Strategies for controlling the medical and socio-economic costs of the Corona pandemic

Claudius Gros\textsuperscript{[1]}, Roser Valenti\textsuperscript{[1]}, Kilian Valenti\textsuperscript{[2]}, Daniel Gros\textsuperscript{[3,4]}

\textsuperscript{[1]}Institute of Theoretical Physics, Goethe University, 60438 Frankfurt a.M., Germany
\textsuperscript{[2]}Vivantes Klinikum Spandau, 13585 Berlin, Germany
\textsuperscript{[3]}Department of Economics, University of California, Berkeley, USA
\textsuperscript{[4]}CEPS (Centre for European Policy Studies), 1000 Brussels, Belgium

April 1, 2020

Abstract

In response to the rapid spread of the Coronavirus (COVID-19), with ten thousands of deaths and intensive-care hospitalizations, a large number of regions and countries have been put under lockdown by their respective governments. Policy makers are confronted in this situation with the problem of balancing public health considerations, with the economic costs of a persistent lockdown. We introduce a modified epidemic model, the controlled-SIR model, in which the disease reproduction rates evolve dynamically in response to political and societal reactions. Social distancing measures are triggered by the number of infections, providing a dynamic feedback-loop which slows the spread of the virus. We estimate the total cost of several distinct containment policies incurring over the entire path of the endemic. Costs comprise direct medical cost for intensive care, the economic cost of social distancing, as well as the economic value of lives saved. Under plausible parameters, the total costs are highest at a medium level of reactivity when value of life costs are omitted. Very strict measures fare best, with a hands-off policy coming second. Our key findings are independent of the specific parameter estimates, which are to be adjusted with the COVID-19 research status. In addition to numerical simulations, an explicit analytical solution for the controlled continuous-time SIR model is presented. For an uncontrolled outbreak and a reproduction factor of three, an additional 28\% of the population is infected beyond the herd immunity point, reached at an infection level of 66\%, which adds up to a total of 94\%. 

1
Introduction

The rapid spread of the Coronavirus (COVID-19) has led the World Health Organization (WHO) to declare the outbreak a pandemic in March 2020. The epidemic started in the Hubei province of China, around the turn of the 2019/20, spreading thereafter to the entire country and to almost the entire world [1]. In late February 2020, early March, contagion spread rapidly first throughout Europe and then through the US. In response to the growth of infections [2] and in particular to the exponential increase in deaths, most of Europe has been put under lockdown, with the US adopting similar measures. The toll on economic activity is considerable. Restaurants are closed or empty, only 10% of the normal number of flights operate, investment plans are being delayed, etc. Discounting the negative economic impact of epidemic containment policies [3], financial markets have become extremely volatile.

Policy makers are confronted with difficult choices. If no action is undertaken, the number of infected increases exponentially together with the death toll. But prolonged total lockdowns may lead to an economic collapse. The resulting need to trade off economic costs against avoided death, or, more generally, public health, is difficult, but not new per se. With regard to the COVID-19 outbreak, it is, hence, necessary to estimate the medical costs and put a price on the value of lives saved by ‘social distancing’. We provide a first, bottom up, estimate of the hospitalization costs which could be avoided, and, we use the implicit ‘value of a life’ as applied in medical frameworks and other areas of public policy to calculate the economic equivalent of lives saved by different policy options.

The societal and political response to a major outbreak like COVID-19 is highly dynamic, changing often rapidly with increasing case numbers. To capture this aspect, we use a standard epidemic model modified in one key point: the reproduction rate of the virus is not constant, but evolves over time alongside with the disease. Our starting model is the SIR (Susceptible, Infected, Recovered) model [4], which describes the evolution of a contagious disease for which immunity is essentially permanent. The negative feedback-loop inherent in the model investigated, the controlled-SIR model, leads to a ‘flattening of the curve’ [5, 6] proportional to the control strength. This is illustrated in Fig. 1 where we show results representing an epidemic outbreak with and without the control pattern proposed here.

Calibrating the numerical simulations with regard to up-to-date COVID-19 data, we find that the overall costs of the endemic could be highest at a medium level of reactivity. Strict social distancing measures reduce the combined medical, economic and value of life costs most effectively. For some parameter combinations a hands-off policy fares also better than a medium level of reactivity, but mainly when the value of life lost is disregarded. Our parameters will need to be readjusted as new research emerges, however, we are confident that the qualitative conclusions presented here stand.

We start by reviewing recent studies on the economic aspects of controlling contagious diseases. For the weighing of the economic costs of a disease transmission against the cost of treatment, and the loss of life, a framework has been established [7, 8]. A critical input is here the value of Statistical Life (VSL), which corresponds to the monetary value of an avoided premature death [9, 10, 11, 12]. This framework has been applied to the Corona pandemic in several recent contributions in which
the evolution of the endemic is taken as exogenous [13, 14], relying on certain premises, as those in [15], for the infection rates\(^1\). Further studies discuss the relative effectiveness of different control measures (such as school closures versus a lockdown) [15, 19], and the possible course of the disease for European countries [20]. The authors of Ref. [21] suggest to derive disease transmission rates from economic principles of behavior, which would allow to measure the cost of the Corona pandemic under different policy settings. Another strand of the literature takes the pandemic as given, and as the basis for scenarios for the economic impact and for the financial-market volatility [22, 23, 24, 25].

