

The true case fatality of COVID-19: An analytical solution

Syamantak Khan

Department of Radiation Oncology, Stanford University. drskhan@stanford.edu

I. Supplementary Tables:

TableS1: COVID-19 related data from countries with more than 50,000 testings till 14 April 2020

Country	Number of Testing	Number of detected cases	Number of Deaths	Number of Test/ Total Case	Observed Case fatality (%)	Country	Number of Testing	Number of detected cases	Number of Deaths	Number of Test/ Total Case	Observed Case fatality (%)
USA	3058078	610206	25830	5.01155	4.233	Peru	102216	10303	230	9.92099	2.23236
Russia	1400000	21102	170	66.34442	0.80561	Belgium	102151	31119	4157	3.28259	13.3584
Germany	1317887	131359	3294	10.03271	2.50763	Thailand	100498	2613	41	38.46077	1.56908
Italy	1073689	162488	21067	6.6078	12.96527	Hong Kong	96709	1013	4	95.46792	0.39487
UAE	648195	4933	28	131.3998	0.56761	Ireland	90646	11479	406	7.89668	3.53689
Spain	600000	172541	18056	3.47743	10.46476	Chile	87794	7917	92	11.0893	1.16206
S. Korea	527438	10564	222	49.92787	2.10148	South Africa	87022	2415	27	36.03395	1.11801
Turkey	443626	65111	1403	6.81338	2.15478	Malaysia	84791	4987	82	17.00241	1.64428
Canada	437475	26897	898	16.26483	3.33866	Japan	78702	7645	143	10.29457	1.8705
UK	382650	93873	12107	4.07625	12.89721	Kazakhstan	76904	1232	14	62.42208	1.13636
Australia	366493	6400	61	57.26453	0.95313	Denmark	74210	6511	299	11.39763	4.59223
France	333807	143303	15729	2.32938	10.97604	Singapore	72680	3252	10	22.34932	0.3075
Iran	287359	74877	4683	3.83775	6.25426	Belarus	71875	3281	33	21.90643	1.00579
India	244893	11487	393	21.31914	3.42126	Azerbaijan	71736	1197	13	59.92982	1.08605
Switzerland	199000	25936	1174	7.67273	4.52653	Romania	70097	6879	351	10.19	5.10249
Portugal	182707	17448	567	10.47152	3.24966	Uzbekistan	70000	1165	4	60.08584	0.34335
Austria	151796	14226	384	10.67032	2.69928	Pakistan	69928	5837	96	11.98013	1.64468
Saudi Arabia	150000	5369	73	27.93816	1.35966	Bahrain	69359	1528	7	45.39202	0.45812
Poland	148321	7202	263	20.59442	3.65176	New Zealand	64399	1366	9	47.14422	0.65886
Netherlands	134972	27419	2945	4.92257	10.74073	Brazil	62985	24920	1489	2.52749	5.97512
Israel	117339	12046	123	9.74091	1.02109	Sweden	54700	11445	1033	4.77938	9.02578
Czechia	131542	6111	161	21.52545	2.63459	Qatar	52622	3428	7	15.35064	0.2042
Norway	128569	6623	139	19.4125	2.09875						
Vietnam	121821	266		457.9737	0						

^a The data was obtained from worldometers.info, which uses official reports, government's communication channels, and local media when deemed reliable as the sources of COVID-19 related data.

II. Supplementary Figures:

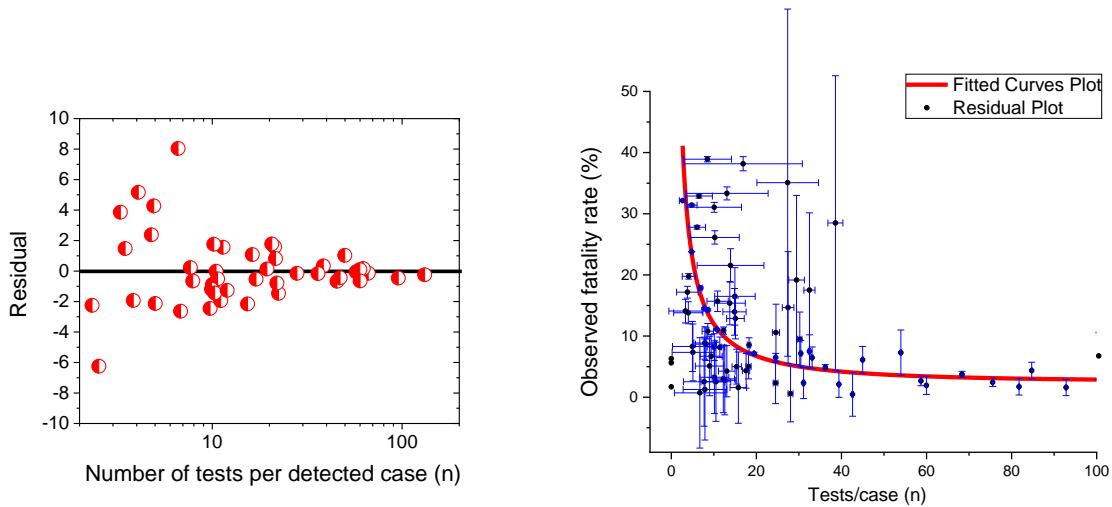


Figure S1: Residual plot for fitting the dataset with equation 7. The residual R squared value of the fitting was found to be 0.95 and can be improved by excluding a few outliers.

