Evolution of disease transmission rate during the course of SARS-COV-2: patterns and determinants

Jie Zhu[a], Blanca Gallego[a,*]

[a] *Centre for Big Data Research in Health, UNSW Sydney, Sydney, 2052, NSW, Australia*

Abstract

To date, many studies have argued the potential impact of public health interventions on flattening the epidemic curve of SARS-CoV-2. Most of them have focused on simulating the impact of interventions in a region of interest by manipulating contact patterns and key transmission parameters to reflect different scenarios. Our study looks into the evolution of the daily effective reproduction number during the epidemic via a stochastic transmission model. We found this measure (although model-dependent) provides an early signal of the efficacy of containment measures. This epidemiological parameter when updated in real-time can also provide better predictions of future outbreaks. Our results found a substantial variation in the effect of public health interventions on the dynamic of SARS-CoV-2 transmission over time and across countries, that could not be explained solely by the timing and number of the adopted interventions. This suggests that further knowledge about the idiosyncrasy of their implementation and effectiveness is required. Although sustained containment measures have successfully lowered growth in disease transmission, more than half of the 101 studied countries failed to maintain the effective reproduction number close to or below 1. This resulted in continued growth in reported cases. Finally, we were able to predict with reasonable accuracy which countries would experience outbreaks in the next 30 days.

1 Introduction

Mathematical and computational models of disease outbreaks are used to generate knowledge about the biological, behavioral and environmental processes of disease transmission, as well as to forecast disease progression. Public health responders rely on insights provided by these models to guide disease control strategies [1]. As a recent example, in mid March this year, the British government changed its SARS-CoV-2 response policy following a brief on simulation results from Ferguson et al. [2], which indicated an unacceptable forecasted number of deaths in the absence of more stringent control measures.

To date, different models have been applied to model the spatial and temporal dynamics of SARS-CoV-2 (see [3] for a review). They range from simple deterministic population-based models [4–6], that assume uniform mixing, to complex agent-based models [7,8] in which individuals defined by different attributes related to their susceptibility, infectiousness and social interactions transmit the pathogen to each other given rise to heterogeneous transmission patterns.

Irrespective of their complexity, the accuracy of these models is constrained by the validity of the epidemiological parameters that underpin them. For example, the model parameters controlling the risk of infection together with the social contact between infectious and susceptible individuals determine the transmission rate, which in turns influences the peak and duration of the epidemics. By manipulating these parameters, modelers can represent the

*Corresponding author.
Email addresses: elliott.zhu@unsw.edu.au (Jie Zhu), b.gallego@unsw.edu.au (Blanca Gallego)

impact of public health measures such as social distancing (lowering the contact rate) or wearing protective masks (lowering the risk of infection).

This study evaluates the effectiveness of public health interventions by estimating the effective reproduction number (the expected number of secondary infections resulting from an infectious individual) on a daily basis. Future outbreaks are predicted to test model accuracy and to illustrate how the currency of epidemiological parameters can influence the expected infection counts.

2 Method

2.1 Data Sources

By 15th May 2020, there were 188 countries with recorded incidence of SARS-COV-2. 101 of them had at least one death and had implemented one or more public health interventions[9] to contain the spread of disease. We modeled the transmission dynamics of SARS-COV-2 between 22nd of Jan to 15th of May 2020 via fitting a stochastic SEIRD model[10] (illustrated in Figure 1) to three available time series: 1) Daily number of incident cases, 2) Daily number of deaths, and 3) Daily number of recoveries, as recorded, for each country, by the Johns Hopkins Coronavirus Resource Center[11].

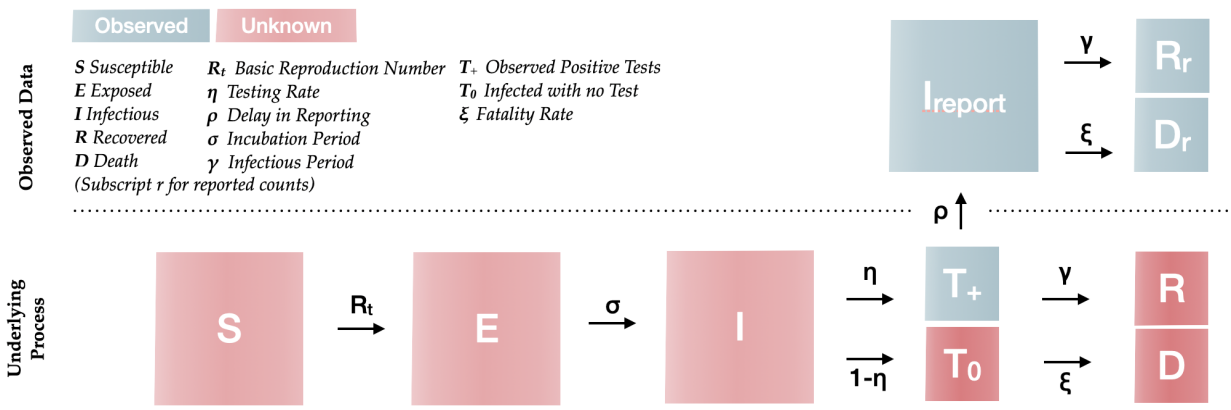


Figure 1. Simulation model structure. The population is divided into the following four classes: susceptible, exposed (and not yet symptomatic), infectious (and symptomatic), removed (i.e., isolated, recovered, or otherwise non-infectious) and dead.

