The first three months of the COVID-19 epidemic: Epidemiological evidence for two separate strains of SARS-CoV-2 viruses spreading and implications for prevention strategies

Two epidemics of COVID-19

Knut M. Wittkowski¹*

¹ ASDERA LLC, New York, NY
* Corresponding author:
E-mail: knut@asdera.com (KMW)

Abstract

About one month after the COVID-19 epidemic peaked in Mainland China and SARS-CoV-2 migrated to Europe and then the U.S., the epidemiological data begin to provide important insights into the risks associated with the disease and the effectiveness of intervention strategies such as travel restrictions and social distancing. Respiratory diseases, including the 2003 SARS epidemic, remain only about two months in any given population, although peak incidence and lethality can vary. The epidemiological data suggest that at least two strains of the 2020 SARS-CoV-2 virus have evolved during its migration from Mainland China to Europe. South Korea, Iran, Italy, and Italy’s neighbors were hit by the more dangerous “SKII” variant. While the epidemic in continental Asia is about to end, and in Europe about to level off, the more recent epidemic in the younger US population is still increasing, albeit not exponentially anymore. The peak level will likely depend on which of the strains has entered the U.S. first. The same models that help us to understand the epidemic also help us to choose prevention strategies. Containment of high-risk people, like the elderly, and reducing disease severity, either by vaccination or by early treatment of complications, is the best strategy against a respiratory virus disease. Social distancing or “lockdowns” can be effective during the month following the peak incidence in infections, when the exponential increase of cases ends. Earlier containment of low-risk people merely prolongs the time the virus needs to circulate until the incidence is high enough to initiate “herd immunity”. Later containment is not helpful, unless to prevent a rebound if containment started too early.

About the Author

Dr. Wittkowski received his PhD in computer science from the University of Stuttgart and his ScD (Habilitation) in Medical Biometry from the Eberhard-Karls-University Tübingen, both Germany. He worked for 15 years with Klaus Dietz, a leading epidemiologist who coined the term “reproduction number”, on the Epidemiology of HIV before heading for 20 years the Department of Biostatistics, Epidemiology, and Research Design at The Rockefeller University, New York. Dr. Wittkowski is currently the CEO of ASDERA LLC, a company discovering novel treatments for complex diseases from data of genome-wide association studies.
# Table of Content

**Introduction** 3

**Materials and Methods** 4
- Data 4
- Methods and models 4
  - Statistical and Bioinformatics Methods 4
  - Epidemiological Models 4

**Results** 5
- Incidence by Country. Norther Hemisphere 5
- Time-course by country/region 7
- Incidence by Country. Southern Hemisphere 13
- Modeling Results for Effectiveness of Containment (Social Distancing) 14
- Responses by country in relation to the turning/decision point 17
  - Asia 17
  - Europe 18
  - North America 18
  - Oceania 18

**Discussion** 19
- Strengths and shortcomings 19
- Evidence for (at least) two different strains of SARS-CoV-2 19
- Changes in infectivity and lethality between China and Europe 21
- Predictions for COVID-19 in North America 22
- A historical perspective 22
- Implications for prevention 22

**Conclusions** 23

**References** 24
Introduction

The first cases of a new coronavirus strain, termed SARS-CoV-2 (Severe Acute Respiratory Syndrome CoronaVirus) by the International Committee on Taxonomy of Viruses (ICTV) (Cascella 2020), were reported on 31-12-2019 in Wuhan, the capital of the Hubei province of China. (Jernigan 2020) As of 2020-04-03, 10:00 CET, 1,082,054 symptomatic cases and 58,142 deaths have been reported from virtually every country in the northern hemisphere (see Section Data). The disease was termed COVID-19 by the WHO on 2020-02-11, and categorized as a pandemic on 2020-03-12, yet the details of the spread and their implications for prevention have not been discussed in sufficient detail.

As the virus moved from China via Europe to the U.S., several administrations have imposed “lockdown” restrictions.

For most of the first three months of the epidemic, much of the response was driven by “fear, stigma, or discrimination” (Ren 2020), including naming SARS-CoV-2 the “China virus” (Rogers 2020), despite the fact that seasonal respiratory zoonotic pathogens typically originate in China, where life-animal markets provide chances for animal viruses to transmit to humans. (Malik 2020) On 03-23, the Veterans Health Administration issued a COVID-19 Response Plan envisioning a pandemic that would “last 18 months or longer and could include multiple waves of illness”. (Stone 2020)

Between 02-14 and 03-16, the Dow Jones fell 31% from 29,440 to 20,188, raising fears for the economy, in general, and retirement savings, in particular. By the end on 03-20, the Dow Jones was down at 19,173 (35%) from 02-14. On 03-26, the U.S. Senate approved a $2T stimulus package.

