 = 1/7$, $a_1 = 1/14$, $a_1 = 1/21$, corresponding to average delays of one, two and three weeks, respectively. These longer delays aim to crudely mimic the standard epidemiological approach to estimating current reproduction numbers from epidemic incidence \cite{54}, which relies on past incidence data combined with an estimated generation time distribution (for estimates for Italy see \cite{55,56}). This implies that the actual ‘current’ knowledge of the current reproduction number is always available with a time lag reflecting the range of the Gamma generation time distribution used for COVID. Similarly, also for the other indicators of epidemic
severity $M_2$ (incidence) and $M_3$ (hospitalization), we considered average delays of 3, 7, 14, 21 days, respectively. Finally, for the indicator of societal impact ($M_4$), we set a baseline delay of 10 days $a_4 = 1/10 \text{d}^{-1}$.

The behaviour of a free epidemic predicted by the adopted SEIRS model in the absence of interventions (figure 1) shows the achievement of a prevalence peak (20.4% of the total population) at day 52 (figure 1a). Over the longer term, recurrent epidemics occur due to immunity loss (figure 1b).

(a) Effects of tiered measures when $M_G$ only reflects epidemic severity

Here we illustrate the behaviour of our tiered system when the $M_G$ index only accounts for epidemic severity (i.e. $M_4$ is not included in $M_G$, by setting $c_4 = 0$). In particular, we first consider the limit cases where $M_G$ includes a single indicator at time, i.e. it either depends on current effective transmission $M_1$ only ($§3a(i)$), on hospitalizations only ($§3a(ii)$) and on infection incidence ($M_2$) only ($§3a(iii)$). Next, we compare these limit cases with situations where more indicators are included in index $M_G$ ($§3a(iv)$).

(i) Intervention measures based only on information on effective transmission

In this case $M_G = M_1$, corresponding to $c_1 = 1, c_2 = c_3 = c_4 = 0$. Interventions initiate at time $t_{st} = 10$ after 10 days of free epidemic growth. Figure 2 reports the results over a 1 year horizon for the different values of the delay rate $a_1$.

During the initial phase (prior to $t_{st}$), the epidemic is uncontrolled and grows exponentially (figure 2a,b) at a speed per generation $R_E(t)$ (figure 2f) which is essentially constant at $R_H$ because the susceptible fraction is nearly 100% (figure 2d). Also the transmission rate $\beta(t)$ is constant in this phase (figure 2c). The transmission indicator $M_1$ (figure 2e) rapidly increases from its initial value $M_1(0) = 1.8$ following the pattern of $R_E(t)$. However, over the short time span considered (10 days), $M_1(t)$ is able to approach the true value of $R_E(t)$ only for small average delays, while for large delays this does not occur, i.e. the delay in $M_1$ causes a ‘delayed discovery’ of the BRN. This point is of interest for the debate on how estimating the BRN during outbreaks of threatening pathogens where decisions are to be taken under urgency. Let us now stick on the case of short...
Figure 2. Temporal trend of a controlled epidemic where tiered interventions ($\omega = 3$) are uniquely informed by the transmission indicator $M_1$. The graphs display the effects of different levels of the delay rate $a_1$: $a_1 = 1/3$ (black), $a_1 = 1/7$ (red), $a_1 = 1/14$ (blue), $a_1 = 1/21$ (green). Initial condition on $M_1$ is set to $M_1(0) = 1.8$. Intervention starts at $t_{\text{int}} = 10$. Time horizon: 1 year. (a) Prevalence $I(t)$; (b) exposed $E(t)$; (c) $\beta(t)$; (d) $S(t)$; (e) $M_1(t)$; (f) $R_E(t) = S(t)\beta(t)/\gamma$. Other parameters and initial conditions as in the text. (Online version in colour.)

delay ($a_1=1/3$) (i.e. the case of fast detection of ‘true’ transmission), but our remarks hold for all scenarios considered.