Regarding the underlying epidemiology models, early on, the study of the dynamics of measles epidemics [26] has shown that human behavior needs to be taken into account [27, 28]. For this purpose, a range of extensions to the underlying SIR model have been proposed, such as including the effect of vaccination, contact-frequency reduction and quarantine [29], human mobility [30], self-isolation [31], and the effects of social and geographic networks [32]. It has been recognized, in particular, that the spontaneous behavioral response of individuals to the progress of an epidemic may lead to a suppression threshold [33]. Statistical approaches provide additional valuable tools

\(^1\)Case fatality rates can be estimated from the outbreak in Wuhan (China) and on the cruise-ship Diamond Princess [16, 17], or, by analogy, from the previous SARS epidemic [18].
for the study of infection processes [34], and of vaccination campaigns [35]. Alternatively, one may use game-theoretical methods to model the effectiveness of social distancing [36].

A range of determining factors have been examined for the ongoing COVID-19 epidemic, in particular the effect of quarantine [37] and that community-level social distancing may be more important than the social distancing of individuals [38]. An agent-based model for Australia found, in this regard, that school closures may not be decisive [39]. Microsimulation models suggest, on the other hand, that a substantial range of non-pharmaceutical interventions are needed for an effective containment of the COVID-19 outbreak [15].

Detailed epidemiology models have the potential to provide the basis for informed political decision making. The drawback is, however, the generally large number of adjustable parameters and the lack of an explicit analytic handling. As an alternative we suggest in the present study a minimal model, the controlled-SIR model. This model is fully determined by two parameters: the underlying disease reproduction factor $g_0$ and a parameter, $\alpha$, which unites together the effect of individual, social and political reactions to disease spreading. The controlled-SIR model can be solved analytically in the continuous-time limit and allows for a direct understanding of the consequences of increased control of an epidemic.

**Results**

In what follows we first introduce the controlled-SIR model that explicitly takes into account societal reactions to the spread of the disease. We then present numerical simulations for parameter values assumed for COVID-19 and discuss analytical results obtained for the continuous-time version of the model. We also show how this model can account for incomplete available data and insufficient disease-testing. We then proceed by presenting a detailed socio-economic analysis of the cost of the epidemic (work-hours lost, medical costs and lives lost) and the cost of ‘social distancing’ measures adopted to control the endemic. These considerations allow for controlled quantitative predictions in our simulations.

**Modeling epidemic control**

The SIR (Susceptible, Infected, Recovered) model [4, 40] describes the evolution of the density of infected individuals $I_t \in [0, 1]$ at time $t$. We start with the discrete time model with $t = 0, 1, \ldots$. One period, from $t$ to $t + 1$, is defined by the time span an infected individual is contagious, which is of the order of two weeks for COVID-19. Recovered (or deceased) individuals are considered to be immune. This implies that the epidemic spreads when infected and susceptible individuals meet, which occurs with the probability $I_t (1 - X_t)$, where $X_t$ is the total number of individuals that are either currently ill (active cases), or that have been infected in the past time (recovered or deceased). The SIR model is hence equivalent to

$$I_{t+1} = g_t I_t (1 - X_t), \quad X_t = \sum_{k=0}^{\infty} I_{t-k}.$$  \hspace{1cm} (1)
with $g_t$ being the reproduction factor at time $t$.

To consider a constant reproduction factor $g_t$ in an epidemic model is an assumption that is arguably a valid basis for the modeling of ‘normal’ diseases, i.e. infectious diseases that do not threaten public health to a significant extent (e.g. influenza). However, the COVID-19 virus has clearly led to massive political and social responses. We propose that the reproduction factor evolves in this situation in line with the progressing of the disease. For this purpose we investigate the following two functional dependencies:

$$g_t = \begin{cases} g_0 (1 + \alpha I_t)^{-1} & \text{(short-sighted)} \\ g_0 (1 + \alpha X_t)^{-1} & \text{(history-aware)} \end{cases}$$

The parameter $\alpha$ describes the strength of the reaction of the society to the spreading of the disease and $g_0$ corresponds to the ‘intrinsic’ or medical infection growth factor in the absence of behavioral reactions, i.e. $g_0$ is the disease transmission by one individual. The first case in Eq. 2, termed here ‘short-sighted’, describes the situation where society reacts only to the percentage $I_t$ of the population currently infected. The inverse functionality in Eq. 2 captures the fact that it becomes progressively more difficult to reduce $g_t$ with increasing social distancing. Reducing $g_t$ only somewhat is comparatively easy, a suppression by several orders of magnitude requires, in contrast, a near to total lockdown.

For the second strategy examined, denoted ‘history-aware’, the reaction is based on the overall history of the epidemic. Society takes the total number of past and current infected cases, $X_t$, into account. At first sight it might appear irrational for society to react to the history of cases, but, as shown in this study, such a ‘history-aware’ reaction is likely to lead to lower overall costs.

Fig. 2 illustrates the capability of short-sighted and history-aware reaction policies to contain an epidemic. The fraction of infected individuals is obtained from solving Eq. 1 and Eq. 2 numerically. While both strategies are able to lower the peak of the outbreak with respect to the uncontrolled ($\alpha = 0$) case, the disease will, however, become close to endemic when the reaction considers only the actual number of cases, and not the overall history of the outbreak.

**Societal reaction to the spread of the disease**

The two strategies investigated here, short-sighted and history-aware, correspond to reaction patterns that are observed for the COVID-19 outbreak [41, 42]. For both cases, the societal reaction described by the parameter $\alpha$ can be thought as a sum of two contributions $\alpha = \alpha_s + \alpha_g$, where $\alpha_s$ quantifies the spontaneous reaction by the population and $\alpha_g$ encodes government interventions.