After three months, enough data are available to discuss important epidemiological characteristics of COVID-19 and the potential impact of interventions. In particular, we have now seen the number of new cases (and deaths) to decline in China and South Korea and to at least stabilize in some European countries. Changes in number of deaths follow the changes in number cases (albeit at a lower level) by about two weeks. Hence, we can discuss both the infectiousness and the lethality of the virus, two important characteristics to assess public health impact of the disease.

One of the key findings published herein is evidence that at least two strains of SARS-CoV-2 with different infectiousness and lethality have evolved, and by following the likely path for each of these strains we can obtain novel insights into the nature of the epidemic and, thus, the effectiveness of prevention strategies. Another finding is that the timepoint when a public health intervention starts in the course of the epidemic (especially the “turning point” where the increase in new cases begins to decline) is crucial for the impact of the intervention.
Materials and Methods

Data
All data were downloaded on 2020-04-04 from the European Centre for Disease Prevention and Control (ECDC) Web site at https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide, where data are collected daily between 6:00 and 10:00 CET. Updates were collected from the Johns Hopkins online tracker available at https://systems.jhu.edu/research/public-health/ncov/. New York City data was downloaded from https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-daily-data-summary.pdf. Population data were accessed from https://www.worldometers.info/world-population/population-by-country/ on 2020-03-12. Data on ages by country were accessed from https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS.

Methods and models

Statistical and Bioinformatics Methods
The data were processed with MS Excel. To avoid biases from inappropriate model assumptions only basic descriptive statistics were employed. In some cases, data from only the day before or after (or both) was averaged (up to a three day moving average) to reduce the effects of apparent reporting artifacts (Darwin’s *natura non facit saltum*,[Berry 1985]) without creating undue biases. The two “smoothers” applied were:

- averaging $x_0$ with a previous $x_{-1}$ (or, rarely, following $x_{+1}$) data and
- applying a moving average of $(x_{-1}, x_0, x_{+1}) \rightarrow (2x_0 - x_{-1} + x_{+1}, x_0, x_0 + x_{+1} + 2x_{+1})/3$

No other changes were applied to the data.

Like China in mid-February, the German Robert-Koch-Institute (RKI) changed the reporting system in mid-March, which resulted in a near 6-fold increase of the data reported on 03-20 over the data reported on 03-19. To account for such isolated event, the data of this and several previous days was averaged (as indicated in the Figures) under the assumption that those changes reflect cases that had previously been diagnosed, but were not reported.

To guide with visual impression, moving averages of the 4 days leading up to the current day were added to the Figures.

Epidemiological Models
If a disease causes immunity after an infectious period of a few days only, like respiratory diseases, an epidemic extinguishes itself as the proportion of immune people increases. Under the SIR model,[Kermack 1991] for a reproduction number[Dietz 1993] (secondary infections by direct contact in a susceptible population) of $R_0=1.5–2.5$ over 7 days (recovery rate: $\gamma=1/7=.14$), the noticeable part of the epidemic lasts about 90–45 days ($R_0/\gamma=\beta=.21–.36$) in a homogeneous population of 10M. The period is shorter for smaller more homogenous and longer for larger, more heterogeneous populations. For a given infectious period $1/\gamma$ (here, e.g., 7 days. SARS and COVID-19 incubation period plus 2 days[Lauer 2020]), $R_0$ also determines how long it will take for early cases to become visible after a single import (150–60 days), the peak prevalence of infections (5–22%), and how many people will become immune (55–90%). To allow for comparisons between models, an arbitrary proportion of symptomatic cases among those becoming infected (.05%) is used and 2% of cases are assumed to die.
Fig 1: SIR Model of SARS. Number of susceptible (blue), infectious (red), resistant (green), case (orange) and dead (gray) people after a population of 10,000,000 susceptible people is exposed to 20 subjects infected carrying a novel virus. Assumptions: $R_0 = 2.2$, infectious period = 7 days. 

The model reflects that the key milestones of the epidemic, the turning point/half maximal point in infections (red, day 68), the turning point/half maximal point in the number of cases (day 75, orange), the peak in infections (day 83, red), the peak in cases (day 70, orange) and the peak in deaths (day 77, gray) follow each other, following the previous milestone after about a week.

Results

Incidence by Country. Norther Hemisphere

Table 1 shows the raw daily incidence by population sizes for countries with epidemiological relevance in the northern hemisphere. Countries within proximity are grouped by their peak incidence (red background).

It should be noted, however, that there is no uniform definition of “cases”. In some countries a case needs to have symptoms, in other countries, it suffices to have antibodies (be immune).