The start of intervention at $t = t_{\text{int}}$ occurs when index $M_G = M_1$ is at its maximum ($M_1 = 1$). This forces (via the response function $F(M_G)$) $R_E(t)$ to suddenly drop below threshold in all scenarios considered, halting the growth of prevalence of both $E$ and $I$ (figure 2a,b). However, the parallel decline in $M_1$ (which is slower and smoother compared to that of $R_E(t)$) reduces the level of severity ‘measured’ by the government, which then reduces the intensity of intervention. This persistent lifting of measures forces $\beta(t)$ to increase so that $R_E(t)$ rapidly returns above threshold (approx. at time $t = 15$) and epidemic growth restarts. The growth of $R_E(t)$ (resulting in epidemic acceleration also in relative terms) continues for the entire period during which $M_1$ declines. This occurs until $t \approx 20$ in the short delay case (and longer under longer delays) at which these opposite trends stabilize around a sort of quasi-equilibrium (figure 2c,f) during which transmission
\( \beta(t) \) remains broadly constant (figure 2c). This causes a phase (lasting until time \( t \approx 75 \) in the short delay case) of fast epidemic exponential growth (given slow susceptible depletion, figure 2d) until \( t \approx 85 \), when the decline in the susceptible proportion causes the decline in \( R_E(t) \) (and \( M_1 \)) to become persistent. Eventually, \( R_E(t) \) will decline below threshold and the epidemic achieves its maximum (at \( t \approx 130 \)), entering in its decline phase. In the subsequent epoch \( R_E(t) \) and \( M_1 \) enter into an aligned pattern where intervention measures are essentially absent, epidemic activity declines to a minimum and the driving force becomes susceptible replenishment due to immunity loss, allowing both \( R_E(t) \) and \( M_1 \) to restart increasing.

Overall, the present scenario, though oversimplified, highlights a factor of potential pitfalls in pandemic intervention, namely the presence of a fast policy response (function \( F(M_C) \)) based on a slow-adjusting transmission indicator (\( M_1 \)). This factor works harmfully when \( R_E(t) \) and \( M_1 \) move in opposite directions (e.g. at times 15–50 in the above example). Indeed, after the intervention has been enacted at time \( t_{st} \), \( M_1 \) continues to decline—suggesting an improving epidemic situation—and eventually goes below threshold, while instead \( R_E(t) \) had returned above threshold causing a new epidemic phase.

Over longer time horizons (figure 3), recurrent epidemics due to immunity loss occur. Issues of diverging trends between \( R_E(t) \) and \( M_1 \) are broadly removed, thereby avoiding incoherent policy responses. The tiered response function (2.10) ensures that some degree of infection control is always active and indeed the long-term prevalence is sharply lower compared to the free epidemic case (see figure 1).

As pinpointed in §2f, the model shows a dependence of equilibria on initial conditions. This can be appreciated by comparing figures 3 and 4 where the only difference in the two underlying parametrizations is represented by the initial condition on indicator \( M_1 \), which was set to \( M_1(0) = 0.0 \) in figure 4 instead of \( M_1(0) = 1.8 \) in figure 3. This difference yields remarkable differences in long-term infection prevalence (figures 3a,b and 4a,b). A further noteworthy consequence of the nontrivial heredity behaviour of the present model can be appreciated by the widely different long-term levels of susceptibles, transmission and prevalence emerging in figure 3. These differences result from the different time-delays embedded in the \( M_1 \) indicator. This is at odds with simple delay systems obtained by heuristically plugging time-delays (both reducible or non-reducible) within a given undelayed ODE system (e.g. [58]). Such delay systems typically maintain the same equilibria of the underlying ODE model and delays only affect their stability properties [43,67], unlike what we observed here.