To monitor public health interventions, we captured government stringent policies using the Oxford Covid-19 government response tracker[9], which contains 17 policies organized into three groups: containment and closure policies, economic policies and health system policies. We explored 12 of these policies, namely those that have a more immediate and direct impact on individual's lifestyle and well-being. Social distancing policies (including travel policies) were further classified into two levels depending on whether the government took a recommended or required stance; and international travel controls were divided into four levels: 1) screening arrivals, 2) quarantine arrivals from some or all regions, 3) ban arrivals from some regions, and 4) ban on all regions or total border closure.

2.2 Epidemic simulation model and cluster analysis

We built an extended version of a SEIRD model that included transitions between reporting states and disease states in order to account for delays in reporting (Figure 2). The model also incorporated uncertainty in reported counts, by explicitly modeling a Poisson process of daily reported infectious, recovered and dead counts, as well as infectious

individuals not yet tested. The incubation period was assumed to be Erlang distributed with mean 5·2 days[12] (SD 3·7) and delay from onset to isolation was assumed to be Erlang distributed with mean 2·9 days (2·1).[13,14].

For each country k, the daily transmission rate ($\beta_{t,k}$) and fatality rate ($\xi_{t,k}$) were modeled as a geometric random walk process, and sequential Monte Carlo simulation was used to infer the daily reproduction number, $R_{t,k}$ (defined as the transmission rate over the assumed incubation period $\beta_{t,k}/\gamma$), as well as the daily fatality rate. The remaining unknown parameters were estimated via grid search. In each country, we assumed the outbreak started with the first infectious case and the entire population was initially susceptible.

To predict reported cases beyond the fitting period, we fixed the model parameters defined above to those fitted on 15th of May 2020 and run the epidemic model for each country for the next 30-day period ending at 15th of June 2020.

Hierarchical clustering was conducted to summarize the patterns of smoothed time-series of $R_{t,k}$ and predictions of reported cases using Savitzky-Golay finite impulse response[15] (FIR) filter of polynomial 4 and frame length 20. Further details on the methodology can be found in the supplementary material.

3 Results

3.1 Patterns of daily reproduction number and public health interventions

Across the 101 selected countries, the estimated R_t peaked on average 29·6 (25·36,33·81) days before the crest in infections. The average peak duration (the time it takes R_t to reduce to 50% of its maximum) was 5·4 (4·84,6·06) days. A hierarchical clustering algorithm identified 5 distinct patterns of R_t (displayed in Figure 2). Descriptive statistics on these patterns are summarized in Table 1, including three metrics over 8 social distancing policies:

- Policy timing defined as days taken from onset to intervention or from peak of R_t to intervention,
- Policy volume, which refers to the number of interventions applied, and
- Policy gap defined as the average number of days between any two policies.

Cluster 1 contains 4 countries: Italy, Spain, Belgium and Sweden. They share the most severe mortality burden and have been featured in a recent study[16]. They are also characterized by high infection rates (4·1 cases per thousand population), high testing rates (35·7 tests per thousand population), and an ageing population (20·1% of the population of these four countries aged 65 or above[17]). This group had the least number of adopted interventions (5·8) and was among the latest in our study cohort to apply the first social distancing policy (34·5 days). The combined average R_t peak, 30·3 (27·22,33·28), and duration, 6·0 (4·48,7·56) days was larger than in other clusters, which resulted in high number of infections.

Cluster 2 contains 12 countries featuring two R_t peaks, the first peak took place during the first week from onset and averaged 4·4 (3·54, 5·35); and the second took place after approximately one month and had an average value of 4·6 (3·95, 5·20). This cluster includes countries such as Australia and the US which started international travel control and contact tracing shortly after the onset, but delayed subsequent social distancing measures by a few weeks.

Cluster 3 includes 20 countries and has an average peak in R_t of 8·6 (7·09, 10·03) on day 7 from onset. This group adopted most interventions (7·3) within the shortest gap between any two interventions (7·1 days). The R_t peak was lower (average of 21·8 (12·59,31·01)) but of similar duration (6·3 days) than that of Cluster 1. Twelve of these countries applied all 8 social distancing policies on average 18·2 (15·22, 20·14) days from the onset which successfully lower the peak in transmission.

Countries in Cluster 4, which included China, Iraq and Argentina, had very high (averaged at 57·3 (44·73, 66·89)) but very short (3·5 (2·46,4·48) days) R_t peaks. These countries adopted their first social intervention 7·8 days after the outbreak which was the most prompt adoption amongst all clusters.

Cluster 5 contains the remaining 49 countries, including 13 of the 28 examined high income countries. The shape of their daily R_t was similar to that of Cluster 4 but with a much lower peak (averaged at 13·1 (10·08, 16·11)) and lower volatility. This group had the lowest infection and fatality rates.

The adoption of social distancing policies took place on average 22·3 (19·39,25·19) days from the first reported case and 5·3 (1·69, 8·84) days before the estimated R_t peak. 76·9% of the examined interventions were applied before the end of peak duration. Adoption timing was much slower in clusters 1 and 2. By 15th of May, the selected countries had adopted 7·0 (6·81,7·29) social distancing measures and around 60% of them had both testing and contact tracing policies in place. Only 32% of the countries had income support and 51% had debt or contract relief policies.

Table 1. Estimated daily reproduction number, observed intervention timing, gap and volume, and reported tests, death and infection across clusters as of 15th May 2020.

