ESTIMATION OF COVID-19 CASES IN FRANCE AND IN DIFFERENT COUNTRIES: HOMOGENEISATION BASED ON MORTALITY

Marc DHENAIN

(1) Académie Vétérinaire de France, 34, rue Bréguet, 75011 Paris, France
(2) Académie Nationale de Médecine, 16 rue Bonaparte, 75006 Paris, France
(3) Centre National de la Recherche Scientifique (CNRS), Université Paris-Sud, Université Paris-Saclay UMR 9199, Laboratoire des Maladies Neurodégénératives, 18 Route du Panorama, F-92265 Fontenay-aux-Roses, France.
(4) Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Institut François Jacob, Molecular Imaging Research Center (MIRCen), 18 Route du Panorama, F-92265 Fontenay-aux-Roses, France.

Correspondance
Marc Dhenain
MIRCen, UMR CEA-CNRS 9199, 18 Route du Panorama, 92 265 Fontenay-aux-Roses CEDEX, France
Tel: +33 1 46 54 81 92; Fax: +33 1 46 54 84 51; email: Marc.Dhenain@cea.fr

Abstract:

Every day the authorities of different countries provide an estimate of the number of persons affected by Covid-19 and a count of fatality. We propose to use the fatality reported in each country to provide a better estimate ($C_{t0\text{-estimated}}$) of the number of cases at a given time t_0.

$$C_{t0\text{-estimated}} = (F_{t0} / F_{r\text{-est}}) * (1 + [C_{(est\text{-}d)} / C_{(est\text{-}3d)}])^6$$

With F_{t0}: number of actual fatalities reported in a country at time t_0; $F_{r\text{-est}}$: estimated fatality rate; $C_{(est\text{-}d)}$: estimated fatalities 18 days before t_0; $C_{(est\text{-}3d)}$: estimated fatalities 21 days before t_0.

Based on $C_{t0\text{-estimated}}$ calculated using a fatality rate of 2%, we assessed the number of cases April 10th, 2020 in Belgium, China, France, Germany, Iran, Italy, South Korea, Netherlands, Spain, United Kingdom and USA. This number reached 2,872,097 in France and 924,892 persons in Germany. The proposed formulas also make it possible to evaluate the impact of policies to prevent the spread of epidemic on the appearance of new cases.

Key-Words*: *Covid-19, Estimated number of cases, Mortality, Prevalence

Version submitted on April 13th 2020

A French first version of this article is "in press" as
Dhenain Marc, Estimation du nombre de cas de Covid-19 en France et dans différents pays : homogénéisation basée dur la mortalité, Bulletin de l'Académie Vétérinaire de France, 2020 (provisionally accepted), https://academie-veterinaire-defrance.org/bavf-coronavirus/

INTRODUCTION

The Sars-CoV-2 coronavirus infection that causes Covid-19 has spread worldwide leading to significant deaths (European Center for Disease Prevention and Control 2020). Every day the authorities of different countries provide and distribute worldwide an estimate of the number of affected persons and a count of fatalities (Dong et al. 2020, https://ourworldindata.org/covid-testing, https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series).

Knowing the number of affected subjects is critical for implementing strategies to protect populations and for ending the crisis. Figures reported by different countries reveal strong differences and only partly reflects the reality (Table I). For example, the day their death toll approached 3,000 people, France had 44,550 people affected versus 80,537 for China and 122,171 for Germany. Calculating the case fatality rate (F_r) on a given day (t_0) is another way to objectify differences between countries. At first sight,

$$F_r = Ft_0 / Ct_0$$

With Ft_0 = number of fatalities reported on day t_0; Ct_0 = number of cases reported on day t_0.

The day when the death toll of different countries was the closest to 3,000 people, three countries (Germany, South Korea, and the United States) had fatality rates close to 2%; seven countries (Belgium, France, Iran, Italy, the Netherlands, Spain, and the United Kingdom) had rates between 6% and 12%, and China had an intermediate value of 3.7% (Table I).