The first contribution, $\alpha_s$, takes into account societal behavioral changes happening when a substantial fraction of the population spontaneously adopts social distancing (avoiding hand-shakes, restaurants, cinemas, etc.), f.i. in response to media reports about the severity of the outbreak. Voluntary social distancing led to substantially reduced restaurants and cinemas attendances even before governments decided to step in with mandatory school closures, curfews and other drastic measures. An important aspect of the spread of COVID-19 is the distinct reactions of societies in
Control of epidemic peak. Shown is the timeline of actual infected cases during an epidemic outbreak with an intrinsic reproduction factor $g_0 = 3.0$, which is close to COVID-19 estimates [18]. The simulation is obtained by iterating Eq. 1, with one iteration corresponding to two weeks, taken as the average duration of the illness. Short-sighted control, which responds to the actual number of cases, see Eq. 2, is able to reduce the peak strain on the hospital system, but only by prolonging substantially the overall duration. History-aware control, which takes the entire history of the outbreak into account, is able to reduce both the peak and the duration of the epidemic.

In Asia the wearing of masks becomes a convention, which is likely to correspond to a higher $\alpha_s$, while such a measure tends to be resisted by a majority of European populations. The second contribution to the control factor, $\alpha_g$, captures the role of government interventions. Measures ranging from forbidding large events, to school closures and, finally, to lockdowns, become politically possible when the number of individuals infected increases and surpasses critical levels. The aim of our investigation is not to evaluate the effectiveness of specific measures, which has been done elsewhere [19, 15], but to assess the dynamical effect of the societal reaction encoded in the feedback parameter $\alpha$, on the overall evolution of the epidemic. As most social distancing measures are costly, both for the economy and the society, it is reasonable to assume that their strictness is increased only when necessary, viz in relation to the severeness of the outbreak. The latter is measured in Eq. 2 either by the number of actual cases, $I_t$, or by the cumulative case count $X_t$. 

different countries.
Analytical results for history-aware control

Epidemic outbreaks may be modeled either using discrete time steps, typically of the order of days or weeks, or by a continuous-time formulation. Both have their merits, with the minimal discrete controlled-SIR model, Eq. 1, being very intuitive. Rigorous analytic results can be derived, on the other hand, for the continuous-time controlled-SIR model, as detailed out in the Methods section.

The point at which the number of infected individuals does not increase anymore is usually called ‘herd immunity’. It is reached for $\alpha = 0$ when $g_t(1 - X_t) = 1$ in Eq. 1. Beyond this point, the fraction of actual cases $I_t$ declines, which implies, by definition, that $I_t$ is maximal at the herd immunity point, $I_t \rightarrow I_{\text{max}}$, and that the herd immunity point corresponds to the time of maximal stress for the hospital system. When controlled, i.e. $\alpha > 0$ the curve is flattened and $I_{\text{max}}$ lower.

For the continuous-time model one finds the relations,

$$I_{\text{max}}|_{\alpha \gg 1} \approx \left(\frac{g_0 - 1}{g_0\alpha}\right)^2, \quad I_{\text{max}}|_{\alpha = 0} = 1 - \frac{1 + \log(g_0)}{g_0},$$

(3)

The second relation is a well-known result, which translates into a peak load of $I_{\text{max}}|_{\alpha = 0} \approx 0.3$ for an unchecked epidemic with $g_0 = 3$, as for COVID-19 [17]. The first relation in Eq. 3, which is particular to the controlled-SIR model, holds in the regime of strong control, viz for large $\alpha$. It states that the maximal medical load is suppressed proportionally to $1/\alpha$.

The epidemic dies out quickly beyond the maximum $I_{\text{max}}$, as illustrated in Fig. 2. This implies that a finite fraction $1 - X_{\text{tot}}$ of the population is never infected. When uncontrolled, about 94% are eventually infected for $g_0 = 3$, as discussed in the Methods section, with only 6% remaining unaffected by the disease. For the continuous-time controlled-SIR model one finds the scaling relations

$$X_{\text{tot}}|_{\alpha \gg 1} \approx 2 \frac{g_0 - 1}{\alpha}, \quad X_{\text{tot}} \approx \frac{2g_0}{g_0 - 1} I_{\text{max}},$$

(4)

which hold for $\alpha \gg 1$, when assuming an infinitesimal starting infection fraction $X_0 = I_0$. The second relation can be used to predict the final reach of the outbreak, in terms of $X_{\text{tot}}$, once $I_{\text{max}}$ is known, viz once the point of maximal medical load has been reached.

Available data: incomplete disease monitoring

Disease monitoring is confronted with the problem that the number of unreported cases is difficult to estimate [43]. Predominantly only symptomatic cases with serious conditions are tested when testing capacities reach their limits. For example, as of mid-March 2020, the degree of testing for COVID-19, as measured by the proportion of the entire population, varied by a factor of 20 between the United States (340 tests per million) and South Korea (6100 tests per million) [44]. According to some estimates [45] the true incidence might be higher by up-to a factor of ten than the numbers reported in the official statistics as ‘positive’.