The Hubei province in China (with the capital Wuhan), South Korea, Iran, Italy (especially the Lombardy region), and Spain have the highest peak incidence, followed by the countries neighboring Italy. In several European countries, incidence is declining. Only in Spain (peak incidence) did the incidence exceed 130/M/d.
Table 1: Incidence by Country. Dates: Feb 13 (peak intensity in Mainland China, mostly Hubei and neighboring provinces), Feb 19 to Mar 22. Countries with low population size (PSz) or low number of cases (Total) are hidden. Red background indicates countries/dates with high incidence per capita. Countries/regions are grouped by proximity among each other and distance from Mainland China.

<table>
<thead>
<tr>
<th>Country</th>
<th>Region</th>
<th>Date</th>
<th>Population</th>
<th>Total Population</th>
<th>Total/M</th>
<th>PSz</th>
<th>PSz/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>Mainland China</td>
<td>Feb 13</td>
<td>1.430</td>
<td>2.927</td>
<td>2.20</td>
<td>11.4</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Hubei</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Henan</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Sichuan</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Jiangsu</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Zhejiang</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Guangdong</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Shandong</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Anhui</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Fujian</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Hebei</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Shanxi</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Shaanxi</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Neurologically impaired</td>
<td>Feb 22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Knut M. Wittkowski: Two epidemics of COVID-19 (2020-04-04) - 6 -
Time-course by country/region

Among the Hubei population of 58.5M, the incidence rose from the first case reported in late 2019 to about 60 new cases per million people per day by 02-05 and then steadily declined (Fig 1) from ~4000 on 02-05 to below 50 cases per day since 03-08. Since 03-31 no cases were reported in Vietnam and only isolated cases in Cambodia.

Fig 2: COVID-19 cases in Mainland China. Blue: cases/M/d, red: deaths/M/d. Around Feb 13, the case definition was expanded, resulting in additional cases from previous days being added. Hence, the 02-13 cases have been truncated. Most cases were seen in the Hubei province of 58.5M people (see Table 1 for population sizes).

By Mid-January 2020, the first cases of COVID-19 were seen in other Asian countries, but incidence remained below about 3/M/d outside of continental China; only Malaysia/Brunei and Singapore have raised above10/M/d. Japan with the largest proportion of people >65 years of age in the world (28%) remained below 1.1/M/d until recently, when incidence increased slightly to ~2.0/M/d. Overall, the steady increase could indicate the begin of a new outbreak.

Fig 3: COVID-19 cases in Maritime Asia. See Fig 1 for legend.

In continental South Korea (population 51M, cumulative incidence 199/M), the incidence rose to a peak of about 14/M/d between 02-29 and 03-02, before declining to continuing low rate of less than 150 cases per day (~2/M/d) since 03-12 (Fig 2a).

In Iran (cumulative incidence 633/M), incidence began to rise about a week after South Korea. The top incidence before a new wave started on 03-23 (~15.5 cases/M/d) was about the same. The increase thereafter with a pea around 35–40M/d may indicate a “rebound” into a population not sufficiently immunized by the previous wave(s). Lethality in Iran was notably higher and followed the increase in cases (see the SIR model, above) with a delay of several days (Fig 2b).
From 03-19 to 03-20, several European countries have seen a more than two-fold increase in the number of cases reported (Germany: 570%, San Marino: 340%, Ireland: 260%, Switzerland: 240%, Austria: 202%). As *natura non facit saltum* (Darwin: nature doesn’t jump), such abrupt increases must be, at least in part, the result of reporting or other artifacts. In Germany, for instance, the reporting system was changed between 03-16 and 03-19, so that the number likely includes cases previously reported only through a parallel system. France, Italy, and Spain also reported an unusual increase by 27–35 percent. Since 3-26, all these countries reported the incidence to decline. France, however, has seen a spike on 04-04.

Italy’s neighbors Spain and Switzerland now have a higher cumulative incidence per capita than Italy (Table 1, Fig 3), all higher than that in the Hubei province (their all with a population of 50–60 M).
In Italy (Fig 5), incidence has been declining from the 100/M/d peak for 13 days, after rising for about 4 weeks from 1/M (02-26~03-22). In Hubei and South Korea, instead incidence took only 2 weeks to peak (01-19~02-05, Fig 2, and 02-21~03-06, Fig 4) and then declined by at least 85% within another 2 weeks. The much slower and decline in Italy (compared, e.g., to China and Hubei)
is consistent with insufficient herd immunity having been developed before the peak incidence was reached (Fig 1). Other European countries peaked in 03-27/28 (Fig 6). On 04-04, France added previously unreported cases from nursing homes. Otherwise, the incidence among Italy’s neighbors is declining, while the incidence in other European countries is still increasing.

