Estimating the infection fatality ratio from COVID-19 and the impact of the lockdown in France

L Roques, BioSP, INRAE

With:
Etienne Klein, Julien Papaïx, Samuel Soubeyrand, BioSP INRAE, France
Antoine Sar, Medicentre & Research center, Moutier, Switzerland

Modeling the propagation of Covid-19
18, 19, 20 May 2020
Introduction
December 2019, a few people in France were affected by an illness whose symptoms were at the crossroads of those caused by influenza and pneumonia [Bondy's patient: retrospective PCR tests].

This low noise signal was not detected by the epidemiological surveillance mechanisms in France. China was aware of the emergence of this new disease, but no health alert was issued.

This first battle, that of early detection, was lost in France and abroad.
China, the first epicenter of the disease, gave the alert when the epidemic signal was already strong.

The WHO has been slow to assign symbolic pandemic status to the COVID-19 (11 March), while the pathogen was spreading all over the world.

In France, the basic reproduction number of the epidemic was significantly above 1 and many hospital structures were approaching the breaking point.

The second battle, that of the containment of the epidemic, was lost in metropolitan France, and the lockdown has been set on 17 March 2020.
A new round then began, with infectious individuals distributed heterogeneously over the territory.

The inertia of the epidemic dynamics was fully expressed in the Grand-Est and the Paris region where deaths accumulated despite the lockdown.

However, from April the circulation of the virus at the population level was strongly mitigated, indicating that the effective reproduction number was below the value 1.

The third battle was won: the slowdown in transmissions. Despite this slowdown, more than 25,000 deaths have been registered in France by early May.

...to be continued at the end of the talk
Part I: the early stage of the epidemic

Article

Using Early Data to Estimate the Actual Infection Fatality Ratio from COVID-19 in France

Lionel Roques 1,*, Etienne K Klein 1, Julien Papaïx 1, Antoine Sar 2 and Samuel Soubeyrand 1
Our goal

By mid-March, our objectives were:

- to estimate the fatality rate of COVID-19
- to estimate the basic reproduction number

Case fatality rate (ratio): \( \text{CFR} = \frac{\text{number of deaths}}{\text{number of confirmed cases}} \)

variable over time and across countries (4% in France at the beginning of the epidemic, 19% by early May)

Infection fatality rate (ratio): \( \text{IFR} = \frac{\text{number of deaths}}{\text{real number of cases}} \)

very informative but cannot be computed directly, stable under constant medical care conditions.
Lower and upper bounds for the IFR:

\[
\frac{\text{number of deaths}}{\text{total population}} \leq IFR \leq \frac{\text{number of deaths}}{\text{number of confirmed cases}} =: CFR
\]

Upper bounds: CFR in France (10 May)=19%, CFR in South Korea (late April)=2.3%

Lower bounds: deaths/pop in France (17 May)=0.04%, Lombardy (17 May)=0.15%

Means that values that are reported outside this range, for populations with the same age structure, are unrealistic (neglecting for genetic variations).
Mechanistic-statistical framework

Data
- Heterogeneous
- Noisy
- Not the process itself
- Confirmed cases
- Hospital deaths
- Number of tests

Epidemiological models
- Typically SIR models and their extensions
- Parameters $\Theta$, to be estimated

Describe a latent process (not observed)

$$(S, I, R) = F(\Theta)$$
Mechanistic-statistical framework

**Data**
- Heterogeneous
- Noisy
- Not the process itself
- Confirmed cases
- Hospital deaths
- Number of tests

**Epidemiological models**
- Typically SIR models and their extensions
- Parameter $\Theta$, to be estimated

Describe a latent process (not observed)

$$ (S, I, R) = F(\Theta) $$

How to bridge the gap between models and data?
# Mechanistic-statistical framework

<table>
<thead>
<tr>
<th>Data</th>
<th>Probabilistic observation model</th>
<th>Epidemiological models</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Heterogeneous</td>
<td>Data are random variables that depend on the latent process</td>
<td>- Typically SIR models and their extensions</td>
</tr>
<tr>
<td>- Noisy</td>
<td>Data $\sim P(S, I, R)$</td>
<td>- Parameter $\Theta$, to be estimated</td>
</tr>
<tr>
<td>- Not the process itself</td>
<td></td>
<td>Describe a latent process (not observed)</td>
</tr>
<tr>
<td>- Confirmed cases</td>
<td></td>
<td>$(S, I, R) = F(\Theta)$</td>
</tr>
<tr>
<td>- Hospital deaths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Number of tests</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Mechanistic-statistical framework