Metric (% of adoption)	Overall (101)	Cluster 1 (4)	%	Cluster 2 (12)	%	Cluster 3 (20)	%	Cluster 4 (14)	%	Cluster 5 (51)	%
Onset to R peak (days)	26·3 (22·38,30·25)	30·3 (27·22,33·28)		45·7 (37·08,54·25)		21·8 (12·59,31·01)		18·1 (12·03,24·25)		25·5 (19·7,31·24)	
R peak duration (days)	5·4 (4·84,6·06)	6·0 (4·48,7·56)		9·9 (7·76,12·02)		6·6 (5·22,7·92)		3·7 (2·58,4·79)		4·4 (3·84,4·96)	
R standard deviation	4·1 (3·15,5·08)	2·8 (2·14,3·53)		1·4 (1·26,1·55)		2·5 (2·01,2·97)		12·7 (8·64,16·79)		3·1 (2·35,3·91)	
R mean	2·9 (2·63,3·17)	2·9 (2·6,3·14)		2·3 (2·1,2·43)		2·7 (2·4,2·91)		4·8 (3·44,6·18)		2·6 (2·38,2·86)	
Incidence Peak to R peak (days)	29·6 (25·36,33·81)	14·5 (4·3,24·7)		10·3 (1·75,18·92)		37·1 (31·02,43·08)		37·7 (26·66,48·77)		30·1 (23·72,36·55)	
School closing (99%)	17·1 (14·13,20·03)	33·0 (23·13,42·87)	75%	45·1 (40·38,49·79)	100%	15·1 (12·53,17·67)	100%	8·6 (6·79,10·5)	100%	12·6 (9·08,16·22)	100%
	-9·2 (-12·63,-5·71)	3·7 (-2·87,10·2)	75%	-0·6 (-10·2,9·04)	100%	-6·7 (-15·66,2·26)	100%	-9·5 (-16·06,-2·94)	100%	-12·8 (-17·64,-8·01)	100%
Workplace closing (93%)	23·7 (19·93,27·39)	34·3 (22·13,46·54)	75%	55·2 (46·54,63·79)	100%	18·9 (16·06,21·64)	100%	15·4 (6·84,23·93)	93%	19·2 (14·75,23·59)	90%
	-1·9 (-6·01,2·13)	5·0 (-4·26,14·26)	75%	9·5 (-3·7,22·7)	100%	-3·0 (-11·44,5·54)	100%	-0·9 (-11·94,10·1)	93%	-5·2 (-10·88,0·45)	90%
Cancel public events (96%)	17·5 (14·63,20·4)	33·3 (23·19,43·48)	75%	46·8 (44·38,49·12)	100%	14·8 (11·98,17·6)	95%	8·6 (7·12,10·16)	100%	13·0 (9·92,16·04)	96%
	-8·3 (-11·76,-4·75)	4·0 (-2·88,10·88)	75%	1·1 (-7·78,9·95)	100%	-7·3 (-16·65,2·13)	95%	-9·5 (-16·07,-2·93)	100%	-11·3 (-16·29,-6·36)	96%
Restrictions on gatherings (93%)	21·8 (18·47,25·19)	41·3 (27·2,55·3)	100%	48·9 (47·27,50·55)	92%	16·5 (14,19)	90%	16·8 (6·88,26·62)	85%	17·4 (13·37,21·36)	96%
	-4·1 (-8·38,0·19)	11·0 (-1·58,23·58)	100%	2·9 (-7·11,12·93)	92%	-5·5 (-15·85,4·85)	90%	0·0 (-13·03,13·03)	85%	-7·4 (-13·36,-1·41)	96%
Close public transport (52%)	23·0 (18·5,27·57)	72·0	25%	50·3 (44·1,56·57)	25%	23·2 (18·74,27·72)	65%	14·2 (11·96,16·48)	64%	21·0 (14·28,27·79)	52%
	-3·3 (-9·5,2·97)	46·0	25%	-5·7 (-32·21,20·88)	25%	-2·0 (-16·13,12·13)	65%	-0·3 (-4·65,3·99)	64%	-6·4 (-15·62,2·8)	52%
Stay at home requirements (84%)	25·8 (22·32,29·26)	36·0 (23·25,48·75)	75%	51·2 (45·57,56·88)	75%	26·9 (20·36,33·52)	90%	19·7 (13·63,25·82)	78%	20·9 (16·57,25·29)	86%
	0·3 (-4·24,4·88)	6·7 (-2·82,16·16)	75%	5·3 (-7·89,18·55)	75%	4·4 (-8,16·89)	90%	0·2 (-10·24,10·6)	78%	-2·8 (-8·95,3·36)	86%
Restrictions on internal movement (88%)	25·5 (22,29·08)	30·0 (12·36,47·64)	50%	51·1 (47·47,54·7)	100%	21·8 (17·28,26·25)	85%	21·8 (14·7,28·84)	93%	21·0 (16·15,25·93)	88%
	-1·0 (-4·91,3)	1·0 (-10·76,12·76)	50%	5·4 (-3·67,14·51)	100%	-1·2 (-12·18,9·82)	85%	3·7 (-4·76,12·15)	93%	-4·0 (-9·62,1·62)	88%
International travel controls (99%)	13·6 (11·2,15·9)	34·5 (16·07,52·93)	100%	20·5 (9·45,31·64)	92%	14·1 (10·42,17·78)	100%	11·2 (7·08,15·34)	100%	10·8 (8·16,13·49)	100%
	-12·7 (-17·11,-8·37)	4·3 (-11·21,19·71)	100%	-26·6 (-40·12,-13·15)	92%	-7·7 (-18·09,2·69)	100%	-6·9 (-15·33,1·47)	100%	-14·6 (-20·75,-8·54)	100%
Testing policy (64%)	33·8 (27·48,40·18)	38·7 (13·84,63·49)	75%	61·9 (44·41,79·39)	83%	32·7 (23·24,42·13)	80%	25·4 (13·79,36·96)	57%	26·4 (16·83,35·88)	55%
	6·8 (0·18,13·42)	9·3 (-15·08,33·74)	75%	14·2 (-6·71,35·11)	83%	15·1 (4·15,25·97)	80%	6·1 (-8·67,20·92)	57%	-0·6 (-11·32,10·04)	55%
Contact tracing (60%)	19·5 (14·02,25·03)	10·8 (3·4,18·1)	100%	26·0 (3·74,48·26)	75%	15·9 (3·81,27·99)	50%	17·9 (6·75,29·05)	71%	20·6 (12·7,28·44)	55%
	-7·9 (-15·59,-0·15)	-19·5 (-26·31,-12·69)	100%	-18·1 (-43·75,7·53)	75%	-9·2 (-31·36,12·96)	50%	0·6 (-11·63,12·83)	71%	-5·5 (-17·15,6·22)	55%
Income support (32%)	34·9 (29·53,40·29)	39·0 (30·46,47·54)	75%	51·8 (46·52,57·15)	50%	23·7 (17·78,29·65)	35%	24·2 (12·08,36·32)	36%	36·5 (26·76,46·33)	21%
	9·7 (4·14,15·3)	7·3 (-2·29,16·96)	75%	9·0 (3·01,14·99)	50%	8·4 (0·58,16·28)	35%	4·4 (-4·99,13·79)	36%	14·0 (-0·6,28·6)	21%
Debt/contract relief (51%)	34·7 (28·75,40·71)	45·0 (17·56,72·44)	50%	58·8 (48·62,68·94)	75%	33·7 (20·94,46·39)	45%	10·5 (7·27,13·73)	43%	31·6 (23·87,39·29)	51%
	5·6 (-0·34,11·49)	14·0 (-15·4,43·4)	50%	14·7 (7·61,21·72)	75%	14·4 (-3·16,32·05)	45%	-12·8 (-27·6,1·93)	43%	3·0 (-5·26,11·18)	51%
Policy gap (days)	8·6 (7·34,9·84)	8·7 (2·95,14·36)		13·1 (9·13,17·17)		7·1 (5·06,9·05)		9·9 (7·31,12·4)		7·8 (5·87,9·67)	
Policy volume	7·0 (6·81,7·29)	5·8 (3·17,8·33)		6·8 (6·3,7·36)		7·3 (6·68,7·82)		7·1 (6·65,7·64)		7·1 (6·76,7·43)	
Policy timing (days from R peak)	-5.3 (-8.84,-1.69)	8.4 (3.21,13.66)		0.1 (-9.12,9.4)		-3.2 (-12.37,5.95)		-3.7 (-10.85,3.39)		-8.8 (-14.05,-3.63)	
Policy timing (days from onset)	22·3 (19·39,25·21)	38·6 (32·55,44·62)		49·8 (47·41,52·21)		19·0 (16·24,21·71)		15·0 (12·5,17·45)		17·9 (14·25,21·47)	
Tests/'000	14·4 (10·12,18·75)	35·7 (22·6,48·89)		21·5 (12·55,30·37)		19·1 (8·17,30·02)		10·2 (2·99,17·36)		10·5 (4·01,16·91)	
Reported deaths/'000	0·08 (0·042,0·115)	0·56 (0·388,0·736)		0·13 (0·027,0·228)		0·08 (0·018,0·145)		0·13 (-0·041,0·302)		0·01 (0·008,0·019)	
Reported infections/'000	1·4 (0·92,1·94)	4·1 (3·1,5·03)		1·5 (0·7,2·39)		1·7 (0·59,2·71)		2·1 (-0·51,4·81)		0·9 (0·41,1·41)	
Death/Infection	0·04 (0·037,0·051)	0·14 (0·118,0·157)		0·06 (0·028,0·083)		0·04 (0·033,0·05)		0·05 (0·03,0·063)		0·03 (0·026,0·042)	