Fig 6: COVID-19 cases in Europe. Early onset/high lethality (IT and neighbors, top), total (middle), and late onset/low lethality (other European countries, bottom). See Fig 1 for legend.
Fig 7: COVID-19 cases in selected European countries. See Fig 1 for legend. Data in Germany is based on cases reported electronically to the Robert-Koch-Institute (RKI) and transmitted to the ECDC, but the RKI also provides two sources of data on its Web site that are difficult to reconcile with these data. To account for a temporary peak in Germany on 03-20/21, when the reporting system was changed, the 03-14..03-21 were averaged. 6082 cases have been added for Germany based on (https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationenberichte/2020-04-04-de.pdf?__blob=publicationFile). To account for temporary peaks in Spain and France on 04-01, the data from 03-30..04-01 were averaged in both countries. On 04-04, France added additional data from nursing homes (https://www.reuters.com/article/us-health-coronavirus-france-toll/french-coronavirus-cases-jump-above-chinas-after-including-nursing-home-tally-idUSKBN21L3BG).
Germany is notable for reporting only 1,109 deaths among 87,495 cases (1.3%, by 04-04), less than even Austria and much less than France, and, in particular, Italy and Spain.

Of particular interest is Sweden, where no “social distancing” policy was implemented. Still, there was no difference in the shape of the epidemic or the height of its peak to the other Scandinavian countries:

![Swedish COVID-19 cases](image1)

![COVID-19 cases in Scandinavia](image2)

**Fig 8: COVID-19 cases in the Scandinavia.** See Fig 1 for legend.

The epidemic in North America started later, especially in the US (except for a few isolated cases likely imported directly from Asia). The increase in incidence is consistent with the dynamics of an earlier epidemic and the cumulative incidence in the US (840/M) is now higher than in Europe (546/M), although the epidemic arrived later. The US may have reached the “turning point”, where the rate of new cases begins to decline (when 50% of the peak incidence is reached) at about 03-27, compared to Europe’s 03-22. Levels in Canada remain below those in the U.S. In New York City, which is particularly hard hit, the number of new cases seem to be leveling off and lethality should follow suit with a delay of about a week.
Incidence by Country. Southern Hemisphere

Most parts of the southern hemisphere (with the possible exception of Chile) have seen only few cases, but Oceania (Australia and New Zealand) shows evidence for an epidemic with a peak incidence on 03-26/27. As in Canada and, earlier, in Germany and other European countries,
isolated spikes tend to reflect delayed reporting, rather than changes in incident trends. The number of deaths (20/5204 = 0.4%) is extremely low (Fig 10).

![Fig 10: COVID-19 cases in the Oceania (Australia and New Zealand). See Fig 1 for legend.](image)

**Modeling Results for Effectiveness of Containment (Social Distancing)**

The effect of reducing the reproduction number by reducing the number of contacts ("containment", "Social Distancing") depends on when it starts in the course of the epidemic. Fig 11 shows the effect of a one-month intervention cutting $R_0$ in half starting at the point of the peak prevalence of infectious subjects. Compared to Fig 1, the duration of the epidemic is shortened, albeit at the price of reducing the R/S ratio, so that a subsequent epidemic with the same or a similar virus (cross-immunity) could start earlier.

![Fig 11: SIR Model of SARS, Window of Opportunity for Fast Eradication of the Epidemic. (see Fig 1 for legend).](image)

The gray area indicates the period where containment can give a "coup de grace" to a respiratory disease epidemic. The more narrow bell curve with a post-peak intervention indicates the reduction in number of infections and, thus, deaths. (spreadsheet for model calculations available from [https://app.box.com/s/pa444z1csxcvkg11oohjm3dg66qj](https://app.box.com/s/pa444z1csxcvkg11oohjm3dg66qj))

Fig 12 shows a one-month intervention starting about two weeks earlier, at the turning point where the curve of the new cases changes from increasing faster to increasing more slowly. This intervention reduces the number of deaths, but the epidemic is extinguished two months later and the R/S ratio ("herd immunity") is further decreased.
Fig 12: SIR Model of SARS, Window of Opportunity for Maximal Reduction of Total Deaths. (see Fig 1 for legend). The gray area indicates the period where containment can have the most impact on total number of deaths. However, the epidemic is not eradicated. (spreadsheet for model calculations available from https://app.box.com/s/pa446z1csxcvfksgj13ooihm3b9g66qj)

The SIR model is flexible enough to reflect a “phased in” intervention, which starts with a period of lower intensity (higher $R_0$) to allow people to adjust, before rolling in a more restrictive intervention (higher $R_0$). The Fig 13a shows the intervention of Fig 12 starting with a week of reduced intensity. Compared to Fig 12, reducing intensity during the period where herd immunity reduces the rebound, but – somewhat counterintuitively – also the number of deaths. Extending the duration of the intervention (almost) extinguish the disease (and further reduce number of deaths by 10%).
Fig 13: SIR Model of SARS, Phased in Restrictions. (see Fig 1 for legend). The gray areas indicate the periods of low (5 d) and high intensity (22/40 d) restrictions.

