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 mortality. We propose to use the mortality reported in each country to provide a better estimate ($C_{t0\text{-estimated}}$) of the number of cases at a given time t_0.

$$C_{t0\text{-estimated}} = (M_{t0} / M_{r\text{-est}}) * (C_{t0} / C_{t0\text{-}18d})$$

With M_{t0}: number of deaths reported in a country at time t_0; $M_{r\text{-est}}$: estimated mortality rate; C_{t0}: number of cases reported in a country at time t_0; $C_{to\text{-}18d}$: number of cases reported in the country at time t0 minus 18 days.

Based on this index and an estimated mortality rate of 2%, we assessed the number of cases April 2, 2020 in Belgium, China, France, Germany, Iran, Italy, South Korea, Netherlands, Spain, United Kingdom and USA. This number reaches 3,538,549 in France and 809,896 persons in Germany. The proposed formula also makes 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

A French 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 (in press), 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 mortality (Dong et al. 2020, 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 approximately 45,000 people affected versus 80,537 for China. Germany reported only 1,107 deaths when 84,794 people were affected. Calculating the case fatality rate (CFR) or mortality rate (M_r) on a given day (t_0) is another way to objectify differences between countries. At first sight,

$$M_r = Mt_0 / Ct_0$$

With Mt_0 = number of deaths 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 mortality rates close to 1.6%; seven countries (Belgium, France, Iran, Italy, the Netherlands, Spain, and the United Kingdom) had rates between 6% and 9.1%, and China had an intermediate value of 3.7% (Table I).

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 four assumptions: 1. The number of deaths reported by each country is reliable; 2. The mortality rate (M_r) is known and similar in different countries; 3. The average time between the onset of symptoms and death is known; 4. The time-dependent increase in the number of cases reported in databases during the average time between the onset of symptoms and death reflects the time-dependent increase in the actual number of cases during the same time-period.

METHODS

Patients who die on any given day were infected much earlier, and thus the denominator of the mortality 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 mortality rate is thus:

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

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). We can therefore calculate an adjusted mortality rate ($M_{r\text{-}18}$) by taking into account this average delay of 18 days (Flaxman et al. 2020).

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

With $C_{t0\text{-}18d}$ = number of cases reported on day t_0 minus 18 days. The calculation of $M_{r\text{-}18}$ reveals widening gaps between countries compared to M_r with variations ranging from 2.1% (South Korea) to more than 700% for Spain. When comparing M_{t0} and $C_{t0\text{-}18d}$ in different countries (Fig. 1), we see a linear relationship between mortality at t_0 and the number of cases at $t_{0\text{-}18days}$ for most countries (except China and South Korea). The slope of this relationship varies widely between countries, which is consistent with variable $M_{r\text{-}18}$. The values of $M_{r\text{-}18}$ based on the cases reported by the different countries are therefore unreliable. Several highly controlled international studies (residents of mainland China, travelers returning from mainland China, repatriated from China, passengers on the *Diamond-Princess* cruise ship) reported $M_{r\text{-}18}$ of 0.7 to 3.6% (Verity et al. 2020). Based on this last study, an estimated mortality rate can be proposed $M_{r\text{-}est}$ = 2%.

By knowing the number of deaths on day t_0, we can therefore estimate the number of cases ($C_{(est\text{-}18d)}$) eighteen days before:

$$C_{(est\text{-}18d)} = M_{t0} / M_{r\text{-}est}$$

The increase in the number of reported cases over the past eighteen days (“progression rate”: P_{18d}) is known in each country. It is equal to:

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

If we assume that this progression rate reflects the time-dependent increase in the actual number of cases during the same time-period, then the number of cases estimated at t_0 in each country is:

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

RESULTS

Using the previous equation 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 currently or previously affected by Covid-19 on April 2nd 2020 in different countries (Table II). This estimation was 3,538,549 in France, 809,896 in Germany, 7,435,552 in Spain, 4,319,749 in the United Kingdom and 20,615,926 in the United States.

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 III). It must be doubled if the mortality rate used goes from 2 to 1%. Note that some authors suggest that the real mortality 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 mortality rate of 15%, then the estimated number of cases drops to 471,807 for France, but it becomes lower than the number of cases actually reported for China or South Korea, which is not consistent (Table III).

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.* 1,114,685 for France (Table III)) although it remains high compared to figures reported in databases.

