



HAL
open science

Estimation du nombre de reproduction de l'épidémie de COVID-19 en France

Gonché Danesh, Samuel Alizon

► **To cite this version:**

Gonché Danesh, Samuel Alizon. Estimation du nombre de reproduction de l'épidémie de COVID-19 en France. [Rapport de recherche] 1, Centre national de la recherche scientifique (CNRS); Institut de recherche pour le développement (IRD); Université de Montpellier (UM), FRA. 2020. hal-02882680

HAL Id: hal-02882680

<https://hal.science/hal-02882680>

Submitted on 9 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Préambule

Contexte

Nombre de reproduction de base

Méthode

Résultats

Discussion

Sources et remerciements

Rapport : Estimation du nombre de reproduction de l'épidémie de COVID-19 en France

Groupe de modélisation de l'équipe ETE
(<https://www.mivegec.ird.fr/fr/contact/160-francais/equipes/1205-ete>)
(Laboratoire MIVEGEC (<http://mivegec.ird.fr>), CNRS, IRD, Université de Montpellier)

14 mars 2020 (mis à jour le 01 avril 2020)

Pour des estimations plus fiables du nombre de reproduction, voir notre nouveau Rapport 5 (Rapport5_R.html)

Préambule

Le contenu de ce rapport n'est en aucun cas officiel et est avant tout à destination pédagogique. De plus, les calculs réalisés ici sont faits avec des hypothèses volontairement simplifiées.

En matière de santé publique et pour toute question, nous recommandons de consulter et suivre les instructions officielles disponibles sur <https://www.gouvernement.fr/info-coronavirus> (<https://www.gouvernement.fr/info-coronavirus>)

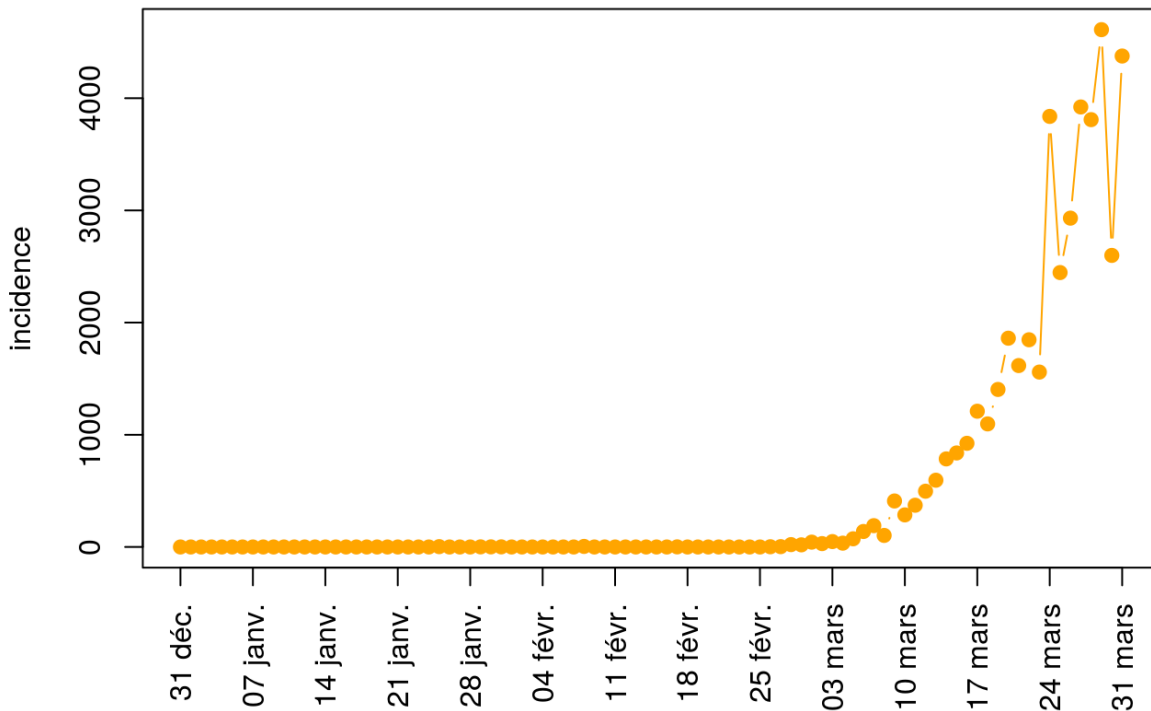
L'Organisation Mondiale de la Santé (OMS) dispose aussi d'un site très complet <https://www.who.int/fr/emergencies/diseases/novel-coronavirus-2019> (<https://www.who.int/fr/emergencies/diseases/novel-coronavirus-2019>)