Patients who die on any given day were infected much earlier, and thus the denominator of the fatality rate should be the total number of patients infected at the same time as those who died (Baud et al. 2020). This is particularly true as the rates of evolution of the pandemic evolve differently in various countries: in March 2020, the number of people affected increased sharply from day to day in France, while it was stable in China.

A better estimate of fatality rate is thus:

$$F_{r\text{-}xday} = F_{t0} / C_{t0\text{-}xdays}$$

With $C_{t0\text{-}xdays}$ = number of cases reported on day t_0 minus x days, with x = average time-period from onset of symptoms to death.

An average duration of 18 days is reported between the onset of symptoms and the death of Covid-19 patients (Ruan et al. 2020; Verity et al. 2020; Zhou et al. 2020). Thus the adjusted Fatality rate ($F_{r\text{-}18}$) that takes into account this average delay is (Flaxman et al. 2020).

$$F_{r\text{-}18} = F_{t0} / C_{t0\text{-}18d} \qquad \text{(Table I)}$$

With C_{t0-18d} = number of cases reported on day t_0 minus 18 days. The calculation of F_{r-18} reveals widening gaps between countries compared to F_r with variations ranging from 2.3% (South Korea) to more than 700% for Spain.

When comparing F_{t0} and C_{t0-18d} in different countries (Fig. 1), we see a linear relationship between mortality at t_0 and the number of cases at $t_{0-18days}$ for all countries (Pearson linear correlation test, $p<0.05$ except for Belgium ($p=0.07$) due to the small number of points ($n=3$)). The slopes of the regression lines fitting the data vary widely between countries, which is consistent with variable F_{r-18}. The values of fatality rate F_{r-18} based on the cases reported by the different countries are therefore unreliable, in part because the number of cases reported in different countries is not reliable (different testing strategies in different countries).

How to assess more precisely the number of people affected using a similar method for different countries? We offer a simple method using the number of deaths reported by each country to estimate and compare the actual rate of people affected by Covid-19. This method relies on three first assumptions: 1. The number of deaths reported by each country is reliable; 2. The fatality rate (F_r) is known and similar in different countries; 3. The average time between the onset of symptoms and death is known (here considered 18 days). Based on these assumptions, one can calculate the number of cases presented eighteen days before a given day (t_0). Two methods are then proposed to infer the number of cases, eighteen days later, at time t_0. The first one relies on the time-dependent increase in the number of cases reported in databases during the average time between the onset of symptoms and death. The second one models the data based on daily rate of changes of the number of estimated cases 18 days before t_0.

METHODS

One way to estimate the number of Covid-19 cases is to infer the number of cases based on the number of death and the fatality rate calculated from well-controlled studies using the following formula

$$C_{(est-18d)} = F_{t0} / F_{r-est} \qquad \text{(Eq. 1, Table II)}$$

With $C_{(est-18d)}$: number of cases estimated 18 days before t_0; Ft_0 = number of deaths reported on day t_0; F_{r-est} = estimated fatality rate from well-controlled studies. F_{r-est} can be assessed from well-controlled international studies based on residents of mainland China, travelers returning from mainland China, repatriated from China, passengers on the *Diamond-Princess* cruise ship (reported values of 0.7 to 3.6% (Verity et al. 2020)). Here, based on this last study, we proposed to use F_{r-est} = 2%.

Knowing $C_{(est-18d)}$ eighteen days before t0, one needs to assess the progression rate of the cases during the 18 last days to estimate the number of cases at day t_0 ($C_{t0-estimated}$). One of the difficulties is that this progression evolves with time and is different in different countries. We tested two methods to assess this progression.

Method 1. One can assume that progression of cases (P_{18d}) reflects the time-dependent increase in the reported number of cases during the same time-period. In that case,

$$P_{18d} = C_{t0} / C_{t-18d}$$

With C_{t0}: number of cases reported in a country at time t_0; C_{to-18d}: number of cases reported in the country at time t0 minus 18 days. Thus,

$$C_{t0-estimated} = C_{(est-18d)} * P_{18d} \qquad \text{(Table II)}$$

A potential limitation of this estimation is that it assumes that the progression P_{18d} reflects the time-dependent increase in the actual number of cases.