Interestingly, since $C_{t0\text{-}estimated} = C_{(est\text{-}18d)} * P_{18d}$, this model takes into account the progression of the number of cases reported during the last eighteen days (P_{18d}). It can thus be used to model how policies to prevent disease spreading modulates P_{18d}, in order to understand the impact of these policies on the appearance of new cases. For example, for France on April 2, P_{18d} = 13.14, while on March 26, 2020, when regulation of progression rate based on population containment was still not fully efficient, P_{18d} = 25.89. We can therefore estimate that the containment made it possible to reduce P_{18d} from 25.89 to 13.14. The number of cases estimated on April 2 using P_{18d} = 25.89 would have been 6,974,156 cases. Thus, the containment from March 26 to April 2 prevented the appearance of 3,435,607 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 offer a simple method to estimate the number of people affected by Covid-19 at a time t_0 in a country, using the equation

$$C_{t0\text{-}estimated} = (M_{t0} / M_{r\text{-}est}) * (C_{t0} / C_{t0\text{-}18d})$$

With M_{t0}: number of deaths reported in a country at time t_0; $M_{r\text{-}est}$: estimated mortality rate; C_{t0}: number of cases reported in a country at time t_0; $C_{t0\text{-}18d}$: number of cases reported in the country at time t_0 minus 18 days. This analysis is based on four assumptions: 1. The number of deaths reported by each country is reliable, 2. The estimated mortality rate among people affected is known ($M_{r\text{-}est}$, here considered as 2%), 3. The average time between the onset of symptoms and death is known (here considered 18 days). 4. The time-dependent increase in the number of cases reported in databases during average time between the onset of symptoms and death reflects the time-dependent increase in the actual number of cases during the same time-period. Our analysis suggests that the number of Covid-19 cases in several country greatly exceeds the number of cases presented in international databases (3,538,549 versus 59,105 for France on April 2, 2020). These very high figures are consistent with those evaluated with another method by (Flaxman et al. 2020). For example, we report 7.4 million cases in Spain while Flaxman reports 7.0 million. 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 mortality 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 mortality rate. One of the limitations of our model is that mortality rates can change from one country to another, for example depending on the distribution of the population in different age groups that have different sensitivities to Covid-19.

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. 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

FIGURE AND TABLES

Country	Day (t_0)	Deaths at t_0 (M_{t0})	Cases at t_0 (C_{t0})	Mortality rate M_r	Cases at t_{0-18j} (Ct_{0-18d})	Mortality rate M_{r-18d}
Belgium	April 2	1 011	15 348	6.6%	886	114.1%
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 2	1 107	84 794	1.3%	5 795	19.1%
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 2	169	9 976	1.7%	8 162	2.1%
Netherlands	April 2	1 339	14 697	9.1%	1 135	118.0%
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: Mortality 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 2nd 2020). (https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series)

Country	Deaths at t_0 (April 2) (M_{t0})	Estimated cases at $t_{0\text{-}18d}$ $C_{(est\text{-}18d)}$	Reported cases at $t_{0\text{-}18d}$ ($Ct_{0\text{-}18d}$)	Reported cases at t_0 (Ct_0)	Progression rate from t_{-18} to t_0 P_{18d}	Estimated Cases (April 2) $C_{t0\text{-}estimated}$
Belgium	1 011	50 550	886	15 348	17.32	**875 667**
China	3 322	166 100	81 303	82 432	1.02	**169 030**
France	5 387	269 350	4 499	59 105	13.14	**3 538 549**
Germany	1 107	55 350	5 795	84 794	14.63	**809 896**
Iran	3 160	158 000	13 938	50 468	3.62	**572 101**
Italy	13 915	695 750	24 747	115 242	4.66	**3 239 973**
South-Korea	169	8 450	8 162	9 976	1.22	**10 328**
Netherlands	1 339	66 950	1 135	14 697	12.95	**866 929**
Spain	10 348	517 400	7 798	112 065	14.37	**7 435 552**
United Kingdom	2 921	146 050	1 140	33 718	29.58	**4 319 749**
USA	5 926	296 300	3 499	243 453	69.58	**20 615 926**

Table II: Estimation of the number of cases in different countries April 2nd (t_0) using a delay of 18 days between symptom occurrence and death and an estimated mortality rate ($M_{r\text{-}est}$) of 2% (note that rounding was not carried out during the calculation of $C_{t0\text{-}estimated}$ presented in the last column. This explains that the presented values are slightly different from the values of $C_{(est\text{-}18d)} * P_{18d}$).