Contexte

Depuis décembre 2019 l'épidémie de COVID-19 causée par un coronavirus jusqu'alors inconnu (SRAS-Cov-19) se propage. Initialement présente dans la région de Wuhan en Chine, elle a officiellement été déclarée comme pandémie le 11 mars 2020 par l'Organisation Mondiale de la Santé (OMS).

En France, des premiers cas isolés ont été détectés en janvier 2020 auprès de voyageurs revenant de Chine. Depuis le 25 février 2020, l'incidence (nombre de nouveaux cas par jour) est strictement positive et croissante, suggérant une transmission locale du virus. Le 11 mars, 2281 cas avaient été recensés. Le 12 mars, l'État français a annoncé la fermeture des structures scolaires à partir du lundi 16 mars.

Nombre de nouveaux cas journaliers de COVID-19 en France



Les données d'incidence dans la figure ci-dessus ne sont pas estimées mais proviennent des rapports de l'OMS.

Nombre de reproduction de base

À l'aide de méthodes statistiques, nous avons estimé le **nombre de reproduction de base** (noté R_0) de l'épidémie en France, c'est-à-dire le nombre d'infections secondaires engendrées par une personne infectée au cours de toute sa période infectieuse.

Ce nombre est clé en santé publique car il détermine l'ampleur de l'épidémie. Si le R_0 passe sous le seuil de 1, soit du fait des mesures de contrôle, soit parce qu'une proportion suffisante de la population est immunisée, l'épidémie décroît.

Ce nombre d'infections secondaires a tendance à diminuer en cours d'épidémie. D'une part car de moins en moins de personnes sont sensibles (une fois guéries, une **mémoire immunitaire** se met en place qui protège contre les ré-infections). D'autre part, car des mesures sanitaires (notamment la **limitation des contacts**) sont implémentées. À une date t on peut alors mesurer un nombre *instantané* (ou *efficace*) que l'on notera ici $R(t)$.

Méthode

Cette section est plus technique et vise à expliciter les hypothèses de travail réalisées.

Pour estimer ces valeurs, nous avons utilisé le nombre de nouveaux cas rapportés chaque semaine en France par l'OMS (données récupérées sur le site <https://ourworldindata.org/coronavirus-source-data> (<https://ourworldindata.org/coronavirus-source-data>)), ainsi que les données brutes de 28 paires infecteur/infecté compilées dans un article de Nishiura *et alii* (2020, *Int J Infect Dis*). Enfin, nous avons estimé le R_0 à partir de plusieurs dates de début d'épidémie, afin de déterminer dans quelle mesure celui-ci était impacté par les cas importés (majoritaires en début d'épidémie).

Les estimations ont été faites à l'aide du logiciel R (cran.r-project.org/) et des fonctions *est.R0.TD()* et *est.R0.ML()* du package R0 (<https://cran.r-project.org/web/packages/R0/>).

Les hypothèses sous-jacentes pour le calcul du R_0 sont les suivantes :

- le temps de signalement des cas est négligé,
- la stratégie de dépistage en France est supposée constante,
- la structure spatiale est négligée,
- les cas importés ne sont pas distingués des cas non-importés,
- les incidences utilisées sont celles disponibles depuis le 21 janvier 2020 et jusqu'au 16 mars 2020 (date de la mise en place des mesures de confinement) pour $R_0(t)$ et au 01 avril 2020 pour $R(t)$.