Method 2. One can calculate the daily rate of change (R_d) or 3-day rate of change (R_{3d}) of the estimated $C_{(est-18d)}$ in a given country. The last day when this calculation is possible is 18 days before t_0.

$$R_{3d} = C_{(est-d)} / C_{(est-3d)}$$

With $C_{(est-d)}$: estimated $C_{(est-18d)}$ a given day; $C_{(est-3d)}$: estimated $C_{(est-18d)}$ three days before Assuming that the progression of the estimated cases follows an exponential model then

$$C_{t0-estimated} = C_{(est-18d)} * (1+ R_{3d})^6$$

6 represents the period of the model as 6 * 3 days = 18 days

A potential limitation of this measure is related to the fact that R_{3d} evolves with time due to immunization of population and containment measures. We thus propose to freeze R_{3d} the last day when it can be measured.

RESULTS

Using the (Eq. 1) and international databases (Dong et al. 2020, https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series), we estimated the number of people that have been affected by Covid-19 eighteen days before April 10^{st} 2020 in different countries (Table II). This estimation was 659,850 persons in France and 138,350 in Germany. We then proposed two different methods to infer the number of cases 18 days latter (Fig. 2). Method based on P_{18d} evaluation based on the reported number of cases during the same time-period provided different results according to the countries and led to a time-dependent decrease of number of cases in some countries (*e.g.* Germany (Fig. 2C) or USA (Fig. 2D)), which is not consistent. Method based on the evaluation of R_{3d} provided a better correspondence with the estimated cases in all tested countries (Fig. 2). We thus retained results from the R_{3d} method for further investigations. Estimation of the number of cases based on this method, April 10th was 2,872,097 in France, 924,892 in Germany, 1,811,469 in Spain, 4,240,198 in the United Kingdom and 9,035,229 in the United States (Table II, Fig. 2, Fig. 3A). This analysis was based on $M_{r\text{-}est}$ = 2%. The estimated number of cases must be halved if the mortality rate used drops from 2 to 4% (Table II). It must be doubled if the fatality rate used goes from 2 to 1%. Note that some authors suggest that the real fatality rate for Covid-19 could be 5.6 to 15.6% (Baud et al. 2020), which is much higher values than those we used. If the calculation uses a fatality rate of 15%, then the estimated number of cases drops to 382,946 for France, but it becomes lower than the number of cases actually reported for some countries (e.g. 2,242 *versus* 10,450 for South Korea), which is not consistent (Table II).

In our study, we set the delay between the onset of symptoms and death at 18 days based on robust data from the literature (Ruan et al. 2020; Verity et al. 2020; Zhou et al. 2020) and delays used in other models (Flaxman et al. 2020). Lowering this delay, for example to 12 days, sharply decreases the number of estimated cases (*e.g.* 1,645,302 for France (Table II)) although it remains high compared to figures reported in databases.

Using estimations based on R_{3d} model, with delay of 18 days between the onset of symptoms and death and fatality rate of 2%, we could thus compare estimated cases of Covid-19 in different countries (Fig. 3A), proportion of cases in different countries (Fig. 3B), as well as notification indexes which is the ability to report cases (Fig. 3C-D). These data highlight strong discrepancies between countries. It suggests a high proportion of affected persons in Belgium. It also shows notification indexes that varies from 60 to 80% in Korea while it is below 5% in most countries.

Interestingly, since $C_{t0\text{-}estimated}$ takes into account the 3-day rate of change, during the last eighteen days, it can be used to model how policies to prevent disease spreading modulates the appearance of new cases. For example, for France the R_{3d} used for estimation of cases April 10th was 0.28 (measured March 23rd), while five day before on March 19th, when population containment was less efficient, R_{3d} was 0.50. We can therefore estimate that 5 days containment made it possible to reduce R_{3d} from 0.50 to 0.28. The number of cases estimated on April 10th using Rd = 0.50 would have been 7,516,104 cases. Thus, the containment from March 10th to April 10th prevented the appearance of 4,644,007 new cases in France, as well as associated fatalities.