(ii) Intervention measures based only on information on hospitalizations

This case corresponds to \( c_1 = 0, c_2 = 0, c_3 = 1, c_4 = 0 \). Simulation over a 1 year horizon (figure 5) of an intervention initiated at time \( t_{st} = 30 \) bring multiple (dumped) epidemic waves. These multiple waves are especially pronounced when the average delay embedded in the \( M_3 \) indicator is long, e.g. for \( 1/\alpha_3 = 21 \) d, and in this case they exhibit a quasi-period in the range of 150 days which is dramatically different from the quasi-period induced by immunity loss in a basic uncontrolled SEIRS model (figure 1). These waves are triggered by the combination of the strong (via function (2.10)) but delayed intervention response to epidemic alert embedded in indicator \( M_3 \), which further includes the hospitalization delay. Waves based on the \( M \)-index approach have been known in behavioural epidemiology after the works in [43,46,66–68]. The figure also shows the pattern of \( Z_3 \), the normalized indicator of \( M_3 \), which sets to its maximum (\( Z_3 = 1 \)) during the entire initial invasion phase, then declines with the onset (at \( t_{st} = 30 \)) and success of intervention and eventually oscillates. However, for small delays (\( \alpha_1 = 3 \)) the control action becomes smooth and further waves are minimized. This is confirmed by the longer term simulations reported in the electronic supplementary materials.

Earlier initiation of intervention (at \( t_{st} = 10 \)) dramatically reduces the impact of both the first epidemic wave, a fact widely known in the COVID-19 literature, and also of subsequent waves, thereby conferring further ‘smoothness’ to the control action (figure 6).
Figure 3. Temporal trend of a controlled epidemic where tiered interventions ($\omega = 3$) are uniquely informed by the transmission indicator $M_1$. The graphs display the effects of different levels of the delay rate $a_1$; $a_1 = 1/3$ (black), $a_1 = 1/7$ (red), $a_1 = 1/14$ (blue), $a_1 = 1/21$ (green). Initial condition on $M_1$ is set to $M_1(0) = 1.8$. Intervention starts at $t_{st} = 10$. Time horizon: 10 years. (a) Prevalence $I(t)$; (b) exposed $E(t)$; (c) $\beta(t)$; (d) $S(t)$; (e) $M_1(t)$; (f) $R_E(t) = S(t)\beta(t)/\gamma$. Other parameters and initial conditions as in the text. (Online version in colour.)

(iii) Intervention measures based only on information on infection incidence

This case corresponds to $c_1 = 0, c_2 = 1, c_3 = c_4 = 0$. Results of a simulation over a 1 year horizon (figure 7) keeping $t_{st} = 30$ show that the trajectories of $M_2, Z_2$ and $R_E(t)$ tend to broadly align, allowing a high level of control. In particular, $R_E(t)$ suddenly drops below unit as intervention starts (at time $t_{st}$), and rapidly sets to one thereafter, thereby speeding up the emergence of a constant long-term trend which neutralizes also the oscillatory pattern arising from immunity loss. Unlike in the previous section, the range of values of the exponential delays considered in indicator $M_2$ does not have the potential for oscillations. This is not unexpected, given that the hospitalization indicator $M_3$ essentially represents past incidence further delayed through the delay between infection and hospitalization, in which a further (humped) delay occurs. Again, longer term simulations (not reported) confirm that epidemic activity is brought to full control in the short time span of about 150 days apparent from figure 7.
Figure 4. Temporal trend of a controlled epidemic where tiered interventions \((\omega = 3)\) are uniquely informed by the transmission indicator \(M_1\). All parameters and initial conditions as in figure 3 with the exception of the initial condition on \(M_1\), set to \(M_1(0) = 0\). (a) Prevalence \(I(t)\); (b) exposed \(E(t)\); (c) \(M_1(t)\); (d) \(R_{E}(t) = S(t)\beta(t)/\gamma\). Other parameters and initial conditions as in the text. (Online version in colour.)