Résultats

R_0

Nous avons estimé un R_0 classique en fixant arbitrairement une date de début d'installation de l'épidémie en France (donc le moment à partir duquel les cas importés ont peu d'influence).

Dans ce modèle simpliste, le R_0 est supposé constant au cours du temps (on néglige donc la baisse du nombre de personnes sensibles à mesure que l'épidémie avance ainsi que les politiques de prévention).

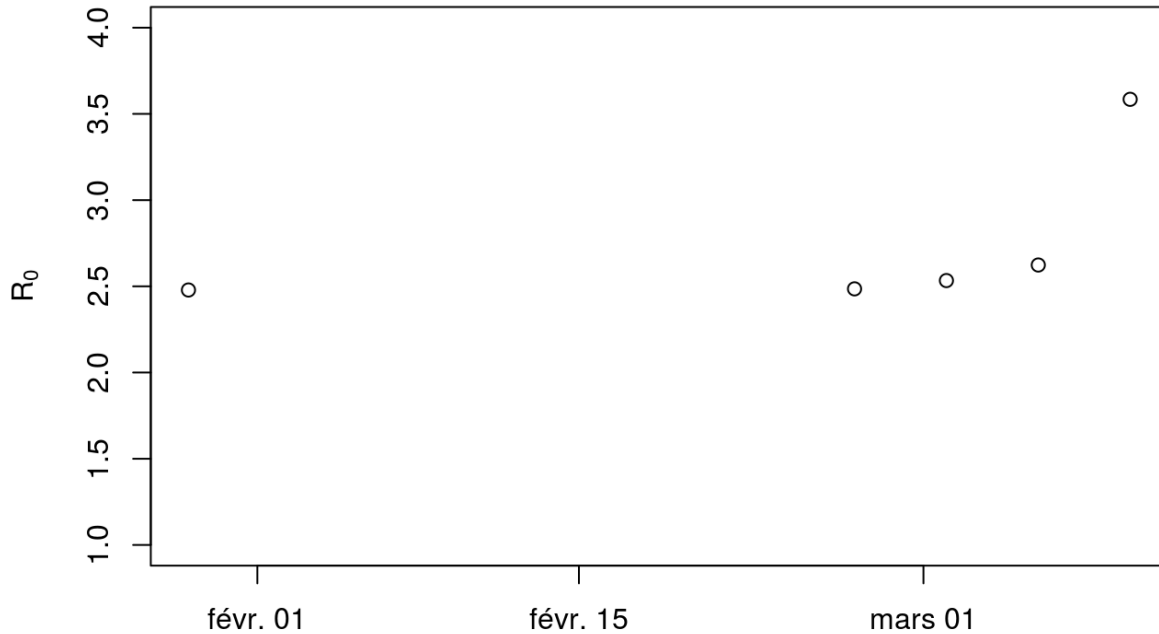
Le résultat de la fonction *est.R0.ML* est le suivant pour des données intégrées entre le 27 février (date du début de la courbe d'incidenc) et le 16 mars (date de début du confinement) :

```
## [1] "R0 = 2.49 [2.39 ; 2.58]"
```

Il semblerait donc que l'épidémie se soit propagée très rapidement en France puisqu'une personne infectée en infectait en moyenne plus de 2,5.

Cette estimation de R_0 dépend de la date à laquelle l'épidémie s'est installée en France (c'est-à-dire ne dépendait plus des cas importés) ainsi que des données disponibles à ce jour. Le graphique ci-dessous indique la valeur du R_0 selon cette date d'origine.

