

# Impact du confinement sur la dynamique épidémique du COVID-19 en France

Front. Med., 05/06/2020

Lionel Roques, Etienne K. Klein, Julien Papaïx, Antoine Sar, Samuel Soubeyrand

(traduction post-éditée par N. Bacaër, suggestions d'amélioration : nicolas.bacaer@ird.fr)



## Résumé

L'épidémie de COVID-19 a été signalée dans la province du Hubei en Chine en décembre 2019 puis s'est propagée dans le monde entier pour atteindre le stade de la pandémie au début de mars 2020. Depuis lors, plusieurs pays sont entrés en confinement. À l'aide d'un formalisme mécaniste-statistique, nous estimons l'effet du confinement en France sur le taux de contact et le nombre effectif de reproduction  $R_e$  du COVID-19. On obtient une réduction d'un facteur 7 ( $R_e = 0,47$ , IC 95% : 0,45-0,50), par rapport aux estimations réalisées en France au stade précoce de l'épidémie. Nous estimons également la fraction de la population qui serait infectée au début du mois de mai, à la date officielle à laquelle le confinement devrait être assoupli. On retrouve une fraction de 3,7% (IC 95%: 3,0–4,8%) de la population totale française, sans tenir compte du nombre d'individus guéris avant le 1er avril, qui n'est pas connu. Cette proportion est apparemment trop faible pour atteindre l'immunité collective. Ainsi, même si le confinement a fortement atténué la première vague épidémique, le maintien d'une valeur faible de  $R_e$  est crucial pour éviter une deuxième vague incontrôlée (initiée avec beaucoup plus de cas infectieux que la première vague) et ainsi éviter la saturation des installations hospitalières.

## 1. Introduction

L'épidémie de COVID-19 a été signalée dans la province du Hubei en Chine en décembre 2019, puis s'est propagée dans le monde entier pour atteindre le stade de la pandémie au début de mars 2020 (1). Pour ralentir l'épidémie, plusieurs pays se sont fermés avec différents niveaux de restrictions. Dans la province du Hubei, où le confinement a été établi bien avant les autres pays (le 23 janvier), l'épidémie a atteint un plateau, avec seulement de nouveaux cas sporadiques au 15 avril [d'après les données du Johns Hopkins University Center for Systems Science and Ingénierie (2)]. En France, les premiers cas de COVID-19 ont été détectés le 24 janvier, et le confinement a été fixé le 17 mars. Ce confinement national se traduit par d'importantes restrictions de mouvement, avec un confinement obligatoire à domicile sauf pour les trajets essentiels, y compris les achats de nourriture, les soins, 1 h d'activité sportive individuelle et de travail lorsque le télétravail n'est pas possible, et fermeture des frontières de l'espace Schengen. Cela comprend également les fermetures d'écoles et d'universités ainsi que de tous les lieux publics non essentiels, y compris les magasins (à l'exception des magasins d'alimentation), les restaurants, les cafés, les cinémas et les discothèques.