Figure 5. Temporal trend of a controlled epidemic where tiered interventions (with \(\omega = 3\)) are uniquely informed by the hospitalization indicator \(M_3\). The graphs display the effects of different levels of the delay rate \(a_3\): \(a_3 = 1/21\) (black), \(a_3 = 1/14\) (red), \(a_3 = 1/7\) (blue), \(a_3 = 1/3\) (green). Intervention starts at \(t_{st} = 30\). Time horizon: 1 year. (a) Prevalence \(I(t)\); (b) exposed \(E(t)\); (c) \(M_3(t)\); (d) \(Z_3(t)\); (e) \(R_{E}(t) = S(t)\beta(t)/\gamma\). Other parameters and initial conditions described in the text. (Online version in colour.)
Figure 6. Temporal trend of a controlled epidemic where tiered interventions (with $\omega = 3$) are uniquely informed by the hospitalization indicator $M_3$. The graphs display the effects of different levels of the delay rate $\alpha_3$: $\alpha_1 = 1/21$ (black), $\alpha_2 = 1/14$ (red), $\alpha_3 = 1/7$ (blue), $\alpha_4 = 1/3$ (green). Intervention starts at $t_{\text{int}} = 10$. Time horizon: 1 year. (a) Prevalence $I(t)$; (b) exposed $E(t)$; (c) $M_3(t)$; (d) $Z_3(t)$; (e) $R(t) = S(t)\beta(t)/\gamma$. Other parameters and initial conditions described in the text. (Online version in colour.)

Figure 7. Temporal trend of a controlled epidemic where tiered interventions ($\omega = 3$) are uniquely informed by the incidence indicator $M_2$. The graphs display the effects of different levels of the delay rate $\alpha_2$: $\alpha_1 = 1/3$ (black), $\alpha_2 = 1/7$ (blue), $\alpha_3 = 1/14$ (red), $\alpha_4 = 1/21$ (green). Intervention starts at $t_{\text{int}} = 30$. Time horizon: 1 year. (a) Prevalence $I(t)$; (b) exposed $E(t)$; (c) $M_2(t)$; (d) $Z_2(t)$; (e) $R(t) = S(t)\beta(t)/\gamma$. Other parameters and initial conditions described in the text. (Online version in colour.)
(iv) Intervention measures based on a combination of indicators

Here, we consider the case $c_1 = 0.5, c_2 = 0, c_3 = 0.5, c_4 = 0$, where the composite index $M_G$ depends non-trivially on both indicators $M_1$ (transmission) and $M_3$ (hospitalizations). Results are reported in figure 8, showing the effects of different combinations of values of the delay rates $a_i$ affecting the underlying indicators. We keep $t_{st} = 30$. Results indicate a beneficial effect of averaging more indicators, in that some harmful features of the underlying subcases are removed (e.g. multiple waves are not allowed even for large delays in $M_3$) and, overall, a better control result is achieved.

(b) Inclusion of societal costs

We consider the case where the composite index $M_G$ depends on both transmission ($M_1$, under the short delay $1/a_1 = 3$ d) and societal costs ($M_4$) with $c_1 = 0.5, c_2 = 0, c_3 = 0$, and three values of $c_4 : 0.25, 0.5, 1$. Intervention is started at $t_{st} = 30$ and concerns on socio-economic costs arise at $t_s = t_{st} + 15 = 45$ d. The results are shown in figure 9, which also reports the trend of the adopted indicators $M_1, M_4$, of their normalizations $Z_1, Z_4$, and of the overall composite index $M_G$. Compared to the corresponding case only considering the transmission indicator $c_1 = 1$ (§3a(i)), the pattern of $M_G$ is more articulated but, overall, the growing concern with indirect epidemic costs speeds up the relaxation of control interventions so that the epidemic dramatically relapses. The extent of the relapse is larger the larger the weight attributed to societal costs, causing the epidemic to restart earlier than in the previous case where $c_1 = 1$ (§3a(i)). As shown in the electronic supplementary materials, quite unsurprisingly, over the longer term the infection establishes an essentially constant trend, coarsely overlapping those of a free epidemic.
despite the presence of a background of persistent epidemic control activities (which have societal costs per se).