R[0] selon la date d'origine de l'épidémie



En revanche, la date de fin d'intervalle, elle, joue plus. En effet, si au lieu d'arrêter notre jeu de données 16 mars nous l'arrêtons au 9 mars, tout en commençant au 27 février, alors le résultat de la fonction *est.R0.ML* est le suivant :

```
## [1] "R0 = 3.81 [3.5 ; 4.13]"
```

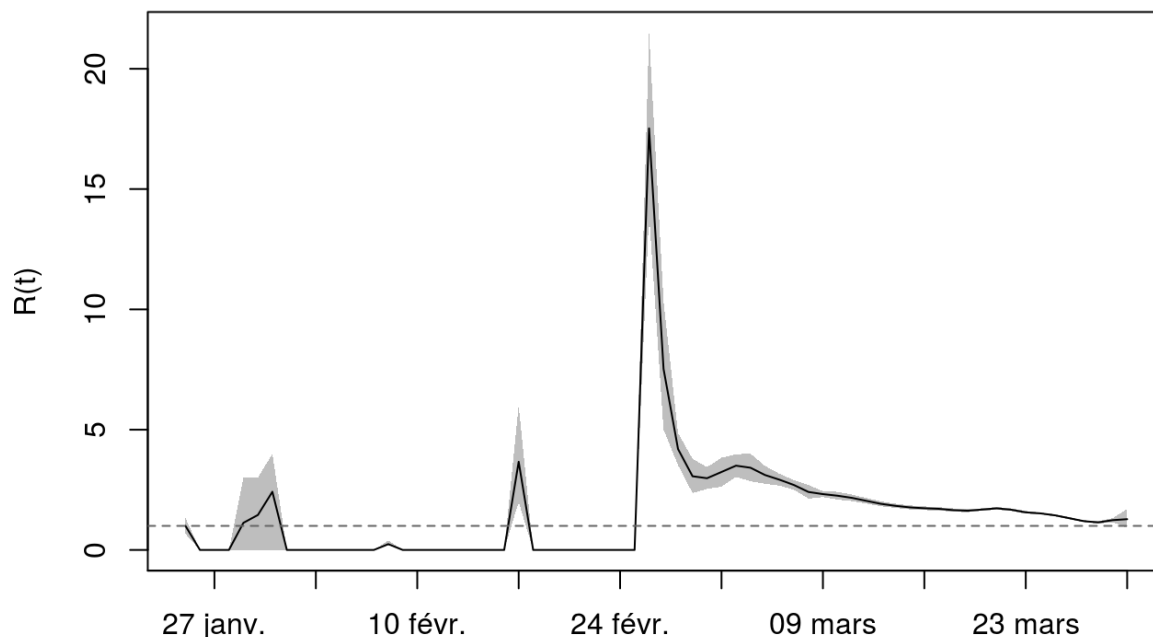
On voit que les valeurs de R_0 dépendent peu de la date d'origine choisie pour le calcul (entre le 29 janvier et le 27 février) mais qu'en revanche elles dépendent de la date de fin d'intervalle. En effet, plus l'intervalle est grand, plus le R_0 est faible. Ceci fait sens car plus l'épidémie avance, plus des mesures sont mises en place pour lutter contre l'épidémie et plus l'immunité se construit. Il semble donc intéressant de calculer les variations du nombre de reproduction.

Nombre de reproduction temporel (ou efficace)

Comme indiqué ci-dessus, le nombre de reproduction de base varie du fait de l'immunité qui se construit dans la population et des mesures de santé publique. Nous avons donc estimé le nombre de reproduction efficace ($R(t)$) afin de détecter ces variations temporelles.

Sur le graphique suivant, la zone grisée correspond à l'intervalle de confiance et la courbe noire à la médiane. Attention, ce n'est pas le nombre de cas qui est représenté mais plutôt la rapidité de propagation de l'épidémie. Si l'on passe sous la barre de 1 (matérialisée avec des tirets), l'épidémie est en phase de décroissance.