DISCUSSION

Evaluating the number of Covid-19 cases present in a country is critical to predict the end of the crisis or containment. We proposed to estimate the number of cases on the estimation of the number of cases estimated 18 days before a given day t_0. Then, from that estimate, two methods were proposed to infer the number of cases 18 days later. The first one relies on the time-dependent increase in the number of cases reported in databases during the average time between the onset of symptoms and death. This method led to a large number of cases and to day to day variations of estimated numbers, that are due to daily variations of reported cases, a bias already reported by previous studies (Flaxman et al. 2020).

The second method provided more plausible results. It can be presented as

$$C_{t0\text{-estimated}} = (F_{t0} / F_{r\text{-est}}) * (1+ [C_{(est\text{-}d)} / C_{(est\text{-}3d)}])^6$$

With F_{t0}: number of fatalities reported in a country at time t_0; $F_{r\text{-est}}$: estimated fatality rate; $C_{(est\text{-}d)}$: estimated $C_{(est\text{-}18d)}$ 18 days before t_0; $C_{(est\text{-}3d)}$: estimated $C_{(est\text{-}18d)}$ three days earlier. This analysis is based on four assumptions: 1. The number of deaths reported by each country is reliable, 2. The estimated fatality rate among people affected is known ($F_{r\text{-est}}$, here considered as 2%), 3. The average time between the onset of symptoms and death is known (here considered 18 days). 4. The three-day rate of change of the estimated cases (R_{3d}) does not change during the last 18 days. Because of the containment, the three-day rate of change is continually decreasing, which probably leads to an overestimation of the actual cases.

Our analysis suggests that the number of Covid-19 cases in several country greatly exceeds the number of cases presented in international databases (2,872,097 versus 124,869 for France on April 10th, 2020). The very high values of estimated cases that we report are consistent with those evaluated with another method by (Flaxman et al. 2020). For example, we report 1.8 million cases in Spain while Flaxman reports 7.0 million on March 28th. Our calculation relies on a relatively simple method while that of Flaxman et al. uses more complex analyzes (hierarchical semi-mechanistic Bayesian model). Our model used a fatality rate of 2% while several strongly controlled international studies reported rates of 0.7 to 3.6% (Verity et al. 2020). Values from 0.5 to 4% could thus be other reasonable options to estimate fatality rate. One of the limitations of our model is that fatality rates can change from one country to another, for example depending on the distribution of the population of different age groups that have different susceptibility to Covid-19. Also, it is possible that death rate change over time in a given country, for example because of the saturation of hospitals. We fixed a single value for the time between symptom occurrence and death (18 days). In reality, this time is variable with a 95% credible interval of 16.9 to 19.2 or more according to (Verity et al. 2020). We however considered that using such interval would make the model more complicated without strongly adding reliability compared to other potential sources of errors that come from other assumptions. Note that our analysis relies solely on the number of deceased people with confirmed cases of Covid-19 and misses deaths not reported as Covid-19-related. It is critical for all countries to be able to provide highly reliable values of fatalities as they will be the only exact figure to reflect the impact of the disease in the future. Finally, note that to know the number of actual cases in a country at a given time, we must subtract from the estimates presented here the number of people healed, including those whose disease has not been identified.

REFERENCES

Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection [published online ahead of print, 2020 Feb 19]. Lancet Infect Dis. 2020. doi: 10.1016/S1473-3099(20)30195-X.

Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time [published online ahead of print, 2020 Feb 19]. Lancet Infect Dis. 2020;S1473-3099(20)30120-1. doi:10.1016/S1473-3099(20)30120-1. Database available on: https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series Consulted 2020/04/04

Flaxman S, Mishra S, Gandy A, Unwin JT, Coupland H, Mellan TA et al. 2020. Estimating the number of infections and the impact of nonpharmaceutical interventions on COVID-19 in 11 European countries. Imperial College London (30-03-2020). doi: 10.25561/77731

European Centre for Disease Prevention and Control. Situation Update-Worldwide. Disponible à: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases. Consulted 2020/04/04

Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China [published online ahead of print, 2020 Mar 3]. Intensive Care Med. 2020;1–3. doi:10.1007/s00134-020-05991-x

Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis [published online ahead of print, 2020 Mar 30]. Lancet Infect Dis. 2020;S1473-3099(20)30243-7. doi:10.1016/S1473-3099(20)30243-7

Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z et al. Clinical course and risk factors for mortality of adult in patients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020 ; 395: 1054-1062. doi: 10.1016/S0140-6736(20)30566-3

FIGURES AND TABLES

Country	Day (t_0)	Fatalities at t_0 (F_{t0})	Cases at t_0 (C_{t0})	Fatality rate F_r	Cases at t_{0-18j} (Ct_{0-18d})	Fatality rate F_{r-18d}
Belgium	April 10	3 019	26 667	11.3%	3743	80.7%
China	March 5	3 015	80 537	3.7%	70 513	4.3%
France	March 30	3 024	44 550	6.8%	2 281	132.6%
Germany	April 10	2 767	122 171	2.3%	29 056	9.5%
Iran	April 1	3 036	47 593	6.4%	12 729	23.9%
Italy	March 18	2 978	35 713	8.3%	1 128	264.0%
South-Korea	April 10	208	10 450	2.0%	8 961	2.3%
Netherlands	April 10	2 511	23 097	10.9%	4 749	52.9%
Spain	March 24	2 808	39 885	7.0%	400	702.0%
United Kingdom	April 2	2 921	33 718	8.6%	1 140	256.2%
USA	March 30	2 978	161 807	1.8%	1 163	256.1%

Table I: Fatality rates in different countries when the number of deaths approached 3,000 people (or the last figure available when the 3,000 deaths were not reached April 10th 2020).
(https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series)

					Method 1&2	Method 1 Progression rate from actual cases		Method 2 Estimation from 3-days rate of change				
Country	Population (million)	Reported cases at t_0 (April 10) (Ct_0)	Reported deaths at t_0 (April 10) (F_{t0})	Reported cases at t_{0-18d} (March 23) (Ct_{0-18d})	Estimated cases at t_{0-18d} (March 23) $C_{(est-18d)}$	Progression rate from t_{-18} to t_0 $P_{18d} = Ct0 / Ct\text{-}18d$ (April 10) P_{18d}	Estimated Cases (April 10) $C_{t0\text{-}estimated}$	R_{3d} (April 10) $C_{t0\text{-}estimated}$	Estimated cases (April 10) $C_{t0\text{-}estimated}$	Estimated cases (April 10) $C_{t0\text{-}estimated}$	Estimated cases (April 10) $C_{t0\text{-}estimated}$	Estimated cases (April 10) $C_{t0\text{-}estimated}$
Delay							**t_{0-18d}**	**t_{0-18d}**	**t_{0-18d}**	**t_{0-18d}**	**t_{0-18d}**	**t_{0-12d}**
M_{r-est}							**2%**	**2%**	**2%**	**4%**	**15%**	**2%**
Belgium	11,476	26 667	3019	3 743	150 950	7,12	1 075 443	0,48	**1 609 257**	**804 628**	**214 568**	**683 894**
China	1,384,688	82 941	3340	81 498	167 000	1,02	169 957	0,00	**168 508**	**84 254**	**22 468**	**168 003**
France	67,795	124 869	13197	19 856	659 850	6,29	4 149 618	0,28	**2 872 097**	**1 436 049**	**382 946**	**1 645 302**
Germany	83,073	122 171	2767	29 056	138 350	4,20	581 717	0,37	**924 892**	**462 446**	**123 319**	**482 669**
Iran	82,022	68 192	4232	23 049	211 600	2,96	626 033	0,09	**360 726**	**180 363**	**48 097**	**294 241**
Italy	60,360	147 577	18849	63 927	942 450	2,31	2 175 668	0,10	**1 674 559**	**837 279**	**223 275**	**1 364 771**
South-Korea	51,709	10 450	208	8 961	10 400	1,17	12 128	0,08	**16 811**	**8 406**	**2 242**	**14 306**
Netherlands	17,282	23 097	2511	4 749	125 550	4,86	610 619	0,20	**365 882**	**182 941**	**48 784**	**240 570**
Spain	46,935	158 273	16081	35 136	804 050	4,50	3 621 909	0,14	**1 811 469**	**905 735**	**241 529**	**1 345 788**
United Kingdom	65,761	73 758	8958	6 650	447 900	11,09	4 967 851	0,45	**4 240 198**	**2 120 099**	**565 360**	**2 041 524**
USA	328,240	496 535	18586	43 847	929 300	11,32	10 523 638	0,46	**9 035 229**	**4 517 615**	**1 204 697**	**4 086 903**