Fig 14 shows the detrimental effect of an intervention that starts even earlier, about two weeks before the turning point (Fig 12). Even if the intervention is extended from one to four months, no herd immunity is created and, thus, the epidemic rebounds and will run eight months, instead of three (Fig 1) or less (Fig 11).

Fig 14: SIR Model of SARS, Effect of Early Social Distancing / Lockdown. (see Fig 1 for legend). It is assumed that a highly effective intervention reduces R₀ by 50% for 4 months, beginning after the appearance of a novel type of cases is noticed. The proportion of symptomatic cases (0.05%/d, i.e., .35% of infected people will become cases and the proportion of cases to die (2%) may change, but the issues discussed here are broadly independent of these assumptions. (spreadsheet model calculations available from https://app.box.com/s/pa446z1csxc0ksg130oojim3b2gk6qj)

The impact of the interventions discussed on the infections (and, thus, the cases and immunity) depend mostly on the homogeneity of the population, but the number of deaths is highly dependent on how the people respond to the intervention. If parents respond to closure of schools by
asking the grandparents to take care of the children, for instance, more of the elderly will get infected and the number of deaths will increase in ways the SIR model cannot predict.

In summary, there is a narrow window-of-opportunity for interventions to “flattening the curve” (reducing \( R_0 \)) to be successful in terms of public health:

- Starting after the peak prevalence (of infections) has little effect (not shown). The curve goes down, but is not “flattened”.
- Starting at the peak prevalence gives the epidemic a “coup de grace”, shortening its duration, albeit at the price of reducing the R/S ratio. The curve is narrower, but also not “flattened” (Fig 11).
- Starting at the peak incidence “flattens” the curve without broadening it and maximizes the number of deaths prevented during the current epidemic (unless behavioral adaptations reduce that effect), but reduces herd immunity and, thus, the chance of another epidemic coming sooner (Fig 12).
- Starting before the peak incidence “flattens the curve”, but also broadens it and causes a rebound, unless the intervention is continued for many more months (Fig 14).

It is herd immunity that stops the spread of an infectious disease, so, in general, one would want to let the epidemic initially run its natural course (or even accelerate it, as people have traditionally done with “measles parties”) to build immunity as fast as possible. If the aim were to reduce the duration of the epidemic and its impact on the economy (and also increase the time until the next epidemic can spread), one would wait until the prevalence of infectious people (I) reaches its peak (in the above model: day 83, red).

Without repeated broad testing, however, that peak prevalence of infections cannot be directly observed, but it is known to be followed about one week by peak number of diagnoses (new cases). This is too late to make a decision, but the SIR model shows that the peak in diagnoses is preceded by two weeks by the “turning point” in cases where the curve of the new cases changes from increasing ever faster to increasing ever more slowly (day 76). The turning point can be estimated from the observed cases in time to making a decision. (It is also about 50% of the peak number of new cases, which one might be able to predict.) Hence, peak prevalence (of infections) follows the turning point/half peak (in number of cases) by about a week. **The window of opportunity for starting an intervention is the week following the turning point in number of diagnoses (new cases) per day.**

**Responses by country in relation to the turning/decision point**

**Asia**

China initially (on 01-23) only closed means of transportation in Wuhan and other Hubei cities, causing 100,000s to leave the city before, in the afternoon, closing major highways. Family outdoor restrictions were issued on 02-01..05 (the week following the 01-31 turning point), but schools and non-essential companies were not closed until 02-20, **three weeks after the turning point**, when interventions are not effective anymore. A month later, on 03-22, the lockdown in the Italian region of Lombardy was tightened to ban sports and other physical activity, as well as the use of vending machines.

South Korea initially requested people in the city of Daegu, where the Shincheonji Church has its center, to “self-quarantine” from 02-23 (the peak of the first wave). Nationwide “social distancing” was not advised until 02-29 (a day before the peak of the main wave, or **three days after the turning point**). Hence, the social distancing prevented herd immunity from developing, as seen
by the low number of cases becoming infected (including some deaths) for now more than six weeks (Fig 12).

Europe

European governments, aiming to take advantage of being forewarned, ordered “social distancing” much earlier in the epidemic.

In Italy, the government ordered a nationwide close of restaurants, bars, and most stores more than a week earlier, on 03-11,  four days before the turning point where the intervention prolongs the epidemic and may cause a rebound (Fig 14).