Does not include countries implementing interventions not listed in Oxford Covid-19 government response tracker as of 15th May 2020.

The first section (as indicated by the black horizontal bars) records the patterns of estimated daily reproduction number. R peak is defined as the day of observing the peak daily reproduction number. Incidence peak is the day of observing the maximum daily reported cases.

The second section records the 8 social distancing measures and the third section records the economic and health measures. Among the two rows within each measure, the first row records the average days from onset to policy implementation and the second row records the average days from the peak day of estimated daily reproduction number to the day of policy implementation with 95% confidence interval. The % columns record the percentage of countries implemented the corresponding intervention within the strata.

The fourth section records the pattern of interventions. Policy Timing (from onset) records the days from onset to the adoption of the social distancing policies averaged over countries and 8 social distancing policies overall and for each cluster. Policy Timing (from R peak) records a similar interval but replaces the onset date with the day of R peak. Policy gap are recorded in days and are defined as the average number of days between any two policies. Policy volume is recorded in absolute number.

The last section records the observed tests, reported death, reported infections per thousand population and reported death per infection.

3.2 The impact of public health interventions

Public health interventions had substantial but varied impact on the estimated daily reproduction number across countries over time.

Policy timing appeared to have an impact on the duration of the peak of transmission rate. For example, cluster 4 has the highest average R_t but its peak duration is the shortest among the five clusters, only 3·7 (2·58,4·79) days. Countries in this cluster responded faster to the outbreak adopting social distancing measures on average 15·0 (12·5,17·45) days from onset. In contrast, the longest transmission peak was found in cluster 2 at 9·9 (7·76,12·02) days. Countries in cluster 2 were on average 34·8 days slower at implementing their social distancing policies compared to countries in cluster 4. The correlation between the number of policies adopted before the peak in transmission rate and the duration of this peak across clusters was -0.26.