This factor can be incorporated in Eq. 2 by positing that the societal reaction depends on the number of known cases, which is equal to the number of real cases multiplied by a fraction, denoted
by $\theta$. The social distancing factor which lowers the transmission rate would then be given by

$$g_t = g_0 \left(1 + \alpha(\theta I_t)\right)^{-1}, \quad g_t = g_0 \left(1 + \alpha(\theta X_t)\right)^{-1},$$

(5)

respectively, for short-sighted and history-aware control. It is apparent that the 'undercounting factor' $\theta$ is equivalent to a rescaling of $\alpha$ by $\theta$ and, thus, equivalent to having a society which reacts little to the spread of the disease. Fig. 3 illustrates how COVID-19 testing parametrized by $\theta$ influences the number of actually infected individuals. The results are obtained from numerical simulations of Eq. 1 and Eq. 5. The parameter values in this simulation were calibrated on actual data for the to-date only case where an entire population of a town was tested, namely the city of Vó in Italy. This experiment found that 3% of the population was infected [46]. We therefore chose for this simulation a value of $\alpha$ that leads to a peak for the proportion of active cases of around 3%.

**Policy: Socio-economic trade-offs**

We are interested in how different policies and societal reaction patterns, as embedded in the parameter $\alpha$, influence the overall costs of the epidemic. This requires an explicitly inter-temporal approach since the cost of restrictions today (lockdowns, closure of schools, etc.) to public life must be set against future gains in terms of lower infections (less intensive hospital care, fewer deaths).
The key question is then, which policy would lead to the best combination of a low number of cases and a low economic cost. We presume that four elements dominate the cost structure: (i) The working time lost due to an infection, (ii) the direct medical costs of infections, (iii) the value of life costs, and (iv) the cost related to ‘social distancing’. The first three are medical or health-related. All costs can be scaled in terms of GDP per capita (GDP\textsubscript{p.c.}). This makes our analysis applicable not only to the US, but to most countries with similar GDP\textsubscript{p.c.}, e.g. most OECD countries.

**Health costs, loss of working time**

A first direct impact of a wave of infections is that a fraction of the population cannot work. Based on the Diamond Princess data [47], where the entire population was tested, we estimate that only half of the infected develop symptoms that require them to stay home for a one- to two-week period and an additional two-week period until they are no longer contagious. About 20% of the total population (or 40% of those with symptoms) develop stronger symptoms requiring one additional period of absence from work [17]. To be conservative, we assume that there are no severe cases or deaths among the working age. This results in a reduction in the work force per year (52 weeks) of around \(0.3 \times 2 + 0.2 \times 3\) = 2.4/52 = 5 percent, for every 1% of the population infected.

**Medical costs, treatment, hospitalization**

There are no rigorous studies yet of the costs of treatment for the COVID-19, but it is estimated that about 20% [48] of the infected individuals require some sort of hospitalization\(^2\), with around 5% needing intensive care and roughly 1% dying [47].

Intensive care with ventilation is the most costly form of life saving in hospital care. In the US, the cost of 2 weeks of an intensive care unit is equivalent to about 1 year (100%) of GDP\textsubscript{p.c.} [50]. In Germany, which might be typical of the rest of Europe, the cost of 2 weeks of intensive care appears to be somewhat lower, around 20,000 euro, or roughly 60% of GDP\textsubscript{p.c.} [51]. We use the German parameter for a conservative estimate of medical costs. The cost of general hospitalization for 2 weeks is assumed to be 12,000 euro, and equivalent to about 30% of GDP\textsubscript{p.c.}. It is estimated that the median time from onset to recovery for mild cases is approximately two weeks and 3-6 weeks for patients with severe or critical disease [52]. We use a conservative estimate of 2 weeks of intensive care and 2 weeks of general hospitalization for severe cases. This results in a medical cost of \((0.05 \times 0.6 + 0.05 \times 0.3 + 0.15 \times 0.3 = 0.09)\), that is 9% GDP\textsubscript{p.c.}.

**Value of lives lost**

Third, the cost of premature death through the disease represents the most difficult contribution to evaluate in financial terms. We will show below that our central results remain valid even without

\(^2\)An average influenza season leads to an hospitalization of about 0.12% of the US population [49]; and one fourth of them require intensive care, with one twentieth (0.13% of all infected) dying [47]. Averaged over the 2010-17, of the order of 35 thousand influenza-related deaths per year have been registered in the US. For Germany, with a quarter of the US population, these numbers would translate into 8-9 thousands influenza deaths per year.
assigning a value to lives lost, but since major contributions [14, 13] are based on an evaluation of
the economic value of lives lost, we show how this point can be incorporated into our framework.
There are two ways to attribute a monetary value on a life saved or lost. The first one, mentioned
above, is based on the concept of a Value of Statistical Life (VSL), which is commonly used in the
impact assessment of public policy which aims at lowering the probability of an avoided premature
death [53]. A typical application scenario for VSL is the case when the probability of death is
very low (e.g. car accidents), but could be lowered even more (seat belts). For COVID-19, a high-
death epidemic, we prefer a medical-based approach3. Putting a monetary value on lives saved is
unavoidable in medical practice that is confronted with the problem of selecting the procedures to
be used to prolong life - a situation that arises for many patients infected by the Coronavirus under
intensive care. A lower bound of 50,000 dollars per year of life lost is typically used in the existing
literature dealing with the cost of medical procedures, with a central range between 100,000 and
300,000 dollars [55, 56]. Given the current US GDP_{p.c.}, these values translate into a range of 1.5 to
4 years of GDP_{p.c.}. Cutler and Richardson [57] argue for a value equivalent to three times GDP_{p.c.}
We use the lower bound of this range for most of our simulations for a conservative estimate of the
value of lives saved.