Clearly, these types of scenarios are affected by many factors. For example, earlier or stronger interventions, combined with later inclusion of societal cost, will improve epidemic control in
Figure 10. Temporal trend of a controlled epidemic where interventions (with $\omega = 3$) are informed by both the transmission indicator ($M_1$) and societal costs ($M_4$) with $c_1 = 0.5$ and three values of $c_4$: $c_4 = 0.5$ (black), $c_4 = 1$ (red), $c_4 = 0.25$ (blue). Time horizon: 1 year. Intervention starts at $t_{init} = 30$ with $\omega = 6$, while social costs are accounted from $t > t_* = 60$. (a) $I(t)$; (b) $E(t)$; (c) $M_1(t)$; (d) $Z_1(t)$; (e) $M_4(t)$; (f) $Z_4(t)$; (g) $M_6(t)$; (h) $R_{eff}(t)$. Other parameters and initial conditions described in the text. (Online version in colour.)

After a first phase, but at the cost of strengthening epidemic relapse when societal costs arise, see figure 10, where a stronger control response ($\omega = 6$) combines with societal costs arising later ($t = t_* = 60$).
4. Concluding remarks

In this paper, we have generalized the $M$-index approach developed in [43,67] to provide a general representation and analyses of the tiered strategy to epidemic mitigation [33,34]. This strategy was enacted by western governments after appreciation of the dramatic societal costs of generalized lockdowns adopted to mitigate the first COVID-19 wave [33,34]. In Italy, tiered measures were maintained in various grades even after vaccines arrival (early 2021) to cope with the various variants of concern and various assessments of their impact were made [40–42].

As tiered strategies currently represent the most agreed approach to pandemic mitigation in a national setting—both without and with vaccination—it is important to fully grasp its characteristics and potential implications. This holds not only in view of future pandemic events, but also in view of the possible appearance of new COVID-19 variants potentially capable of eluding vaccine protection against serious disease.

Tiered strategies share a number of features, namely the adaptiveness of the response—with periodic and automatic upgrading of tiers—relying on multiple epidemiological and socio-economic indicators. Therefore, we represented tiered policies by a composite index $M_G$ combining indicators of epidemic severity (e.g. transmission, incidence, hospitalization) without or with indicators of the societal costs due to interventions. Index $M_G$ was used to inform the social distancing governmental response. The resulting framework was used to explore the main features of tiered strategies for the case of a fast response in transmission to the evolving epidemiological and societal conditions, as has been the case for Italy [41].

A main finding is that control pitfalls may arise when the fast policy response combines with slowly changing indicators of transmission. This is due to the lack of coordination between the indicators of epidemic activity used to tune the policy response and the timing of the response which in turn provides the new input to the indicators. Additionally, multiple epidemic waves can arise when the index $M_G$ incorporates significantly delayed information of the epidemic course. This is especially the case when the index attributes most weight to hospitalizations, which follow the underlying infection events with a well-defined temporal delay. Further, consideration of societal costs always speeds up the relaxation of control measures. Last, the impact of social distancing restrictions critically depends on the weights attributed to the different indicators to inform policy decisions. Summing up, our main conclusions are that the intervention action will be the more successful the sooner it starts and especially the less delayed is the information adopted to inform it. This suggests that—given basic knowledge of transmission—incidence indicators should always be prioritized. The intuition is that all the indicators of epidemic severity that can be built (e.g. transmission, serious disease, quarantines, hospitalization, mortality, etc.), using data provided by the epidemic in real time, are always delayed manifestations of incidence. This is, for example, suggestive of the possible pitfalls arising from using transmission indicators based on old data because resorting to a generation time distribution with a long range. An example of this pitfall is the delayed intervention against the second pandemic wave in Italy: when the epidemic was already out of control because of the collapse of the tracing system, the transmission indicator $R_t$ was still suggesting a mildly worsening situation [70].