Germany ordered schools to be closed and various other restrictions on or before 03-16/17, depending on the state and a national curfew on 03-22. Taken together, “social distancing” started effectively several days before the 03-22 turning point, so that the epidemic is unlikely to reach herd immunity within the near future. Hence the epidemic will not cease for several months, unless restrictions are removed, which may cause a rebound (Fig 14), a catch-22.

North America

In the North America, the virus arrived about a week after Europe. Responses varied not only by country, but in the U.S., like in Germany, also by state.

In NY, one of the epicenters of the U.S. epidemic, social distancing started about 03-15 with closure of schools being announced and 03-17 with restaurants being closed, and intensified on 03-22 with the closure of all non-essential businesses (similar actions were taken in New Jersey and Connecticut). Given the fast rise in number of cases in New York turning on 03-22 (Fig 9c), which was also observed in Detroit, Chicago, and Dallas, see, an intervention at 03-22 started at the turning point (Fig 12). Hence the virus will spread in the population much longer but a reduction of deaths is possible (albeit not guaranteed).

In California, a shutdown was ordered on 03-19 and many states and cities in the US followed suit, mandating schools, restaurants, and “non-essential” businesses. The turning point in the In the U.S. as a whole, however, cannot have been earlier than 03-27 (Fig 9). Hence, social distancing was ordered before the epidemic reached its turning point, and a “flattened”, but much broader with the virus staying in the population until the intervention is changed (or ignored) so that herd immunity can be reached.

Oceania

Other than travel restrictions (14 day quarantine for non-residents and travel restrictions between states, similar to what other countries have enacted) the AU government has not imposed any social distancing requirements and then financed a bailout to prevent a burndown of the market (as in the U.S.). Instead, it provided funds for delivering an AU$20.7B (AU$800 per person) support package (investment, jobs, health) opening fever clinics and funding home delivery of prescription medicines.
Discussion

Strengths and shortcomings

A major strength of this analysis of the epidemiological data is that it does not rely on epidemiological models with questionable assumptions. Instead, the results reflect raw incidences over time as reported by the ECDC, depicted by country or region of neighboring countries.

A shortcoming of such an entirely data-based approach is that it lacks the sophistication and potential additional insights that could come from fitting, e.g., differential equation models. The only modeling assumption made is that curves should be “smooth” (except when reporting artifacts are suspected), but even then, data were redistributed only to the directly neighboring day.

Still, the evidence is strong enough to draw qualitative conclusions about possible scenarios for the spread of SARS-CoV-2 in the near future. Also, the results suggest strategies to explore the variability of the SARS-CoV-2 virus strains and to select prevention strategies.

Evidence for (at least) two different strains of SARS-CoV-2

During the 2003 SARS epidemic the number of new cases peaked about three weeks after the initial increase of cases was noticed and then declined by 90% within a month. Table 1 shows the relevant timepoints for the 2020 COVID-19 epidemic. The SARS-C0V-2 data also suggest that it takes at least a month from the first case entering the country (typically followed by others) for the epidemic to be detected, about three weeks for the number of cases to peak and a month for the epidemic to “resolve”. This data is consistent with the results from the SIR model (see Epidemiological Models).

![Table 2: Epidemiological Timepoints by Country](image)
The 2003 SARS and the 2020 SARS-CoV-2 are not only similar with respect to genetics (79% homology),\(^{(Lu 2020)}\) immunology,\(^{(Ahmed 2020)}\) involvement of endocytosis (also with influenza and syncytial viruses),\(^{(Behzadi 2019)}\) seasonal variation (same season in the northern hemisphere also with influenza, syncytial, and metapneumo viruses),\(^{(Olofsson 2011)}\) evolution (origin in bats, 88% homology),\(^{(Benvenuto 2020; Malik 2020)}\) but also with respect to the duration between emergence and peak of cases as well as between this peak and resolution of the epidemic (Table 1). Based on these similarities, one could predict the COVID-19 epidemic to end before 04-15 in Europe and about two weeks later in the U.S.

The time and height of the peak incidence of cases in the different countries are consistent with the hypothesis that SARS-CoV-2 moved step-by-step westward from China, via other Asian countries to Middle East (Iran, Qatar, and Bahrain), Southern Europe (Italy, followed by its neighbors CH/F/ES/AT/SI), central and northern Europe, and, finally, the US.