What remains to be determined is the number of years lost when a Corona patient dies. First
studies of China epidemic showed that the median age of those who died was only 56 years [58],
whereas an average death-age slightly below 80 years has been reported for Italy [59]. This implies
over 20 years of life lost per death, given a residual life expectancy of 19.9 years at 60 in China,
and about 8 years for Italy. On the cruise ship Diamond Princess [47, 16] which served almost
as a laboratory, most deaths were for individuals with ages between 70-79. Cruise passengers tend
to have fewer acute health conditions than the general population, thus rendering the co-morbidity
argument less prominent. With an average age of around 76 years, the remaining life expectancy
would be 10.5 years for men and over 12 years for women [60]. Weighted by the different propensity
for men and women to be infected and die, the average number of years lost is 11. This implies
that the economic value of the premature deaths should be equal to about 11 times the loss for one
year of life saved (potentially higher for most European countries which tend to have a higher life
expectancy). For each 1% of the population the value of lives lost would thus be equal to 0.01 × 11 ×
the nominal value of one year of life.

The value of life can be measured in terms of multiples of GDP_{p.c.}, which allows to write the
sum of the three types of health or medical costs (loss of working time, hospitalization and value of
lives lost) as a linear function of the percentage of the population infected:

\[ c^\text{med}_t = k I_t \]

with a proportionality factor \( k \) being equal to the sum of the three contributions. Scaling \( k \) with
the GDP_{p.c.} allows for an application and comparison across countries. Using the lower bound of

---

3A central reason we prefer the medical-based approach is that we aim to produce conservative estimates and the
VSL arrives often at much higher values, up to millions of euro or dollars [54].
the central range yields then the following calibration of the medical costs:

\[(0.05 + 0.09 + 0.01 \times 1.5 \times 11) \times \text{GDP}_{p.c.} = 0.305 \times \text{GDP}_{p.c.}\]  

(6)

The upper bound for the value of \( k \) would be substantially higher: \((0.05 + 0.09 + 0.01 \times 4 \times 11) \times \text{GDP}_{p.c.} = 0.58 \times \text{GDP}_{p.c.} \). For the numerical calculations we will use the conservative estimate \( k = 0.305 \) in terms of GDP\(_{p.c.}\).

If we only consider the direct medical costs consisting of loss of working time and hospitalization, without including the value of lives lost, the proportionality factor in Eq. 6 reduces to \( k = 0.14 \) in terms of GDP\(_{p.c.}\).

**Medical costs over the lifetime of the epidemic**

The cost estimates discussed so far, \( c^\text{med}_t \), refer to the per-period cost of the currently infected. For the total cost over the entire endemic we need to calculate the discounted sum of all \( c^\text{med}_t \) over time. Given that a period corresponds to about two weeks, we neglect discounting, which would make little difference even if one uses a social discount rate of 5% instead of using market rates (which may be negative). The total medical costs over the course of the epidemic can be written as the simple sum of the cost per unit of time:

\[C^\text{medical} = \sum_{I_t \geq I_{\text{min}}} c^\text{med}_t = kX_{\text{tot}}.\]  

(7)

The epidemic is considered to have stopped when the fraction of new infections \( I_t \) falls below a minimal value, \( I_{\text{min}} \).

Using the conservative estimate (low value of life) \( k = 0.305 \) it is straightforward to evaluate the total cost of a policy of not reacting at all to the spread of the disease, which would lead in the end to \( X_{\text{tot}} = 0.94 \), as pointed out in the Methods section. A hands-off policy would therefore lead to medical costs of over 28% of GDP\(^4\).

In absolute terms the cost of a policy of doing nothing would amount to 10000 billion euro for a country like Germany. For the US the sum would be closer to 5 Trillion of dollars (25% of a GDP of 20 Trillion of dollars). As it would not be possible to ramp up hospital capacity in the short time given the rapid spread of the disease, the cost would be in reality substantially higher, together with death toll [13, 14, 15]. We abstract from the question of medical capacity (limited number of hospital beds) because we assume that society would react anyway as the virus spreads, thus limiting the peak, and, second, we are interested in the longer term implications of different strategies and not just in their impact on the short-term peak.

\(^4\)It is well documented that there are large differences in the severity of the symptoms following an infection. Isolating the group at risk (essentially the elderly), and waiting for a vaccination to be developed, would reduce medical costs substantially. However, to which extent this strategy could be carried out in practice is to be seen.
We note that even concentrating only on the direct medical cost and working time lost \((k=0.14)\) a policy of letting the epidemic run its course through the entire population would lead to losses of working time and hospital treatment of over 13% of GDP (94% of 14%). By comparison, total health expenditure in most European countries amounts in normal times to about 11% of GDP [61]. Even apart from ethical considerations, to avoid or not potentially hundreds of thousands of premature deaths, there exists thus an economic incentive to slow the spread of the COVID-19 virus.

Given the somewhat contentious nature of the value of lives lost, we present in the middle panel of Fig. 4 and Fig. 5, the medical cost estimates (as a proportion of GDP) without and with including the value of life costs, respectively. As shown in the figures, increasing \(\alpha\) leads to a lower medical cost because the percentage of the population infected will be lower. The difference between a shortsighted and a history-aware control increases for higher values of \(\alpha\). At these \(\alpha\) values the medical cost over the entire endemic would be lower because the overall fraction of infected population is lower. For a strongly reactive society and policy i.e. for \(\alpha \gg 1\) (and the case of historical reaction) one can use Eq. 4 for an explicit solution for the total health cost,

\[
C_{\text{medical}} = kX_{\text{tot}} \bigg|_{\alpha \gg 1} \approx 2k \frac{g_0 - 1}{\alpha}
\]

which implies that the total health or medical costs are inversely proportional to the strength of the policy reaction parameter. Draconian measures from the start, i.e. \(\alpha\) going towards infinity reduce the medical costs to close zero - irrespective of whether one adds the value of lives lost. This can be seen in Figs. 4 and 5 where the medical cost (over the entire epidemic) starts for \(\alpha = 0\) at values close to \(k\) because without any societal reaction 94% of the population would get infected and with increasing \(\alpha\) the medical costs decline monotonously.