**Data**
- Heterogeneous
- Noisy
- Not the process itself
- Confirmed cases
- Hospital deaths
- Number of tests

**Probabilistic observation model**

Data are random variables that depend on the latent process

\[ \text{Data} \sim P(S, I, R) \]

**Epidemiological models**
- Typically SIR models and their extensions
- Parameter \( \Theta \), to be estimated

Describe a latent process (not observed)

\[ (S, I, R) = F(\Theta) \]

**Statistical inference**

Computation of the distribution \( P(\Theta|\text{Data}) \)
Data

The study was published (preprint) at the end of March: we only used data from the beginning of the epidemic, before the 17 March

- Number of positive cases and deaths in France, day by day (Johns Hopkins University Center for Systems Science and Engineering)

- Number of tests carried out (Santé Publique France)

Note: the official data on the number of deaths by COVID-19 in France only take into account hospitalised people (not EHPAD=nursing homes)
Epidemiological model (latent process)

Extremely parsimonious to avoid identifiability issues.

\[
\begin{align*}
S'(t) &= -\frac{\alpha}{N} S(t) I(t), \\
I'(t) &= \frac{\alpha}{N} S(t) I(t) - \beta I(t), \\
R'(t) &= \beta I(t),
\end{align*}
\]

\(N\): total population, \(\alpha\): contact rate (to be estimated), \(1/\beta\): mean duration of the infectious period (10 days, [He et al., Nature 2020])

\(S(t_0) = 67\) millions, \(I(t_0) = 1, R(t_0) = 0\).

\(t_0\): efficient date of introduction (to be estimated).

Impact of death on the SIR dynamics is neglected at this stage.
Probabilistic observation model

Number of cases tested positive on day \( t \): \( \hat{\delta}_t \)

Number of tests carried out: \( n_t \)

Tested population: \( n_t = \tau_1(t)I(t) + \tau_2(t)S(t) \)

**Important point:** \( \tau_1 \neq \tau_2 \), due to an important bias towards infected cases

Individual probability to get a positive test:

\[
p_t = \frac{\sigma \tau_1(t) I(t)}{\tau_1(t) I(t) + \tau_2(t) S(t)} = \frac{\sigma I(t)}{I(t) + \kappa_t S(t)},
\]

\( \sigma = \text{test sensitivity} \approx 0.7 \)  

[Wang et al, JAMA 2020]

\( \kappa_t \): relative probability to get tested for a \( S \) vs a \( I \) (assumed constant)
Probabilistic observation model

The number of cases tested positive on day $t$ $\delta_t$ then follows a binomial law conditionally on the latent $(S, I, R)$ process:

$$\delta_t \sim Bi(n_t, p_t).$$

Assuming that the daily number of confirmed cases $\delta_t$ are independent conditionally on the underlying SIR process, the likelihood is:

$$L(\alpha, t_0, \kappa) := P(\{\delta_t\} | \alpha, t_0, \kappa) = \prod_{t=t_i}^{t_f} \frac{n_t!}{(\delta_t)! (n_t - \delta_t)!} p_t^{\delta_t} (1 - p_t)^{n_t - \delta_t}. $$
Statistical inference

The posterior distribution of the parameters \((\alpha, t_0, \kappa)\) is computed with a Bayesian method, using uniform (non-informative) prior distributions:

\[
\alpha \in (0,1), \ t_0 \in (1,31), \ \kappa \in (0,1)
\]

The posterior distribution is the distribution of the parameters conditionally on the observations:

\[
P(\alpha, t_0, \kappa|\{\hat{\delta}_t\}) = \frac{\mathcal{L}(\alpha, t_0, \kappa) \pi(\alpha, t_0, \kappa)}{C},
\]

\(\pi(\alpha, t_0, \kappa)\) is the prior distribution (uniform), \(C\) normalization constant.
Statistical inference

Numerical computation of the posterior distribution: Metropolis-Hastings (MCMC) algorithm, using four independent chains, each of which with $10^6$ iterations
Pairwise posterior distributions

Joint posterior distribution of $(t_0, \alpha)$
Pairwise posterior distributions

Joint posterior distribution of \((t_0, \kappa)\)
Pairwise posterior distributions