Viruses improve their "survival" if they develop strategies to coexist with the (human) host.\(^{(Woolhouse 2007)}\) Multiple coronaviruses have been found to coexist in bat populations.\(^{(Ge 2016)}\) The emerging COVID-19 data is consistent with the hypothesis that (at least) two SARS-CoV-2 strains have developed. One strain, which traveled through South Korea, remained more infectious, while the other strain, which traveled through other Asian countries lost more of its infectiousness. The strain that passing through South Korea and then Iran and Italy (SKII strain) showed high lethality in Iran and Italy, but less lethality when it traveled to Italy's neighbors, either because of differences in health systems, because the strain mutated back, or because a strain arriving directly from Asia had the advantage of spreading first. Only sequencing samples from these countries can help to answer these questions.

**Fig 15: Hypothetical virus transmission pathways.** Connection width: number of contacts, box colors: infectious-ness, box borders: lethality, dotted connections/borders: unknown. The end date of a box indicates the date of peak incidence, if known (bold date).
Changes in infectivity and lethality between China and Europe

Mainland China is not reporting relevant numbers of new cases anymore and Hubei reports no new cases since 03-19. The number of new cases in South Korea also has declined to low levels since its peak around 02-30. Maritime Southeast Asia continues to show low levels of new cases only (<3/M/d), with the possible exception of Singapore and Malaysia/Brunei. The Philippines are slowly approaching a 3/M/d incidence.

The data are consistent with the same “SKII” strain traveling from China via South Korea and Iran to Italy. Iran was hit about a week after South Korea (around 03-07), with a similar peak incidence, but higher lethality (red bars in Fig 2). The data also suggests a second wave of infections in Iran, which may have peaked on 03-15. Italy was hit a week after the first wave in Iran (which peaked around 03-07/08, Fig 4a). Incidence in Italy reached a substantially higher incidence than reported in Iran. The peak incidence in Italy may have been reached on 03-22 (at about 100/M/d).

Without sufficiently detailed genetic data, it is not clear whether the high lethality in Italy is due to genetic variations in the virus or to Italy having the second oldest populations in the world (after Japan). A 03-20 report by the Istituto Superiore di Sanita, however, implicates that age and comorbidities played a role – among 3200 deaths, mostly in Lombardy and Emilia-Romana, median age was 80 years (IQR 73-85, only were 36 below the age of 50), 98.8% had at least one comorbidity (hypertension: 74%, diabetes: 34%, ischemic heart disease: 30%, atrial fibrillation, 22%, chronic renal failure: 20%, ...).

The epidemiological data does not support the hypothesis that SARS-CoV-2 spread from Munich in Germany to Italy. Instead, the virus may have spread from Italy to its neighboring countries, Switzerland, France, Spain, Austria, and Slovenia, within just a few days of arriving from Iran. The top incidence seems to be less than half of that in Italy and the lethality is lower, too. While Italy has many people 65 years and older (23%, second only to Japan), the relatively small differences or age distribution within Europe (e.g., Germany: 21%) are unlikely to account for much of this difference. A possible explanation (indicated in Fig 6) is that the less virulent strain(s) arriving from other parts of Asia may have had a head start in those countries, so that imported infections from Italy met subjects who had already developed (cross) resistance against both strains.

The parts of Europe not directly hit by the SKII strain leveled off at (about 70/M/d). Overall, incidence in Europe seems to have leveled off at about 80/M/d (and recent reporting irregularities seem to have less of an effect). There is strong evidence for incidence having leveled off and is already decreasing in several countries.
Predictions for COVID-19 in North America

From Table 1, SARS-CoV-2 has arrived in the U.S. almost a week after it arrived in Europe and incidence of COVID-10 peaked around 03-28. U.S. lethality (2.9%) is about a bit lower than Europe’s (3.1%), consistent with this later onset. Hence, there is no evidence for a high lethality strain dominating the epidemic in the U.S. If the U.S. incidence were to peak at about around 04-04, it could reach 120/M/d (40,000 cases per day). South-Korea’s, economy and political system are comparable to the US, and the proportion of people older than 65 years is similar (14% vs 16%) and, based on the shape of the curve in South Korea, where the epidemic completed its course, the cumulative incidence could reach 1,500/M or a total of 500,000 cases, and, at a lethality of about 3%, about 15,000 deaths. A total of 6,046 deaths have already been reported. A number of 15000 U.S. death over the course of the epidemic compares to an expected number of 16,000–78,000 influenza deaths per season from pneumonia and respiratory/circulatory complications alone, which also occur predominantly among people at 65 years of age and older. [Rolles 2018]. The precise number of people dying will, of course, depend on how easily they can access health care to be treated against severe complications (e.g., pneumonia).