Le nombre de reproduction de base  $R_0$  correspond au nombre attendu de nouveaux cas générés par un seul cas infectieux dans une population pleinement sensible (3). Plusieurs études, principalement basées sur des données chinoises, visaient à estimer le  $R_0$  associé à l'épidémie de COVID-19, conduisant à des valeurs de 1,4 à 6,49, avec une moyenne de 3,28 (4). Comme la valeur de  $R_0$  peut être interprétée comme le produit du taux de contact et de la durée de la période infectieuse, et comme l'objectif du confinement et des stratégies de restriction associées est précisément de diminuer le taux de contact, un effet important sur le nombre  $R_e$  des cas secondaires générés par un individu infectieux est à prévoir. Cette valeur  $R_e$  est souvent appelée «nombre effectif de reproduction» et correspond à la contrepartie de  $R_0$  dans une population qui n'est pas totalement sensible (5). Si  $R_e > 1$ , le nombre de cas infectieux dans la population suit une tendance à la hausse, et plus  $R_e$  est grand, plus cette tendance est rapide. Au contraire, si  $R_e < 1$ , l'épidémie s'éteindra progressivement. Il a été démontré que les mesures de contrôle en Chine ont un effet significatif sur l'épidémie de COVID-19, avec des taux de croissance qui sont passés de valeurs positives à négatives (correspondant à  $R_e < 1$ ) dans les 2 semaines (6). L'étude (7) a montré que les politiques de confinement dans la province du Hubei ont également conduit à une croissance sous-exponentielle du nombre de cas, compatible avec une diminution du nombre effectif de reproduction  $R_e$ . En adaptant un modèle épidémique SEIR à des séries chronologiques de cas signalés dans 31 provinces de Chine, Tian et al. (8) ont trouvé un indice de reproduction de base  $R_0 = 3,15$  avant la mise en œuvre de l'intervention d'urgence en Chine, une valeur qui a été divisée par plus de 20 une fois que les mesures de contrôle étaient pleinement efficaces. En utilisant les données des enquêtes de contact pour Wuhan et Shanghai, il a été estimé dans Zhang et al. (9) que le nombre effectif de reproduction a été divisé par un facteur 7 à Wuhan et 11,5 à Shanghai.

Les modèles épidémiologiques standards reposent généralement sur les systèmes SIR (Susceptible-Infected-Removed) d'équations différentielles ordinaires et leurs extensions [pour des exemples d'application à l'épidémie de COVID-19, voir (10, 11)]. Avec ces modèles, et plus généralement pour la plupart des modèles déterministes basés sur des équations différentielles, lorsque la perte d'information due au processus d'observation est lourde, des approches spécifiques doivent être utilisées pour combler le fossé entre les modèles et les données. L'une de ces approches est basée sur le formalisme mécaniste-statistique, qui utilise un modèle probabiliste pour relier le processus de collecte de données et la variable latente décrite par le modèle EDO. Des articles et des manuels sur les jalons ont été rédigés sur cette approche ou des approches connexes (12), qui devient la norme en écologie (13, 14). L'application de cette approche aux données épidémiologiques humaines est encore rare.

Dans une précédente étude (15), nous avons appliqué ce cadre aux données correspondant au début de l'épidémie en France (du 29 février au 17 mars), avec un modèle SIR. Notre objectif principal était d'évaluer le taux de mortalité par infection (IFR), défini comme le nombre de décès divisé par le nombre de cas infectés. Le nombre de personnes infectées n'étant pas connu, cette quantité ne peut être mesurée directement, même maintenant (au 15 avril). Le cadre mécanistico-statistique nous a permis de calculer un IFR de 0,8% (IC à 95%: 0,45-1,25%), ce qui était cohérent avec les résultats précédents en Chine (0,66%) et au Royaume-Uni (0,9%) (16) et inférieure à la valeur précédemment calculée sur les données du navire Diamond Princess (1,3%) (17). Dans cette étude précédente, nous avons également calculé le  $R_0$  en France, et nous avons trouvé une valeur de 3,2 (95% -CI: 3,1–3,3). Bien que le nombre de tests à ce stade soit faible, l'avantage de travailler avec les données du début de l'épidémie est que l'état initial de l'épidémie est connu.

Ici, nous développons une nouvelle approche mécanistico-statistique, basée sur un modèle SIRD (D étant le compartiment des morts), avec comme objectif

- l'estimation de l'effet du confinement en France sur le taux de contact et le nombre effectif de reproduction  $R_e$  ;
- l'estimation du nombre d'individus infectieux et de la fraction de la population infectée au début du mois de mai (à la date officielle à laquelle le confinement devrait être assoupli).