Table II: Estimation of the number of cases in different countries April 10st (t_0) using different methods and an estimated Fatality rate (F_{r-est}) of 2%. Numbers of cases estimated with different methods are provided using delays of 18 or 12 days between symptom occurrence and death.

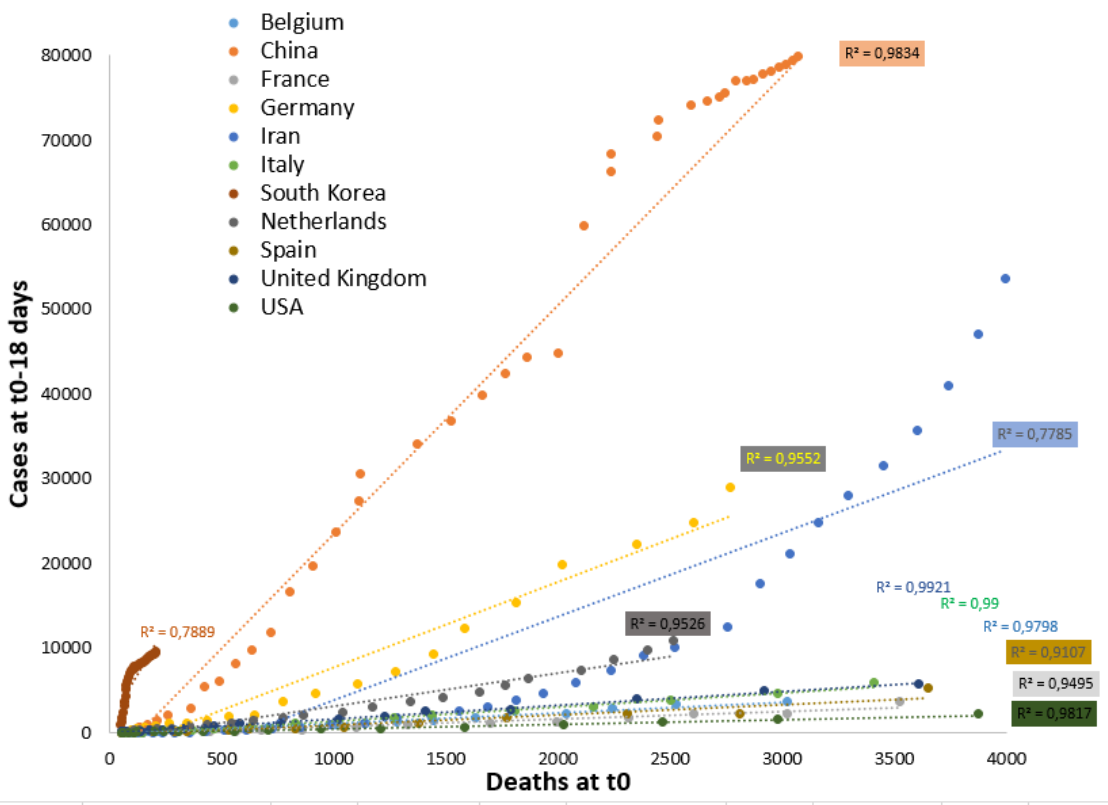

Figure 1: Relationships between deaths a given day (t_0) and the number of cases eighteen days before ($t_{0\text{-}18days}$) in different countries. The figure includes only values between 50 and 4 000 deaths (or less if the number of deaths was lower in the country April 10th).