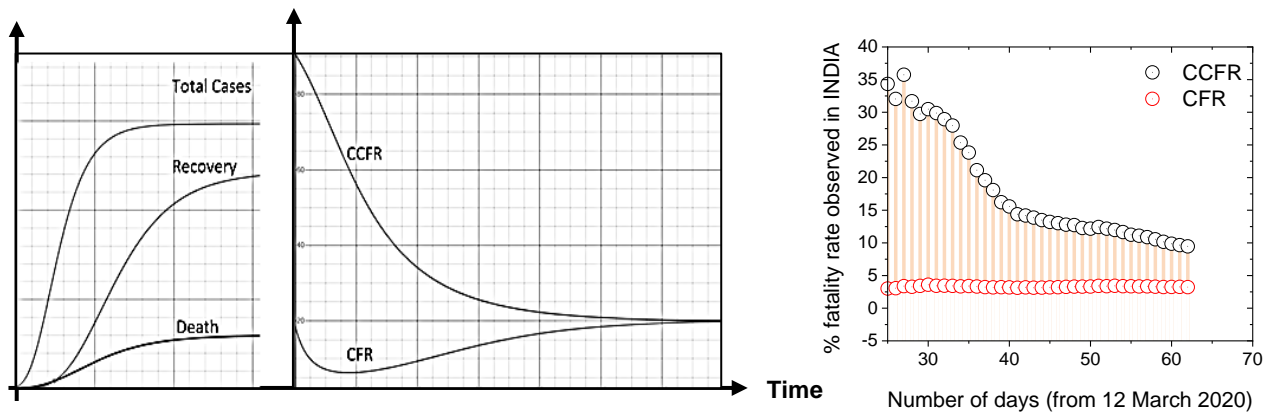


Figure S2: The dynamics of the fatality rates during a pandemic. LEFT: A simulated curve of CCFR and CFR using a modified SIR model (not discussed here) RIGHT: CCFR and CFR of INDIA calculated till 15 May 2020 (Data source: www.covid19india.org)

III. Supplementary Notes:

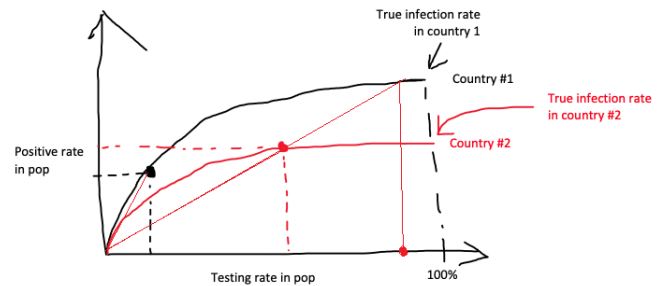
Q1: Why is contact-tracing used in this modeling?

The probability of the spread of infection from one particular individual gradually decreases from the close-contacts (persons who are close, or came close to the infected individual) to the far-contacts (persons who were physically distant from the infected individual). The probability of the spread of the infection approaches to zero for someone who is located at a large distance from the infected individual. To overcome this challenge, the detection strategies (COVID-19 testing) follow an efficient algorithm.

Most of the COVID-19 testing is done by finding close contacts or contact-tracing. When someone is tested positive, her/his family and friends, who came close to that person in the last 10-14 days are tested. However, when more number of suspects are tested (*tests-per-case*), the detection efficiency increases. Then, if they find a secondary infection, her/his close contacts are tested again. It goes on like a chain reaction. The testing need keeps growing according to the epidemic growth and the *tests-per-case* ratio. There is also some random testing being done but, it is a small fraction of the total test. The randomly detected individual also starts new chains of contact-tracing initiating from them. This is why the most affected places are currently testing more than other places.

Q2: How is tests-per-case related to the testing rate within the population?

Answer: When the testing rate is increased within a population, more infections are detected. However, the % positive rate of testing progressively decreases with further expansion of testing. This means, it will increase quickly when there are many undetected cases, but it will gradually stop increasing when most of the infected people have been identified. The aim is to accurately model the rate of detected cases in the population as a function of (1) the testing rate and (2) the true infection rate of the population. The figure here illustrates this expected relationship for two differently affected countries (red and black lines). In this example, country #1 (black dot) has fewer detected cases (per 1 M) than county #2 (red dot), but higher infection rate in reality (due to different testing rate). The observed fatality rate would, therefore, be higher in country #1 than #2. It is interesting to note that the slope of the two points is independent of the population and represents the testing rate more accurately. This slope mathematically represents the %positive rate of testing or a reciprocal of the *tests-per-case* (Thanks to Dr. G Pratz for the hand-drawn schematic).



Q3: Why is tests-per-million not used in this model instead of tests-per-case?

Answer: Although counter-intuitive, the testing rate can be more than 100% in a population. The epidemic grows as a function of time and most of the infected people recover after a few days of infection, the testing is also performed dynamically according to the outbreak status. It is important to notice that, the unaffected population, once tested negative, can also contract the infection in the future. As a result, the same individual needs to be tested twice or thrice if they are suspects at different points of time. Moreover, the PCR testing is not confirmatory (has false negatives), some cases are missed. This means, testing everybody at a given point does not ensure 100% detection efficiency. In other words, the testing is aimed to accurately identify infection within the outbreak region, not in the whole population. A small number of random tests are often useful to detect new outbreaks, but most of the testing is performed where the probability of finding infection is maximum. As a result, the percent of the population tested has limited relevance to the actual detection efficiency of a country. In contrast, the average number of persons tested per confirmed case is free from that bias and represents a robust estimate for a testing rate within the outbreak region. A low value of *tests-per-case* indicates a high %positive rate (total case/total test) in testing, which is an indication of under-testing. This parameter is much more relevant and useful for mathematical modeling of COVID-19 outbreak, detection efficiency, or case fatality rate. For instance, the figure here shows that unlike the *tests-per-case*, *tests-per-million* is not well correlated with the observed case fatality rate.