Country	Population (million)	Reported cases at t_0 (April 2) (Ct_0)	Progression rate t_{-18} to t_0 P_{18d}	Estimated cases (April 2) $C_{t0\text{-estimated}}$	Estimated cases (April 2) $C_{t0\text{-estimated}}$	Estimated cases (April 2) $C_{t0\text{-estimated}}$	Progression rate t_{-12} to t_0 P_{12d}	Estimated cases (April 2) $C_{t0\text{-estimated}}$
Delay				$t_{0\text{-}18d}$	$t_{0\text{-}18d}$	$t_{0\text{-}18d}$		$t_{0\text{-}12d}$
$M_{r\text{-est}}$				**2%**	**4%**	**15%**		**2%**
Belgium	11,476	15 348	17.32	**875 667**	**437 834**	**116 756**	5.45	**275 610**
China	1,384,688	82 432	1.02	**169 030**	**84 515**	**22 537**	1.01	**168 402**
France	67,795	59 105	13.14	**3 538 549**	**1 769 275**	**471 807**	4.14	**1 114 685**
Germany	83,073	84 794	14.63	**809 896**	**404 948**	**107 986**	3.82	**211 288**
Iran	82,022	50 468	3.62	**572 101**	**286 051**	**76 280**	2.45	**386 897**
Italy	60,360	11 5242	4.66	**3 239 973**	**1 619 987**	**431 996**	2.15	**1 496 503**
South-Korea	51,709	9 976	1.22	**10 328**	**5 164**	**1 377**	1.13	**9580**
Netherlands	17,282	14 697	12.95	**866 929**	**433 465**	**115 591**	4.05	**270 990**
Spain	46,935	112 065	14.37	**7 435 552**	**3 717 776**	**991 407**	4.42	**2 285 112**
United Kingdom	65,761	33 718	29.58	**4 319 749**	**2 159 875**	**575 967**	6.72	**981 370**
USA	328,240	24 3453	69.58	**20 615 926**	**10 307 963**	**2 748 790**	9.55	**2 830 049**

Table III: Estimation of the number of cases in different countries April 2nd 2020 (t0) using estimated mortality rate (Mr-est) from 2 to 15% and delays between symptom occurrence and death of 18 and 12 days.

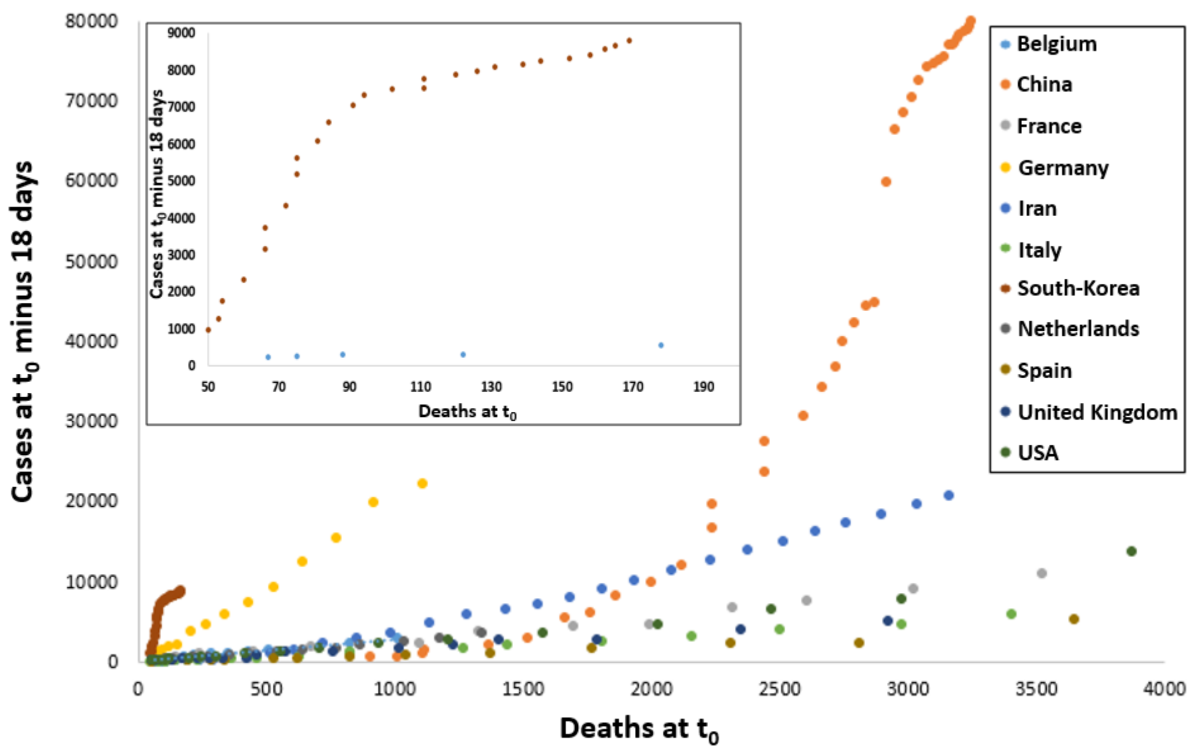

Figure 1: Relationships between deaths a given day (t_0) and the number of cases eighteen days before ($t_{0\text{-}18days}$) in different countries. The data include only figures between 50 and 4 000 deaths (or less if the number of deaths was lower in the country). The internal frame shows data from South Korea and Belgium that presented with less deaths than other countries.