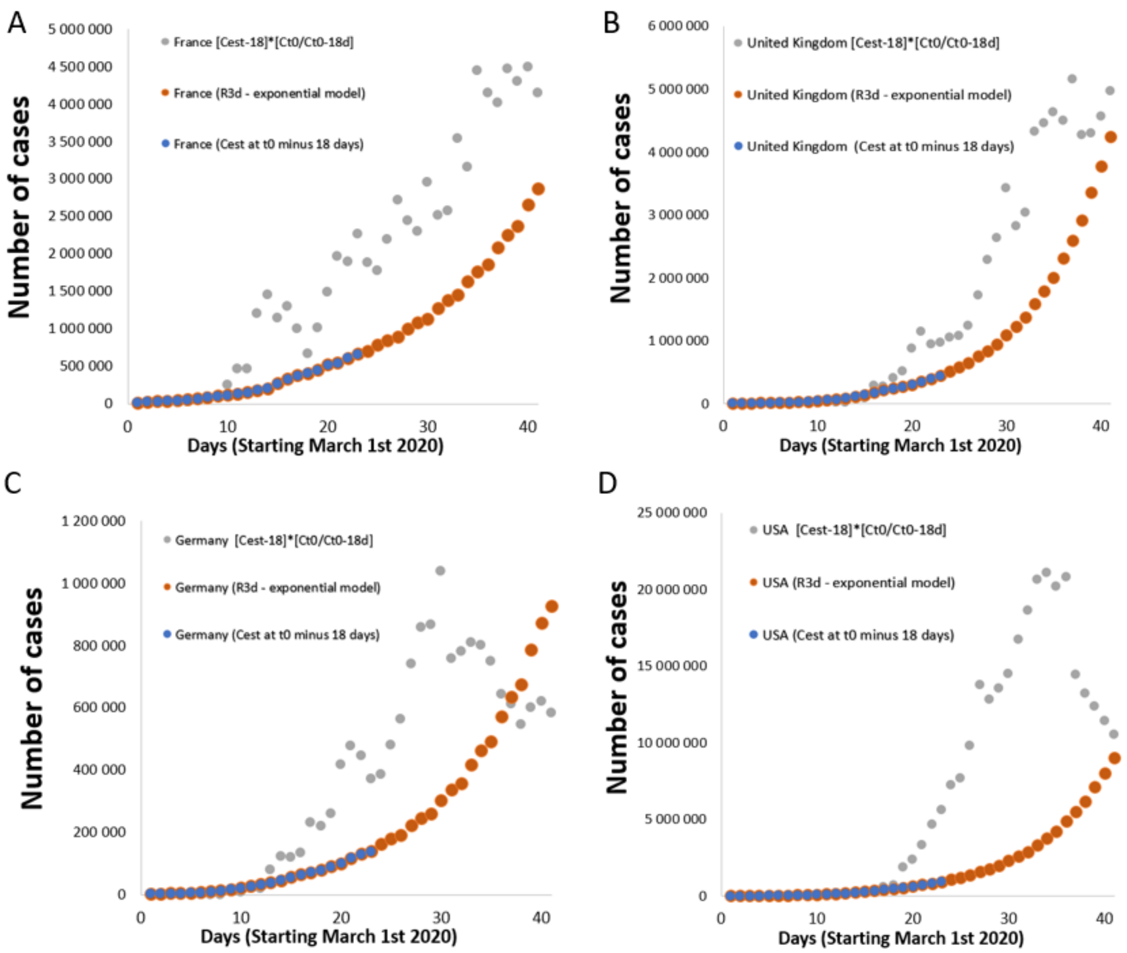

Figure 2: Comparison of evolution of estimated Covid-19 cases in France (A), United Kingdom (B), Germany (C), and USA (D) from March 1st 2020 to April 10th 2020. Estimated values corresponding to C_{est} from fatalities 18 days later are displayed in blue. Gray marks correspond to estimation based on time-dependent increase in the number of cases reported in databases during the average time between the onset of symptoms and death (18 days). This method leads to large number of cases and day-to-day variations. Orange marks correspond to a model based on R_{3d}. It provides curves that follow-up values corresponding to C_{est} from fatalities estimated 18 days before a given day (blue marks).