## 2. Matériels et méthodes

### 2.1. Les données

Nous avons obtenu le nombre de cas positifs et de décès en France, jour par jour auprès de Santé Publique France (18), du 31 mars au 14 avril. Nous avons obtenu des données hebdomadaires sur le nombre d'individus testés (dans les laboratoires privés et les hôpitaux) de la même source. Nous avons supposé que pendant chacune de ces semaines, le nombre de tests par jour était constant. Cette hypothèse est cohérente avec les faibles variations entre le nombre de tests durant la première semaine (111 690) et la deuxième semaine d'observation (132 392). Les données sur le nombre de cas positifs n'étant pas totalement fiables (moins de cas le week-end avec rebond le lundi), nous avons lissé les données avec une moyenne mobile sur 5 jours. Les données officielles sur le nombre de décès par COVID-19 depuis le début de l'épidémie en France ne prennent en compte que les personnes hospitalisées. Environ 728 000 personnes en France vivent dans des maisons de retraite médicalisées [EHPAD, source: DREES (19)]. Le nombre de décès dans ces structures n'a été signalé que récemment et ne peut être obtenu au jour le jour. Les dernières données de Santé Publique France indiquent un nombre total de  $10^{6,43}$  décès à l'hôpital et 6 524 décès en maison de retraite au 15 avril. Le nombre total de décès correspond donc à environ 1,6 fois le nombre de décès à l'hôpital. Le même facteur avait été estimé dans Roques et al. (15) sur la base d'un ensemble de données locales de la région du Grand Est français.

## 2.2. Cadre mécanistique-statistique

Le cadre mécaniste-statistique consiste en la combinaison d'un modèle mécaniste qui décrit le processus épidémiologique, d'un modèle d'observation probabiliste et d'une procédure d'inférence.

### 2.2.1. Modèle mécaniste

La dynamique de l'épidémie est décrite par le modèle compartimental SIRD suivant:

$$S'(t) = -\frac{\alpha}{N}S(t)I(t), \quad I'(t) = \frac{\alpha}{N}S(t)I(t) - (\beta + \gamma)I(t), \quad R'(t) = \beta I(t), \quad D'(t) = \gamma I(t), \quad (1)$$

avec  $S$  la population sensible,  $I$  la population infectieuse,  $R$  la population rétablie,  $D$  le nombre de décès dus à l'épidémie et  $N$  la population totale. Par souci de simplicité, nous supposons que  $N$  est constant, égal à la population française actuelle, négligeant ainsi l'effet des faibles variations de la population sur le coefficient  $\alpha / N$ . Le paramètre  $\alpha$  est le taux de contact (à estimer) et  $1 / \beta$  est le temps moyen jusqu'à ce qu'un infectieux se rétablisse. Sur la base des résultats de Zhou et al. (20), la période médiane d'excrétion virale est de 20 jours, mais l'infectiosité a tendance à se décomposer avant la fin de cette période: les résultats de He et al. (21) indiquent que l'infectiosité débute 2 à 3 jours avant l'apparition des symptômes et diminue significativement 8 jours après l'apparition des symptômes. Sur la base de ces observations, nous supposons ici que la durée moyenne de la période d'infectiosité est de  $1 / \beta = 10$  jours. Dans Li et al. (22), la durée de la période d'incubation a été estimée en moyenne à 5,2 jours. Ainsi, la durée moyenne de la période d'exposition non infectieuse est relativement courte (environ 2 à 3 jours) et peut être négligée sans trop de différences sur les résultats, comme le montrent Liu et al. (23). L'inclusion d'un compartiment exposé (comme dans les modèles SEIR) est particulièrement pertinente lorsque les individus exposés peuvent transmettre indirectement la maladie, par exemple par le biais d'insectes vecteurs [par exemple, (24)], ce qui n'est apparemment pas le cas pour les coronavirus. Le paramètre  $\gamma$  correspond au taux de mortalité des infectieux (à estimer).