Joint posterior distribution of \((\alpha, \kappa)\)
Model fit

Maximum posterior estimate \((\alpha^*, t^*_0, \kappa^*)\)

Corresponding probability of being tested positive

\[
p^*_s = \frac{\sigma I^*(s)}{I^*(s) + \kappa^* S^*(s)}
\]

Comparison between the expectation of the observation model \((\sum_{s=1}^{t} n_s p^*_s)\) and the data (cumulated confirmed cases)
Basic reproduction number

The basic reproduction number $R_0$ can be computed directly, based on the formula $R_0 = \alpha / \beta$

Using the distribution of $\alpha$, we get the following posterior distribution: $R_0 = 3.2$ (95% CI 3.1 – 3.3)
Infection Fatality Rate

We assume that the cumulated number of deaths satisfies:

\[ D'(t) = \gamma(t)I(t), \]

with \( \gamma(t) \) the rate at which the infectious die.

The distribution of \( \gamma(t) \), at each time \( t \) is computed as the ratio between the data (smoothed \( D'(t) \)) and the posterior estimate of \( I(t) \).

Then:

\[ IFR_t := \frac{\gamma(t)}{\gamma(t) + \beta}. \]

It is the fraction of ‘removed’ that die.

On March 17, the IFR is 0.5\% (95\% CI: 0.3\% – 0.8\%)
Taking nursing homes into account

The IFR is computed from hospital death counts.

On 31 March data from the Grand Est Region (Source: ARS Grand Est) indicate that this value should be multiplied by 1.6 to take into account the deaths in nursing homes (EHPAD).

Taking this into account, we get that on March 17:

the IFR is 0.8% (95% – CI: 0.45% – 1.25%)
We get an average factor x8 during the period 29 February-17 March
Comparison with other studies

- IFR consistent with the lower and upper bounds (0.15% in Lombardy, 2.3% in South Korea)
- IFR consistent with Chinese data: 0.66% [Verity et al., Lancet infectious diseases]
- Data from the Diamond Princess: 1.3% [Russell et al., Eurosurveillance]
- New York large scale serological study of late April found an IFR of 0.6%
- Study from Pasteur Institute found an IFR of 0.7%, using data until one month after our study (and $R_0 = 3.3$)

→ Our method was (presumably) efficient to assess the severity of the disease at a very early stage of the epidemic
Part II: the lockdown

Impact of lockdown on the epidemic dynamics of COVID-19 in France

Lionel Roques\textsuperscript{1,*}, Etienne K. Klein\textsuperscript{1}, Julien Papaïx\textsuperscript{1}, Antoine Sar\textsuperscript{2} and Samuel Soubeyrand\textsuperscript{1}
Data

Same type of data as in the first part of the talk, but over the period ranging from 31 March to 14 April.

Epidemiological model (latent process)

\[
\begin{align*}
S'(t) &= -\frac{\alpha}{N} S(t) I(t), \\
I'(t) &= \frac{\alpha}{N} S(t) I(t) - (\beta + \gamma) I(t), \\
R'(t) &= \beta I(t), \\
D'(t) &= \gamma I(t).
\end{align*}
\]

New difficulty: the initial conditions are not known.
Initial conditions

\[ S_0 = \text{number of susceptible cases at the end of March} \ 66 \cdot 10^6 \]
(sensitivity analysis: not much effect of this assumption)

\[ D_0: \text{known (cumulated number of deaths by March 31)} \]

\[ R_0: \text{not known, but no effect on the S, I, D compartments} \]

\[ I_0: \text{will be estimated} \]

Probabilistic observation model

For the number of confirmed cases: same model as in the first part of the talk.