Nombre de reproduction efficace en France



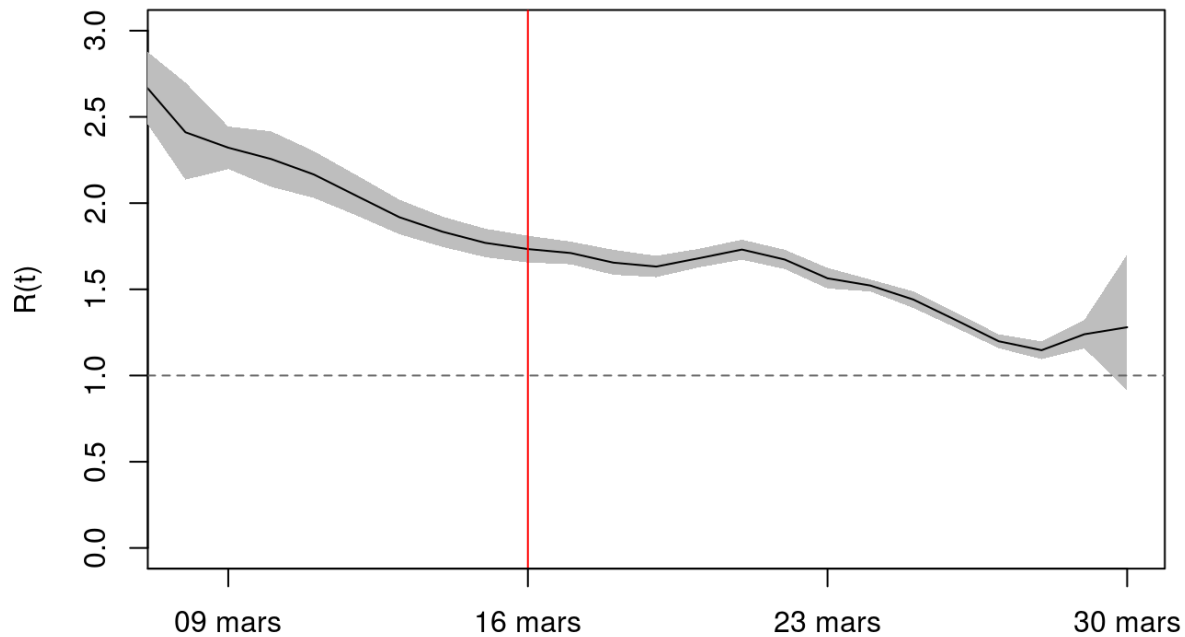
On peut observer des pics initiaux de $R(t)$ correspondant à des clusters de transmission initiés par des cas importés et qui ont été contrôlés. Ces pics sont avant tout dus au nombre limité de dates et peuvent être ignorés : avec des incidences hebdomadaires au lieu d'incidences journalières, ces courbes seraient bien plus lisses.

À partir du 28 février, l'épidémie semble s'installer en France (le pic de valeur de $R(t)$ est non biologique et lié à ce début). Depuis cette date, la valeur du nombre de reproduction de base semble s'être stabilisée sur une valeur supérieure à 2, en cohérence avec notre estimation de R_0 .

Situation avec confinement

Depuis le 16 mars 2020, le gouvernement français a mis en place un confinement à l'échelle nationale. Si on se concentre sur une période de temps plus récente, voici à quoi ressemble le graphique (la ligne rouge est le début du confinement et la zone grisée l'intervalle de confiance à 95 %).

Nombre de reproduction efficace en France



On voit donc que l'effet du confinement met du temps à se manifester sur les données, ce qui est en partie logique (il faut environ 5 jours pour que les symptômes se manifestent, ce qui crée un décalage). Toutefois, le délai à passer sous le seuil de 1, et donc à atteindre le pic épidémique, reflète aussi probablement un délai à la baisse des contacts dans la population.

Attention, cette estimation du nombre de reproduction a un délai potentiellement important car il faut au moins 4 jours pour que les symptômes apparaissent chez une personne infectée, plus un délai pour que cette personne soit testée et, enfin, que les résultats soient enregistrés. Au mieux, la valeur la plus récente reflète donc l'état de l'épidémie la semaine dernière.