Policy gap was also found to be correlated with the duration of the peak in R_t, with a correlation value across clusters of 0.24. We observed that clusters 1 and 2 adopt similar number of policies at similar times from onset. However, the average policy gap in cluster 2 is 13·1 (9·13,17·17) days, compared to 8·7 (2·95,14·36) days in cluster 1. This correlates with a 3·9 days longer peak in transmission rate for countries in cluster 2.

There is not much difference in policy volume across clusters, even though countries in clusters 1 and 2 applied less polices than countries in the other three clusters. The combined effect of less social distancing measures, and longer policy timing and gap may have contributed to larger outbreaks in those clusters. By 15th of May 2020, we saw the mean of R_t had reduced to 1·3 (0·94, 1·74) from its peak of 20·5 (17·79, 23·20).

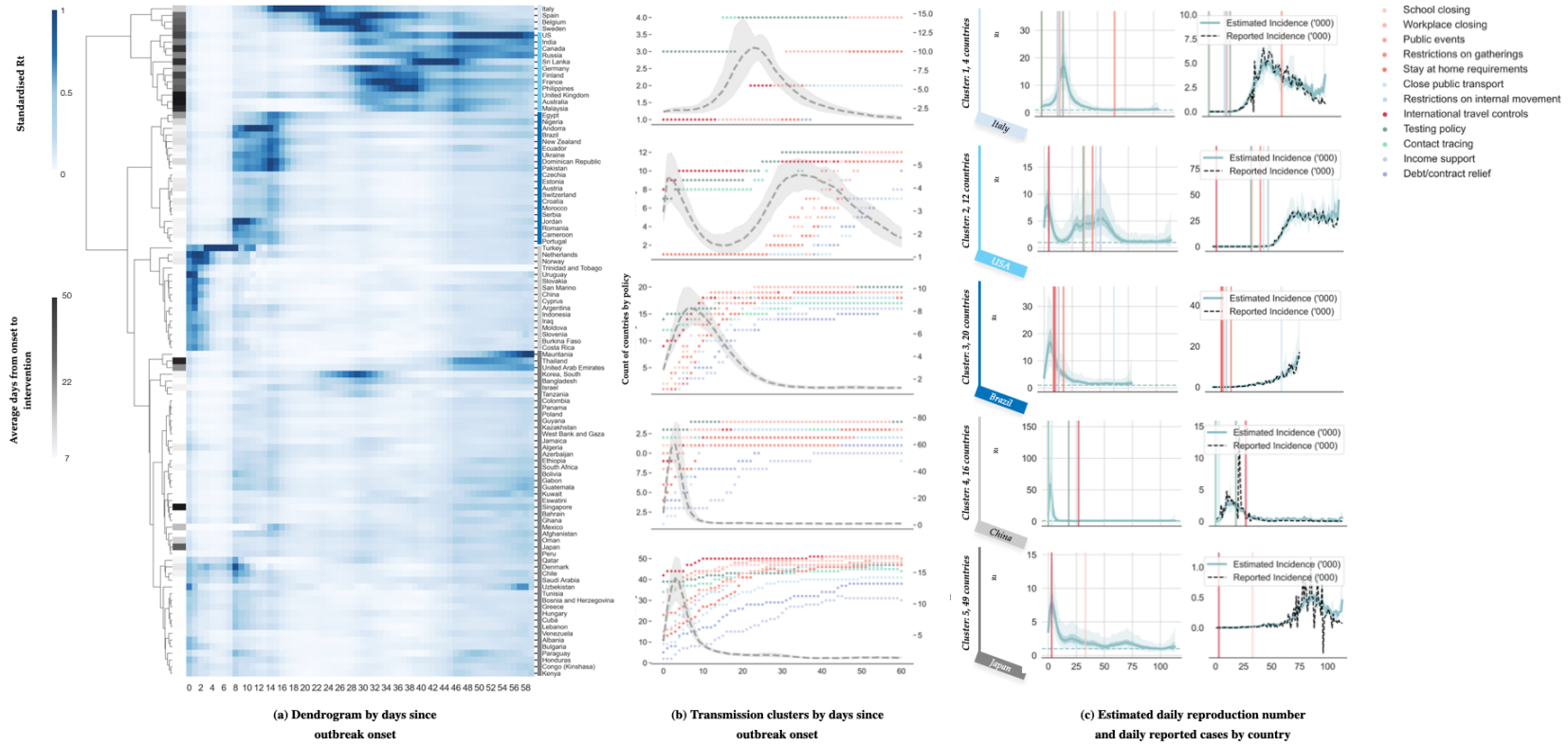
3.3 Forecasting future outbreaks

To examine the potential for future outbreaks we simulated reported counts in the next 30-days (up to 15th of June 2020) using the fitted parameters at the cutoff date (see Figure 4). Average prediction bias is 19·9% (16·48%, 23·36%), with 64 countries achieving lower than 20% bias. Prediction accuracy declines with forecasting time, and the bias on the last seven days of the prediction window (9th to 15th of June 2020) is 23·8% (19·30%, 28·30%). This decline in prediction accuracy is related to the decline in the currency of key transmission parameters. For instance, we found that bias over the period from 15th of May to 15th of June increases to 25·9% (21·29%, 30·61%) and 28·2% (23·18%, 33·17%) when using parameters from 5 and 10 days before the cutoff. The number of predicted new cases is very sensitive to R_t, with an approximate 30% increase in R_t leading to a 350% increase in predicted reported counts.