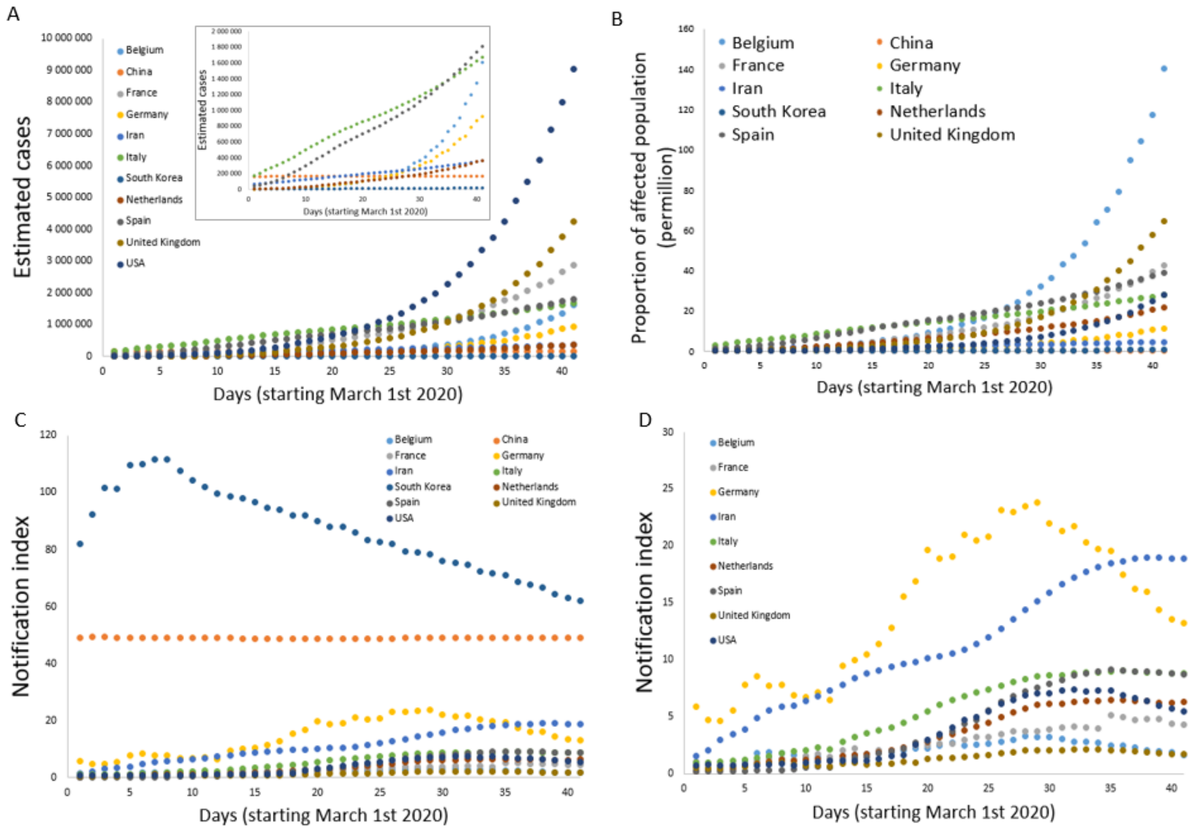

Fig. 3. *Comparison of estimated cases and related parameters in different countries. A. estimated cases in different countries. B. Proportion of affected person compared to the country population. C-D. Notifications indexes reflecting the number of cases reported by different countries compared to the estimated number of cases (percentages).*