Social distancing costs

The economic costs of imposing social distancing on a wider population are at the core of policy discussions and drive financial markets. As mentioned above, social distancing can take many forms; ranging from abstaining from travel or restaurant meals to government interventions enforcing lockdowns, quarantine, closure of schools, etc. This cost is more difficult to estimate. However, a rough estimate is possible if one takes into account that most economic activity involves some social interactions. Limiting social interaction thus necessarily reduces economic activity. This suggests that the economic cost of the social distancing described in Eq. 2 should increase with the reduction in the transmission rate described by \(g_t\).

Without any social distancing, \(\alpha = 0\), the economy would not be affected by the spread of the virus. Stopping all economic social interactions would bring the economy to a halt, but the reproduction rate of the virus would also go close to zero (Eichenbaum et al. [21] make a similar assumption). We thus posit that the (per-time unit) social-distancing economic cost \(c_s^t\) is proportional to the reduction in the transmission rate. The total economic costs \(C_{\text{social}}\) can be written as
Figure 4: **Cost of epidemic control without value of life.** As a function of $\alpha$, the results for ‘history-aware’ and ‘short-sighted’ control defined by Eq. 2. Shown are the costs incurring from social distancing (lowest panel), as given by Eq. 9 with $m = 0.25$, the pure medical costs without value of life costs (middle panel), see (6) and the sum of social and medical costs (top panel). The epidemic has been considered to have stopped when the fraction of actual cases $I_t$ dropped below a threshold of $I_{\text{min}} = 10^{-5}$. Costs are in terms of fraction of GDP, the starting $I_0 = 2 \cdot 10^{-5}$.

The key question is the factor of proportionality, $m$, which links the severity of social distancing to the reduction in economic activity. Popular attention has focused on services linked directly to social contact. There exist indeed some sectors which will completely shut down under a lockdown. However, these sectors (tourism, non-food retail, etc.) account for a limited share of the economy (less than 10% for most countries). Expenditure for food is actually little affected since even under the most severe lockdown, grocery shopping is still allowed and families must consume more food at home as they cannot go out to restaurants.
Figure 5: **Cost of epidemic control including the value of life.** As in Fig. 4 (bottom panels are identical), but with the medical costs including the value of life saves.

The manufacturing sector is less affected by social distancing than the service sector because in modern factories workers are scattered over a large factory floor, making it relatively easy to maintain production while maintaining the appropriate distance between workers. Moreover, some sectors, e.g. finance, can work online with only a limited effect on productivity. The widespread impression that the entire economy stops under a lockdown is thus not correct. The drastic measures adopted in China illustrate this proposition: when all non-essential social interactions were forbidden, industrial production and retail sales fell by 'only' 20-25% [3] while the reproduction factor went from 3 to 0.3, a fall by a factor of ten. Using this experience we calibrate the parameter \( m \) at 0.25.

A reduction in the reproduction factor \( g_t \) to one tenth its normal epidemiological value of \( g_0 \) would thus lead to a loss of GDP of 25% for the time period during which the restriction or social distancing measures are in place. This would imply that a Chinese-type shutdown of the economy to 25% of its capacity for 12 weeks, or 6 incubation periods would cost about \( 0.25 \times (12/52) \) or about 6% of annual GDP. A reduction of GDP by 6% would represent a recession even deeper than the
one which followed the financial crisis of 2009. This is compatible with current forecasts of zero GDP growth in China in 2020 (relative to a baseline of 5-6% before the crisis). But even such a large cost in terms of output foregone would be below the medical cost arising from herd immunity. Even apart from ethical considerations, it would thus appear to make sense to accept a temporary shut down of parts of the economy to avoid the huge medical costs.

A first result is thus that if one compares two extremes: letting contagion run its course (herd immunity) or draconian measures, the social costs are lower in the second case. Small changes to the key parameters, $k$ and $m$, might change the exact values of the costs in terms of overall magnitude, but the ranking appears robust.

We do not consider separately the fiscal cost, i.e. the cost for the government to save millions of enterprises from bankruptcy and ensure that workers have a replacement income when they get laid off. This cost to governments is a transfer within the country from one part of society (taxpayers) to those who suffer most under the economic crisis.

A key issue in the discussion on the economic cost of social distancing is the question about how long these measures need to be maintained. It is sometimes argued that the cost of a policy of social distancing would be unacceptably high because the measures could not be relaxed until the virus had been totally eradicated. However, this pessimism is not warranted by the success of a strategy of ‘testing and tracing’ implemented in some countries (mainly those which had experienced SARS). Such a strategy is, of course, only possible if the starting number of infections is low enough to allow for individual tracing.

We thus make the assumption that when the number of active cases falls below a certain threshold, the costly measures of general social distance containment are no longer needed and can be substituted by pro-active repeated testing coupled to quick follow-up of the remaining few cases which are quarantined and whose contacts are quickly traced. In this case the resulting economic cost is assumed to fall away. The experience of Singapore and Japan suggests that when the infected are less than one per 100,000, general social distancing is no longer required (assuming mass testing has been adopted in the meantime so that the infections can be accurately measured).