#### 2.2.1.1. Conditions initiales

Le modèle démarre à une date  $t_0$  correspondant au 1er avril. Le nombre initial  $I(t_0) = I_0$  d'infectieux n'est pas connu et sera estimé. Le nombre total de guéris au temps  $t_0$  n'est pas non plus connu. Cependant, le compartiment  $R$  n'ayant pas de retour sur les autres compartiments, on peut supposer sans perte de généralité que  $R(t_0) = 0$ , ne considérant ainsi que les nouveaux individus guéris, à partir de la date  $t_0$ . Nous avons fixé  $D(t_0) = 3523$ , le nombre de décès à l'hôpital au 31 mars. La population initiale  $S$  au début de la période, devrait encore être proche de la population totale française: au 31 mars, seuls 52 128 cas avaient été observés en France, soit 0,08 % de la population totale. Un facteur 8 avait été estimé par Roques et al. (15) entre le nombre cumulé de cas observés et le nombre réel de cas au début de l'épidémie. Même si ce facteur peut avoir changé, cela signifie que la proportion de la population totale qui a été infectée au 31 mars est encore faible. Nous pouvons obtenir une limite supérieure du nombre cumulé de cas d'ici le 31 mars en divisant le nombre de décès à l'hôpital à la fin de la période d'observation (10 129 au 14 avril) par l'IFR hospitalier [0,5%, comme estimé dans (15)] conduisant à environ 2 millions de cas. Cela signifie que la valeur de  $S(t_0)$  est comprise entre 65 et 67 millions de cas. Pour notre calcul, nous avons supposé que  $S(t_0) = 66 \cdot 10^6$ , correspondant à environ 98,5% de la population française. Comme le montre la figure S3, nos résultats ne sont pas très sensibles à la valeur de  $S(t_0)$  (du moins lorsque  $S / N$  reste proche de 1).

#### 2.2.1.2. Méthode numérique

Le système EDO (1) a été résolu grâce à un algorithme numérique standard, utilisant le solveur Matlab ® ode45.

### 2.2.2. Modèle d'observation

Le nombre de cas testés positifs au jour  $t$ , noté  $\hat{\delta}_t$ , est modélisée par des lois binomiales indépendantes, conditionnellement au nombre de tests  $n_t$  effectués au jour  $t$ , et à  $p_t$  la probabilité d'être testé positif dans cet échantillon:

$$\hat{\delta}_t \sim Bi(n_t, p_t). \quad (2)$$

La population testée est constituée d'une fraction des cas infectieux et d'une fraction des susceptibles:  $n_t = \tau_1(t) I(t) + \tau_2(t) S(t)$ . Donc,

$$p_t = \frac{\sigma I(t)}{I(t) + \kappa_t S(t)}.$$

avec  $\kappa_t := \tau_2(t) / \tau_1(t)$ , la probabilité relative de subir un test de dépistage pour un individu de type  $S$  vs un individu de type  $I$ . Nous avons supposé que le rapport  $\kappa$  était indépendant de  $t$  sur la période d'observation. Le coefficient  $\sigma$  correspond à la sensibilité du test. Dans la plupart des cas, des tests RT-PCR ont été utilisés et les données existantes indiquent que la sensibilité de ce test utilisant des écouvillons pharyngés et nasaux est d'environ 63 à 72% (25). Nous avons supposé ici  $\sigma = 0,7$  (sensibilité de 70%).

Chaque jour, le nombre de nouveaux décès observés (hors maisons de retraite), noté  $\hat{\mu}_t$ , est modélisée par des distributions de Poisson indépendantes conditionnellement sur le processus  $D(t)$ , avec la valeur moyenne  $D(t) - D(t-1)$  (qui mesure l'accroissement quotidien du nombre de décès):

$$\hat{\mu}_t \sim Poisson(D(t) - D(t-1)). \quad (3)$$

Notez que le temps  $t$  dans (1) est une variable continue, tandis que les observations  $\hat{\delta}_t$  et  $\hat{\mu}_t$  sont à des moments discrets. Par souci de simplicité, nous avons utilisé la même notation  $t$  pour les jours dans les cas discrets et continus. Dans les formules (2) et (3),  $I(t)$ ,  $S(t)$  et  $D(t)$  sont calculés à la fin du jour  $t$ .