A historical perspective

This is not the first, and likely not the last time, that well-intentioned public health policies are inconsistent with our understanding of how epidemics spread. For instance, during much of the HIV epidemic, there was widespread fear that HIV could establish itself in the population as a whole, even though the data (including data showing absence of transmission to the wives of hemophiliacs) [Wittkowski 1995a] and models [Wittkowski 1992; Seydel 1994] contradicted this fear. [Wittkowski 1995b; 1996] These results have been repeatedly confirmed. [Centers for Disease Control and Prevention 2019; Haddad 2019] In the case of heterosexual transmission of HIV one could argue that there was little risk associated with a the public health policy promoting condom use, but in the case of COVID-19 prevention, ignoring models and data may carry substantial risk.

During the AIDS epidemic, epidemiologists had the advantage that, in addition to the date of report, the date of diagnosis was available for analysis so that variations in reporting delays, such as mid-February in China, 03-20 in Germany, and 03-26 in Canada, could be accounted for. Unfortunately, the public COVID-19 data lacks that information.

Implications for prevention

In China and South Korea (and the first wave in Iran), incidence peaked after about 2 weeks and then declined. In Europe it took and in the U.S. it takes twice as long. The shorter duration of the epidemic in China and South Korea however, does not demonstrate the effectiveness of social distancing, because the social distancing started too late to be effective. Instead, the longer duration in Europe is consistent with premature interventions in Europe (aiming to “flatten the curve”) prolonging the epidemic (“broadening the curve”).

A major problem with respiratory diseases is that one cannot stop all chains of infections within families, friends, neighbors, …. Even after a couple of weeks of “lockdown” there will be a few infectious persons, and as long as there are enough susceptible people in the society, this is enough to re-start the epidemic until there are enough immune people in the society to create “herd immunity”. Hence, one would expect the cases to appear in waves (Fig 14, the period of the “lockdown” corresponds to March to May, 2020 in the U.S.). Such waves of cases have been seen in different countries and the longer than expected duration of the epidemic supports the hypothesis that the social distancing / lockdown interventions had some effect, albeit at a high cost for approx. 10% of deaths saved.
A peak incidence of 15/M/d and a total incidence of 155/M are similar to South Korea (14/189) and comparable to less densely populated European countries (Slovakia: 20/168, Finland: 16/203, Greece: 10/116). Hence, the epidemic of Oceania, together with those of China, South Korea, and Iran are consistent with the model results suggesting that a natural COVID-19 epidemic peaks two weeks after the first cases are seen, and then declines with financial and medical assistance from the government to prevent deaths to reduce the burden on the health system and damage to the economy.

The data from Scandinavia (Fig 8) also does not support the effectiveness of social distancing. The epidemic curves of Sweden and the surrounding Norway, Denmark, and Finland are virtually identical, although the latter countries imposed “social distancing” rules, while Sweden did not.

This analysis of the publicly available data suggests that at the time Italy imposed quarantine on the Lombardy and adjacent regions on 03-08, the SKII virus strain had already reached the adjacent countries (Switzerland, France, Spain, Austria, Slovenia). Even though the lockdown started early (03-08), which may have caused a rebound when compliance declines.

Some containment strategies could even be counterproductive in other ways. For instance, the simple model used in Fig 14 does not account for age-stratification. In diseases such as COVID-19, children develop mostly mild forms, elderly people have a high risk of dying. Hence, containment of high-risk groups, like elderly people in nursing homes (see the Washington State example) is highly effective in protecting them from becoming infected and in reducing the pool of children and young adults that would have to be infected to reach herd immunity. A substantial increase in the duration of the epidemic by preventing immunity to develop among the young, however, might make effective containment of the elderly more difficult and, thus, increase the number of deaths among the elderly.

**Conclusions**

Until a vaccine will become available, the only pharmacological strategy to reduce the number of deaths is to reduce the damage the infection (and immune system) does, e.g., by reducing the initial viral load, and making sure that people get treated at the earliest signs of pneumonia.

Aside from separating susceptible populations (elderly and high-risk subjects, e.g., in nursing homes) from the epidemic, which is effective as long as virus is circulating, public health intervention aiming to contain a respiratory disease need to start within a narrow window of opportunity starting at or a week after the curve of the new cases changes from increasing faster to increasing more slowly. Only if stopping the epidemic from generating a sufficient number of immune people is avoided can containment efforts stop after about a month or two (depending on late or early start, respectively), when the ratio of infectious vs immune people is low enough for preventing the disease from rebounding. When the window of opportunity has been missed, containment has only limited impact on the course of the epidemic, but high impact on economy and society.

As epidemics in maritime Asia and the southern hemisphere remain a possibility and some develop within the next weeks, the use of data and scientifically sound models may help to time interventions to optimize their effect.

To determine that time point, case data collected and reported needs to contain not only the date of report, but also the date of “diagnosis” and whether the patient had clinical symptoms or was merely tested positive and whether the patient was positive for circulating virus RNA/DNA (currently infectious) or antibodies (already immune).
References