**Cumulative health and social-distancing costs**

We have so far discussed only the costs of two extremes: letting contagion run its course (herd immunity) or draconian measures from the start. However, as we have emphasized above, most countries show a more gradual reaction, described by the parameter $\alpha$. The total expected medical and social-distancing economic costs can then be calculated as the discounted sum of the per-period costs from any arbitrary starting period. As mentioned above, we abstract from discounting given market interest rates close to zero and short time period over which the endemic plays out. A social discount rate between 3 and 5 [64] would make little difference over the course of one year.

---

5The abrupt slowdown in the economy could of course have further consequences for health, which might offset the benefits from fewer Coronavirus infections. However, this should not be the case because, somewhat surprisingly, several studies found that recessions lead to lower mortality rates [62, 63].
In the absence of discounting, the total medical cost incurred over the duration of the epidemic (here +/- one year) is proportional to the overall fraction $X_{t \to \infty}$ of infected, with the factor of proportionality $k$. However, the sum of the economic costs depends in a non-linear way on the evolution of new cases (short-sighted control) or the percentage of the population infected (history-aware control).

The total cost of different policies can now be compared by computing the sum of future costs under different values for $\alpha$ in Eq. 2 as shown in Figs. 4 and 5. The three panels of the figures show the total cumulated costs of different policy regimes as expressed by the value of $\alpha$. The two key common parameters in these simulations are: $k$ and $m = 0.25$.

We distinguish two cases: In Fig. 4 only the direct medical costs and working time lost are considered. This corresponds to $k = 0.14$. In Fig. 5 the value of lives lost is taken additionally into account, resulting in $k = 0.305$ The respective middle panels show, as discussed further above, that a short-sighted society will incur higher medical costs, because restrictions are relaxed after the peak. By contrast, if policy (and individual behavior) is influenced by the total number of all cases experienced so far, restrictions will not be relaxed prematurely and the medical costs will be lower for all values of $\alpha$.

The bottom panels in Figs. 4 and 5 are identical for comparison purposes. They show the social distancing costs as a fraction of GDP, which represent a more complicated trade-off between the severity of the restrictions and the time they need to be maintained. If neither policy, nor individuals react to the spread of the disease ($\alpha = 0$) the epidemic will take its course and costs are solely medical. This changes as soon as society reacts, i.e. as $\alpha$ increases. Social distancing costs increase initially (i.e. for small values of $\alpha$) more for the history-aware case than for the short-sighted reaction framework. The situation reverses for higher values of $\alpha$ with $\alpha \approx 30$ being the turning point. From there on, the distancing cost from a history-aware reaction falls below that of the short-sighted strategy.

The sum of the two costs is shown in the uppermost panel of Figs. 4 and 5. It is here that one observes a different pattern for the two cases considered (with and without the value of life). In Fig. 4 (without value of life costs) one observes a hump-shaped pattern: For the history-aware reaction function, the cost over the entire epidemic increases for small values of $\alpha$, falling as the policy becomes more reactive. For the short-sighted policy the opposite is the case. When one takes into account also the value of life costs, the hump-shaped pattern in Fig. 4 is nearly not observable. The total costs decrease fast, starting at low values of $\alpha$ for the history-aware reaction, while they increase for the short-sighted policy. What is common to both figures is that for large values of $\alpha$ the short-sighted policy results always in higher costs.

The inverted U shape of the total cost of the virus as a function of $\alpha$ has one important corollary: the 'laissez faire' equilibrium is not the optimum for society. If the government abstains from action leaving it to societal reaction to dampen the peak, the spread of the disease would be governed only by $\alpha_s$, which might bring society close to the hump of the total cost curve. Strong government action, i.e. a high value of $\alpha_g$ could then push the path to the other side of the hump resulting in lower costs. In other words, relying only on individual reaction which aims at lowering the risk to oneself, would be suboptimal. This is of course a general result for all contagious diseases [8, 7] but
we confirm it accounting explicitly for the cost of the measures needed to protect public health.

Discussion

A central result of our study is that strong suppression strategies lead to lower total costs than taking no action, when containment efforts are not relaxed with falling infection rates. A ‘short-sighted’ approach of softening containment with falling numbers of new cases is likely to lead to a prolonged endemic period. With regard to the ‘exit strategy’ discussion, these findings imply that social distancing provisions need to be replaced by measures with comparative containment power. A prime candidate is in this regard to ramp up testing capabilities to historically unprecedented levels, several orders of magnitude above pre-Corona levels. The epidemic can be contained when most new cases can be tracked, as expressed by the factor $\theta$ in Eq. 5. This strategy can be implemented once infection rates are reduced to controllable levels by social distancing measures. Containment would benefit if the social or physical separation of the ‘endangered’ part of the population from the ‘not endangered’ would be organized in addition on a country-wide level, as suggested by community-epidemiology. With this set of actions the vaccine-free period can be bridged.