Previous results suggest the merits of the proposed first attempt (to the best of our knowledge) to provide a stylized endogenous representation of the tiered governmental responses during the COVID-19 crisis. However, such results ought to be considered as first evidence of the potential of the proposed approach, so that more extensive theoretical and simulative work will be necessary to fully clarify the model behaviour, also in view of its complexity. Such further work should explore in depth the robustness of model results to all relevant factors included by our approach, e.g. (i) the type of mean adopted to summarize the chosen indicators into the composite index $M_G(t)$, (ii) the type of indicators included in the composite index and their weights, in particular the interplay between the subset of indicators of epidemic activity and the one related to societal welfare, (iii) the type of delaying kernels and (iv) the normalization adopted to combine the indicators into the composite index. Full conceptual understanding of the proposed approach appears an ineludible pre-requisite for subsequent empirical studies.
The proposed model relies on a number of specific hypotheses whose relevance should be checked in future work. First, the underlying epidemiological framework is highly stylized and robustness to a more detailed representation of the COVID-19 epidemic should be considered. Second, we assumed that the public authorities are capable of accessing error-free measurements of the relevant indicators of epidemic activity, which is plainly not true. However, the fact that policy pitfalls and ambiguities arise also in this highly desirable situation suggests that the model highlights interesting phenomena deserving further scrutiny. Another abstraction is that we assumed that the population is fully compliant and rapidly adapting to the social distancing rules communicated by the government, i.e. we ruled out spontaneous behavioural responses. These spontaneous responses typically occur in reality and have been foreseen and documented for COVID-19 [3,4]. However, their inclusion should not be an alternative to what we have done here: rather it should be added on top of the baseline setup we proposed. It is indeed interesting to note that the adopted \( M \)-index representation of \textit{governmental behaviour} has led to a model formulation analogous to typical \textit{behaviour implicit} formulations of individual responses [43,46,67]. Further, the model has disregarded vaccination. Future pandemic preparedness plans should carefully investigate the coexistence of tiered social distancing measures and vaccination. Also, the response of transmission to governmental interventions is far more complicated than the fast response proposed here. However, adding delayed responses to the information delays would surely yield more important oscillatory behaviours. In any case, this certainly calls for more sophisticated real-world policies, incorporating perhaps model-based predictions.

Finally, from a more speculative standpoint, the model proposed here is, as Vito Volterra would say, a \textit{hereditary} system [65] whose equilibria depend on initial conditions. In other words, the system never loses the memory of its past. This seems to be an unavoidable property of tiered policies based on multiple indicators. In the real world the chosen indicators might have an even more complicated form. For example, the normalizing values of a given indicator \( M(t) \) might take the form \( \max_{t \in [t-T, t]} M(t) \), where the time length \( T \) is established by some \textit{ad hoc} rules by decision-makers. This opens the way for a large number of new problems deserving further analysis in future work.

\textbf{Data accessibility.} The code used to simulate the model and to generate the figures is provided in electronic supplementary material [71].

\textbf{Authors’ contributions.} P.-A.B.: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review and editing; A.C.-M.: conceptualization, methodology, writing—original draft, writing—review and editing; A.d.O.: conceptualization, formal analysis, investigation, methodology, software, writing—original draft, writing—review and editing; P.M.: conceptualization, formal analysis, investigation, methodology, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

\textbf{Conflict of interest declaration.} We declare we have no competing interests.

\textbf{Funding.} No funding has been received for this article.

\textbf{Acknowledgements.} The authors warmly thank three anonymous referees of the journal whose valuable comments allowed us to sharply improve the quality and exposition of the manuscript.

\section*{References}