Hierarchical clustering on the predicted counts identified four clusters of countries (see Figure 3). Cluster 1 contains 24 countries and is characterized by a small increment in the predicted cumulative active cases of 8·0% (5·64%, 10·29%). Its leading R_t defined as the average of median R_t during the week before the cutoff date is 1·2 (0.98 1·34). Cluster 2 has 38 countries with a leading R_t of 1·3 (1·15, 1·43) and an average increment in reported cases of 47·3% (39·87%, 54·64%). Cluster 3 has 5 countries with a leading R_t of 1·4 (1·23, 1·67) and an average growth in predicted counts at 138·2% (92·29%, 184·12%). Lastly, cluster 4 captures the 34 countries with the highest number of predicted future cases with a 169·9%(138·05%, 201·80%) increment, and its leading R_t is 1·7 (1·44, 1·95). By the end of the prediction window, countries such as Mauritania, South Africa, Brazil and Chile are amongst those in this cluster with higher predicted counts.

Amongst the 101 studied countries, 53 had a major outbreak (case growth rate is greater than 50%) between the cutoff date of 15th of May at the end of the prediction window in 15th of June. Our model is able to identify 44 of them (a true positive rate of 83·0%)[1] . We computed the area under the receiver operating characteristic curve (AUROC) as 83·4% by varying the growth rate threshold from 0% to 100%. There is no meaningful pattern between adopted policies before the cutoff date and future outbreaks.

[1]The failed predictions are Ecuador, Ethiopia, Indonesia, Jordan, Kenya, Nigeria, Paraguay, Sri Lanka, and Venezuela



(a) Dendrogram by days since outbreak onset

(b) Transmission clusters by days since outbreak onset

(c) Estimated daily reproduction number and daily reported cases by country

Figure 2. Results of fitted daily reproduction number and public health interventions
a) Dendrogram of the clustered time-series of estimated daily reproduction number by country and days since outbreak and countries. Cluster 1, 2, 3, 4, and 5 are colored as pale blue, light blue, blue, light gray and gray respectively.
b) The dashed gray lines represent the median, light gray shading represents the 5th to 95th quantile of the daily reproduction number estimate of each cluster by days since outbreak onset. Each colored dot indicates the count of countries adopted corresponding policy.
c) The colored lines mark the initiation of the corresponding public health intervention. The solid blue lines represent median, light blue shading represents 50% confidence intervals of the model estimate, and dark blue shading represents 95% confidence intervals of the model estimate. The left figure in each cluster is the estimated daily reproduction number over time. The dashed horizontal line represents a daily reproduction number of one. The right figure in each cluster is the estimated reported cases versus the actual reported cases (black dashed line) by date of onset.