Cette valeur la plus récente du taux de reproduction de base est la suivante :

```
## [1] "R(t) = 1.28 avec 95 % des valeurs entre 0.91 et 1.7."
```

Si la borne supérieure de l'intervalle de confiance de $R(t)$ est strictement inférieur à 1, on peut dire avec 97,5 % de certitude que, selon nos hypothèses et les données utilisées, l'épidémie est en phase de décroissance.

Attention, en plus des délais, cette estimation du nombre de reproduction est aussi très sensible à des variations en termes d'intensité d'échantillonnage (si on teste moins, on détecte moins de cas et $R(t)$ descend mécaniquement).

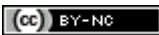
Discussion

En analysant les données d'incidence de France et en les couplant à des données existantes récoltées sur des chaînes de transmission ailleurs dans le monde, nous avons estimé que le nombre de reproduction de base (R_0) initial de l'épidémie de COVID-19 était de l'ordre 2,5. À titre de comparaison, celui de la grippe saisonnière est généralement proche de 1,5. Cette forte valeur s'explique vraisemblablement par le fait que, contrairement à la grippe, il n'y a aucune immunité pré-existante dans la population.

De plus, une analyse des variations temporelles de ce nombre de reproduction indique que ce dernier ($R(t)$) s'est stabilisé autour de valeurs supérieures à 2 au début du mois de mars. Les mesures de confinement en France ont ensuite fait baisser ce R efficace si l'on fait l'hypothèse que l'effort de dépistage est demeuré constant (en dépistant moins, on fait aussi baisser le R).

Ces valeurs obtenues sont qualitativement cohérentes avec celles obtenues par Abbott *et alii* (<https://cmmid.github.io/topics/covid19/current-patterns-transmission/global-time-varying-transmission.html>). Il faut toutefois noter qu'ils utilisent un autre logiciel pour leur inférence (EpiEstim) mais les bases mathématiques et statistiques sont similaires. Leur analyse est aussi plus fine car elle inclut un délai dans le signalement des cas, ainsi que les cas importés. En revanche, ils ont imposé une distribution pour l'intervalle sériel (délai entre le début symptômes chez deux individus, l'un-e ayant infecté l'autre), alors que nous avons utilisé une distribution empirique. Enfin, comme le signale le site Our world in data (<https://ourworldindata.org/coronavirus-source-data>), il existe des incohérences mineurs entre différentes séries temporelles de source officielle.

Sources et remerciements

- L'équipe de modélisation de l'équipe ETE est composée de Samuel Alizon, Thomas Bénéteau, Marc Choisy, Gonché Danesh, Ramsès Djidjou-Demasse, Baptiste Elie, Yannis Michalakis, Bastien Reyné, Quentin Richard, Christian Selinger, Mircea T. Sofonea.
- Les données d'incidence proviennent du site <https://ourworldindata.org/coronavirus-source-data> (<https://ourworldindata.org/coronavirus-source-data>)
- Les données de l'intervalle sériel du site <https://github.com/aakhmetz/nCoVSerialInterval2020> (<https://github.com/aakhmetz/nCoVSerialInterval2020>)
- L'estimation de $R(t)$ et R_0 est faite grâce au package de Obadia, T., Haneef, R. & Boëlle, P. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Med Inform Decis Mak* 12, 147 (2012). <https://doi.org/10.1186/1472-6947-12-147> (<https://doi.org/10.1186/1472-6947-12-147>)
- Contribution à ce travail :
 - conception du travail : ensemble de l'équipe
 - compilation des données et calculs de R_0 et $R(t)$: GD
 - rédaction du rapport : SA
 - validation du rapport : ensemble de l'équipe
- contact : samuel.alizon@cnrs.fr (<mailto:samuel.alizon@cnrs.fr>)
-  (<http://creativecommons.org/licenses/by-nc/4.0/>)
Ce(tte) œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International (<http://creativecommons.org/licenses/by-nc/4.0/>).