For the daily number of deaths: new model.
Probabilistic observation model, number of deaths

New observed deaths (excluding nursing homes): $\hat{\mu}_t$

Modelled by independent Poisson distributions conditionally on $D(t)$, with mean value $D(t) - D(t - 1)$:

$$\hat{\mu}_t \sim \text{Poisson}(D(t) - D(t - 1)).$$

Unknown parameters: $(\alpha, \gamma, \kappa, I_0)$

Assuming that the increments $\delta_t$, $\hat{\mu}_t$ are independent conditionally on the underlying SIRD process, the likelihood is:

$$
\mathcal{L}(\alpha, \gamma, \kappa, I_0) := P(\{\delta_t, \hat{\mu}_t\}|\alpha, \gamma, \kappa, I_0) = P(\{\delta_t\}|\alpha, \gamma, \kappa, I_0) P(\{\hat{\mu}_t\}|\alpha, \gamma, \kappa, I_0)
= \prod_{t=t_i}^{t_f} \frac{n_t!}{(\delta_t)!(n_t - \delta_t)!} p_t^{\delta_t} (1 - p_t)^{n_t - \delta_t} \prod_{t=t_i}^{t_f} e^{-D(t) + D(t - 1)} \frac{(D(t) - D(t - 1)\hat{\mu}_t}{\hat{\mu}_t!},
$$
Statistical inference

The posterior distribution of the parameters \((\alpha, \gamma, \kappa, I_0)\) is computed with a Bayesian method, using uniform (non-informative) prior distributions for \((\alpha, \kappa, I_0)\):

\[
\alpha \in (0,1), \kappa \in (0,1), I_0 \in (1,10^7)
\]

and an informative prior distribution for the death rate \(\gamma\), corresponding to the estimation obtained in the first part of the talk.

Finally the posterior distribution of the parameters is:

\[
P(\alpha, \gamma, \kappa, I_0 | \{\hat{\delta}_t, \hat{\mu}_t\}) = \frac{\mathcal{L}(\alpha, \gamma, \kappa, I_0) \pi(\alpha, \gamma, \kappa, I_0)}{C},
\]

with:

\[
\pi(\alpha, \gamma, \kappa, I_0) = 1_{(\alpha, \kappa, I_0) \in (0,1) \times (0,1) \times (1,10^7)} f_g(\gamma).
\]
Model fit

Maximum posterior estimate ($\alpha^*, \kappa^*, \gamma^*, I_0^*$). We compare the expected observation associated with this estimate and the data.
Effective reproduction number

Computed as $\frac{\alpha}{(\beta + \gamma)} \approx \frac{\alpha}{\beta}$

Using the distribution of $\alpha$ we get the following posterior distribution for $R_e$: $R_e = 0.47$ (95% CI 0.45 – 0.50)

The reproduction number has been divided by 7 compared to the early stage of the epidemic.
- About 3.7% of the population has been infected by 10 May (95% CI: 3%-4.7%), neglecting the $R$ (immune cases) before April 1st.
- We find a factor $\times 18$ between the confirmed cases and actual (estimated) number of cases.
- High inertia of the epidemic: the number of $I$ is still high at the end of the lockdown $\approx 160\,000$ cases.
Comparison with other studies

- Contact surveys data for Wuhan and Shanghai have shown that the reproduction number was divided by a factor 7 in Wuhan and 11.5 in Shanghai during the lockdown [Zhang et al. medRxiv].

- The study of Institut Pasteur [Salje et al., Science, 13 May] estimated that 4.4% of the population has been infected by 10 May.

- The same study shows a reduction by a factor 5 in the reproduction number.
The true (?) story of COVID-19 in France: Chapter 4

The fourth round begins now, with the following initial data:

• an order of magnitude of 100,000 contagious individuals
• 95% of the population still susceptible

Thus, we face a significant contagion potential and we are very far from the herd immunity (about 70% of immune individuals)
The true (?) story of COVID-19 in France: Chapter 4

Below an estimate of $R_e(t)$, computed over the previous 15 days (moving window $(t - 15, t)$):

![Graph showing the effective reproduction rate $R_e(t)$ over time, with a line indicating the value 1 and a natural value around 3.]

It is therefore essential to maintain an effective reproduction rate below the value 1 or very slightly above. It will be a difficult task, as the ‘natural’ value is around 3.
Factors that can help
• barriers to the spread of the pathogen due to ‘local’ group immunity, at the scale of a residence, a nursing home or even a city (for example, serological tests showed an attack rate of 41% at Lycée Jean-Monnet in Crépy-en-Valois).

As shown by Marino Gatto this morning, spatial aspects are very important.

*Ongoing work with O Bonnefon, H Berestycki, V Baudrot and S Soubeyrand, on the spread of immunity*
The true (?) story of COVID-19 in France: Chapter 4

Factors that can help

• Seasonality effects? [Demongeot et al. MDPI Biology, May]

But there may be several confounding variables (e.g., age structure)
Thank you for your attention