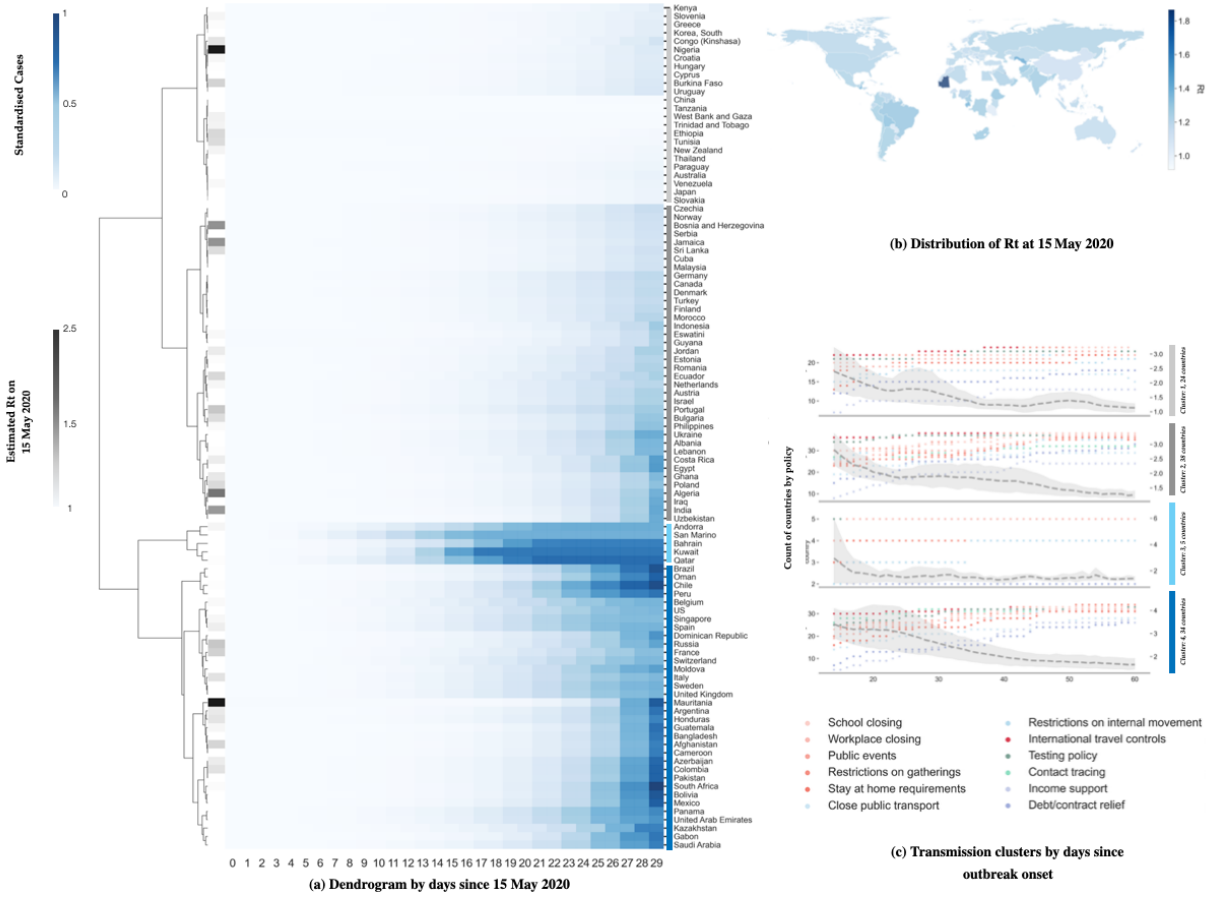


Figure 3. Results of predicted number of daily active infections and public health interventions. a) Dendrogram of the clustered time-series of predicted cumulative active cases by country and days since 15th of May 2020. Cluster 1, 2, 3 and 4 are colored in light gray, gray, light blue and blue, respectively. b) Estimated daily reproduction number on by country. c) Time series of public health interventions and evolution of daily reproduction number by clusters. Each colored dot indicate the count of countries with corresponding policy in place. The dashed gray lines represent median, light gray shading represents 95% confidence intervals of the daily reproduction number estimate.

4 Discussion

This study combined a stochastic epidemic simulation model with hierarchical clustering analysis to unveil the transmission rate of SARS-CoV-2 over time and across the globe as well as the observed impact of public health interventions.

To date, various algorithms have been applied to modeling the spatial and temporal dynamics of SARS-CoV-2, including mathematical simulation-based models[4,5,7,8] and statistical models[6]. For example, a Wuhan study[4] used a deterministic, population-based compartmental model. Baseline model parameters were fitted on observed local and exported cases. The impact of containment measures was explored by assuming various contact patterns given different physical distancing scenarios. The simulation suggested sustaining intense social distancing until the end of April. Similarly, an Italian study[5] used a more complex deterministic model to analyze the effect of social distancing. Initial parameters were inferred from fitting the model outputs to observed counts, deaths and recoveries while preserving a priori information on their relative magnitude. During the course of the simulation, the effect of social distancing

was modeled by changing the value of key model parameters. They found predictions were extremely sensitive to key parameters related to disease transmission, which may be a problem for the prediction of the effect of interventions. In this case, for instance, the basic reproduction number from 29th to 6th of April (when the national lockdown was fully operational) was assumed to be 0·99, compared to our estimated effective reproduction number, which varied from R_t=1·97(1·44, 2·50) to 1·07 (0·76, 1·38) during that period. The difference between these numbers may be attributed to different model structures as well as our consideration for the lag in reporting.

The assumption of population-wide homogeneous parameters limits the ability of these deterministic models to assess the spread of disease and its decline in relation to control measures[18,19]. During the current epidemic, a study in Singapore[7] used agent-based influenza epidemic simulations to recreate a synthetic but realistic representation of the Singaporean population. The effect of distancing measures was then assessed by assuming different transmission rates under given social distancing policies. They recommended implementation of the quarantine of infected individuals, distancing in the workplace, and school closures immediately after having confirmed cases once international travel control has been imposed.

Further, Report 9 by Imperial College London[8] assigned individuals to household, school, workplace and wider community and simulated scenarios by manipulating contact rates within and between groups. They found that a combination of case isolation, home quarantine and social distancing of aged population was the most effective scenario. This agrees with our findings that concentrated implementation of multiple interventions is most effective to contain transmission.

Rather than assessing the intervention impact via simulation, statistical models aim to represent the empirical association between public health interventions and transmission rates and/or counts. For example, Report 13 by Imperial College London[6] used a semi-mechanistic Bayesian hierarchical model to infer the impact of social distancing and travel lock downs on the daily reproduction number in 11 European countries[2]. The daily reproduction number was modeled as a function of the baseline reproduction number before any intervention together with multiplicative relative percent reductions in R_t from interventions. Model parameters were fitted to observed deaths in these countries as a function of the number of infections. This model had heavy assumptions on prior distributions of model parameters and assumed consistent intervention impact across countries to leverage more data for fitting. By 28th of March they found R_t was reduced to around 1 across the 11 countries. Our study found similar reduction in R_t, which was sustained after one and half months at 15th of May, when the average of median R_t was 1·17 (0.91, 1.53) across these countries.

Our results highlight the importance of using real-time estimates of R_t in these types of studies, since it is difficult to set epidemiological parameters under uncertain impact from public health interventions. This is particularly important given that prediction of future counts is very sensitive to simulation parameters. Here we provide a way to update the value of the daily efficient reproduction number in real time in order to facilitate the empirical evaluation of the impact of interventions as they happen as well as more informed simulations of future counts. We analyzed the period from 29th of Jan to 15th of May 2020, which covers a longer period than the aforementioned studies. By 15th of May 2020, R_t in the studied countries had reduced from an average peak of 20·5 (17·79, 23·20) to 1·3 (0·94, 1·74). There were substantial variations in transmission over time and per country associated with differences in the adoption of public health interventions combined with demographics and socio-economic factors.