The estimates on which our results are based will have to be updated when actualized COVID-19 data is available in the future. The WHO-China Joint Mission Report suggests a $g_0$ of $2 - 2.5$ [52], while we use the figures from Liu et al. [18], who predict a reproduction factor of around 3. The numbers for the forecast of health costs are derived in part from the Diamond Princess data [47], for which the population was comparatively healthy. The statistics for symptoms requiring the absence from work may therefore in reality be somewhat higher. The hospitalization and mortality rate are estimates with a substantial uncertainty, due to the high numbers of unregistered and untested infections. Two studies from Japan [65] and China [66] estimate that the number of actual infections may be between 10 to 20 times higher than the number of detected infections, resulting in possibly lower true hospitalization and mortality rates for COVID-19. One of our main goals has therefore been the introduction of a generic framework, which can be updated by future advances in the accuracy of estimates while still presenting specific results with the data available at this time.

As a last note, there is a commonly voiced misconception regarding the meaning of the herd immunity point, which occurs for an infection factor of three when 66% of the population is infected. While falling beyond the herd immunity point, the infected-case counts remain elevated for a considerable time. The outbreak stops completely only once 94% of the population has been infected, as illustrated in Fig. 1. The view that the epidemic is essentially over once the herd immunity point is reached does not reflect the admittedly idealized SIR model.
Methods

Continuous-time SIR model

We denote with \( S = S(t) \) the fraction of susceptible (non-affected) people, with \( I = I(t) \) the fraction of the population that is currently ill (active cases), and with \( R = R(t) \) the fraction of recovered. Normalization demands \( S + I + R = 1 \) at all times. The continuous-time SIR model,

\[
\dot{S} = -rSI, \quad \dot{I} = rSI - \gamma I, \quad \dot{R} = \gamma I,
\]

(10)
obeys normalization, as \( \dot{S} + \dot{I} + \dot{R} = 0 \). Infection and recovery rates are \( r \) and \( \gamma \). The number of infected grows as long as \( \dot{I} > 0 \), namely when \( rS > \gamma \). Herd immunity is consequently attained when the fraction of yet unaffected people dropped to \( S = \gamma/r \). Making connection with the discrete case, (1), the herd immunity condition becomes

\[
1 = g(1 - X) = gS \equiv \frac{r}{\gamma}S, \quad X = I + R = 1 - S,
\]

(11)

which yields \( g = r/\gamma \). Epidemic spreading is observed for \( r > \gamma \).

0.1 SIR model with history-aware control

Control does not affect the medical recovery rate \( \gamma \), but the infection rate \( r \). History-aware control, Eq. 2, corresponds then to

\[
r = \frac{r_0}{1 + \alpha(1 - S)}, \quad \frac{r_0}{\gamma} = g_0.
\]

(12)

A functional relation between \( S \) and \( I \) is obtained considering \( \dot{I}/\dot{S} \), which results in

\[
dI = -dS + \frac{1}{g(S)S} dS = -dS + \frac{1}{g_0} \frac{1 + \alpha(1 - S)}{S} dS.
\]

(13)

Integrating leads to

\[
I = - \left( \frac{\alpha}{g_0} + 1 \right) S + \frac{1 + \alpha}{g_0} \log(S) + c,
\]

(14)

where the integration constant \( c \) is given by the condition \( I(S=1) = 0 \). We hence obtain

\[
I = \frac{\alpha + g_0}{g_0} X + \frac{1 + \alpha}{g_0} \log(1 - X),
\]

(15)

when substituting \( S = 1 - X \). The number of actual cases, \( I \), vanishes both when \( X = 0 \), the starting point of the outbreak, and when the epidemic stops. The overall number of cases, \( X_{\text{tot}} \), is obtained consequently by the non-trivial root \( X_{\text{tot}} \) of (15), as illustrated in Figure 1. Expanding (15) in \( X \), which becomes small when \( \alpha \gg 1 \), one obtains

\[
X_{\text{tot}}\big|_{\alpha \gg 1} \approx 2 \frac{g_0 - 1}{\alpha},
\]

(16)

which is asymptotically exact for large control.
0.2 Peak medical load

The maximal number \( I_{\text{max}} \) of actual cases, the peak medical load, is attained at the herd point, when \( \dot{I} = 0 \),

\[
gS = 1, \quad g_0S = 1 + \alpha(1 - S), \quad S_{\text{herd}} = \frac{1 + \alpha}{g_0 + \alpha},
\]

see (11). For the epidemic state, \( g > 1 \), one has \( S_{\text{herd}} < 1 \), with \( \lim_{\alpha \to \infty} S_{\text{herd}} = 1 \). With (17) we obtain

\[
I_{\text{max}} = \frac{g_0 - 1}{g_0} + \frac{1 + \alpha}{g_0} \log \left( \frac{1 + \alpha}{g_0 + \alpha} \right)
\]

for the maximal number of actual cases. For strong control, \( \alpha \gg 1 \), the expansion

\[
I_{\text{max}}\big|_{\alpha \gg 1} \approx \frac{(g_0 - 1)^2}{g_0 \alpha}, \quad I_{\text{max}}\big|_{\alpha = 0} = 1 - \frac{1 + \log(g_0)}{g_0}
\]

holds. Without control, when \( \alpha = 0 \), close to the entire population is infected simultaneously when \( g_0 \) is large. Combining (19) with (16) one obtains the large-\( \alpha \) relation

\[
X_{\text{tot}} \approx \frac{2g_0}{g_0 - 1} I_{\text{max}}
\]

between the total impact of the epidemic, \( X_{\text{tot}} \) and the maximum number of actual cases \( I_{\text{herd}} \). This relation can be used to estimate \( X_{\text{tot}} \) once the peak has been reached.

Acknowledgments

We thank Erik Gros for carefully reading the manuscript. D.G. thanks the Fulbright foundation for financial support.

References


Ying Liu, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine*, 2020.