A hierarchical clustering analysis systematically characterized the heterogeneity across countries and allowed us to draw inference on the relationship between the patterns of intervention and transmission. For example, there were two transmission peaks in countries such as Australia and the US who started international travel control shortly after the onset, but delayed subsequent social distancing measures by a few weeks. The clustering analysis also questioned the effect of testing and contact tracing in the presence of community transmission and in the absence of social distancing. Italy, Spain, Belgium and Sweden are among the top 8 countries with the highest mortality burden[16] and they are also in the top 10 countries which carried out most tests per thousand population. Our results suggest that the timing and concentration of social distancing interventions is an important consideration to ensure testing efforts have an impact.

[2] 11 countries in the original study are Austria, Belgium, Switzerland, Germany, Denmark, Spain, France,United Kingdom, Italy, Norway and Sweden.

Using the parameters estimated for the 15th of May we were able to predict future counts until 15th of June 2020 with reasonable accuracy. In particular, we predicted the current worrisome increase in counts found in countries like South Africa, Chile and Brazil. The prediction of future infections was very sensitive to changes in R_t, and the use of more current estimates provided more accurate counts.

On average, all the examined countries applied 7·0 (6·81,7·29) out of 8 social distancing policies prior and during the prediction period. However, this fact did not seem to slow future transmission in all countries, which leads us to question the quality of policy enforcement in some countries.

4.1 Limitations

There are several limitations to our analysis. The estimated daily reproduction number is specific to our extended, population-level SEIRD model. This model does not represent the heterogeneity of transmission within a country, and therefore, may be less relevant for countries such as the US and China, which have varied adoption of interventions across regions. Researchers using other models need to estimate their own transmission parameters in a similar manner.

Our stochastic model does not fit well during the initial week of the epidemic outbreak, when only a small number of data points were available across countries. Over time, the availability and reliability of data improved, which was reflected in an increased model prediction accuracy, where the absolute bias of fitted daily incidents is below 15% (please refer to the supplementary material for more details). Nevertheless, we acknowledge that varied data quality across countries may influence the accuracy of our analyses. In particular, WHO's daily Situation Reports shifted its reporting cutoff time on 18 March 2020, which compromised the comparability of its earlier figures. There woas not much difference bewteen the other two sources of disease counts[20] except that Johns Hopkins also included estimates of presumptive positive cases that have been confirmed by state or local labs, but not by national labs. Two well-known problems of SARS-CoV-2 data are under-reporting of cases and delays in reporting. We attempted to account for both problems by introducing a parameter representing the proportion of actual cases detected via testing, as well as a delay in reporting. We used plausible biological parameters for SARS-CoV-2 based on current evidence, but these values might be refined as more clinical evidence becomes available. The standardized intervention measures from Oxford Covid-19 Government Response Tracker, may also suffer from inaccuracies. As the updating frequency of the tracker is performed once a week in most countries, we infer the days with missing policy using the last observed policy in that country.

4.2 Conclusions

This study provides an estimate of the daily effective reproduction number of SARS-CoV-2 across countries and over time as various public health interventions were adopted. This allowed us to look at the evolution of disease transmission in the presence of containment measures. At the end of the study period, predictions regarding future outbreaks were made with reasonable accuracy, which were sensitive to the currency of the chosen epidemiological parameters. Heterogeneity of the impact of social distancing on transmission rates across countries could not be explained solely by the timing and number of adopted interventions, implying that the ways in which these policies were implemented and how people responded to them may have an important role.

[1] WHO. Rolling updates on coronavirus disease (COVID-19); 2019. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.

[2] Ferguson N, Laydon D, Gilani GN, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 2020;.

[3] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. {Annals of internal medicine}. 2020;172(9):577–582.
[4] Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. The Lancet Public Health. 2020;.
[5] Giordano G, Blanchini F, Bruno R, Colaneri P, Filippo AD, Matteo AD, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine. 2020;26(6):855–860. Available from: https://dx.doi.org/10.1038/s41591-020-0883-7.
[6] Flaxman S, Mishra S, Gandy A, Unwin H, Coupland H, Mellan T, et al. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. 2020;.
[7] Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. {The Lancet Infectious Diseases}. 2020;.
[8] Ferguson N, Laydon D, Gilani GN, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 2020;.
[9] Hale T, Webster S, Petherick A, Phillips T, Kira B. Oxford covid-19 government response tracker; 2020. Available from: 2020-05-10.
[10] Dureau J, Kalogeropoulos K, Baguelin M. Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. {Biostatistics}. 2013;14(3):541–555.
[11] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time; 2020.
[12] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. {New England Journal of Medicine}. 2020;.
[13] nCoV 2019 Data Working Group. Epidemiological data from the nCoV-2019 outbreak: early descriptions from publicly available data; 2020.
[14] Liu T, Hu J, Kang M, Lin L, Zhong H, Xiao J, et al. Transmission dynamics of 2019 novel coronavirus (2019-nCoV). 2020;.
[15] Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. {Analytical Chemistry}. 1964;36(8):1627–1639.
[16] Schellekens, Sourrouille DP. COVID-19 Mortality in Rich and Poor Countries : A Tale of Two Pandemics? {Policy Research Working Paper;No 9260 World Bank}. 2020;.
[17] Educational STUN, Organization C. The UNESCO Institute for Statistics (UIS);. Available from: http://uis.unesco.org.
[18] Dye C, Gay N. Modeling the SARS epidemic. Science. 2003;300(5627):1884–1885.
[19] Galvani AP. Emerging infections: what have we learned from SARS? Emerging infectious diseases. 2004;10(7):1351.
[20] Ritchie H, Ortiz-Ospina E, Roser M, Hasell J. COVID-19 deaths and cases: how do sources compare?;. Available from: https://ourworldindata.org/covid-sources-comparison.