### 2.2.3. Inférence statistique

Les paramètres inconnus sont  $\alpha$ ,  $\gamma$ ,  $\kappa$  et  $I_0$ . Nous avons utilisé une méthode bayésienne (26) pour estimer la distribution postérieure de ces paramètres.

### 2.2.3.1. Calcul de la fonction de vraisemblance

La probabilité  $L$  est définie comme la probabilité des observations (ici, les incréments  $\{\hat{\delta}_t, \hat{\mu}_t\}$  conditionnellement aux paramètres. En utilisant les modèles d'observation (2) et (3) et en supposant que les incréments  $\hat{\delta}_t$  et  $\hat{\mu}_t$  sont indépendantes conditionnellement au processus SIRD sous-jacent et que le nombre de tests  $n_t$  est connu, on obtient:

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

avec  $t_i$  la date de la première observation et  $t_f$  la date de la dernière observation. Dans cette expression  $L(\alpha, \gamma, \kappa, I_0)$  dépend de  $\alpha, \gamma, \kappa, I_0$  à travers  $p_t$  et  $D(t)$ .

### 2.2.3.2. Distribution postérieure

La distribution postérieure correspond à la distribution des paramètres conditionnellement sur les observations:

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

où  $\pi(\alpha, \gamma, \kappa, I_0)$  correspond à la distribution a priori des paramètres (détaillée ci-dessous) et  $C$  est une constante de normalisation indépendante des paramètres.

### 2.2.3.3. Distribution antérieure

Concernant le taux de contact  $\alpha$ , le nombre initial de cas infectieux  $I_0$  et la probabilité  $\kappa$ , nous avons utilisé des distributions a priori uniformes indépendantes non informatives dans les intervalles  $\alpha \in (0, 1)$ ,  $I_0 \in (1, 10^7)$  et  $\kappa \in (0, 1)$ . Pour surmonter les problèmes d'identifiabilité, nous avons utilisé une distribution préalable informative pour  $\gamma$ . Cette distribution, disons  $f_g$ , a été obtenue par Roques et al. (15) au stade précoce de l'épidémie ( $f_g$  est représentée sur la figure S1). Dans Roques et al. (15), le nombre de cas infectieux  $I_0$  au début de l'épidémie était connu (égal à 1) et n'avait pas besoin d'être estimé. Ainsi, nous avons estimé dans Roques et al. (15) la distribution du paramètre  $\gamma$  en calculant la distribution de la classe infectieuse et en utilisant la formule  $D'(t) = \gamma I(t)$  avec les données de mortalité (qui n'ont pas été utilisées pour l'estimation des autres paramètres, contrairement à la présente étude). Enfin, la distribution a priori est définie comme suit:

$$\pi(\alpha, \gamma, \kappa, I_0) = \mathbf{1}_{(\alpha, \kappa, I_0) \in (0,1) \times (0,1) \times (1,10^7)} f_g(\gamma).$$

Le calcul numérique de la distribution postérieure est réalisé avec un algorithme Metropolis-Hastings (MCMC), en utilisant 5 chaînes indépendantes, chacune avec  $10^6$  itérations, à partir du mode postérieur. Pour trouver le mode postérieur, nous avons utilisé l'algorithme de minimisation contrainte BFGS, appliqué à  $-\ln(L) - \ln(\pi)$ , via la fonction Matlab® `fmincon`. Afin de trouver un minimum global, nous avons appliqué cette méthode à partir de 4000 valeurs initiales aléatoires. Les codes Matlab® sont disponibles en tant que matériel supplémentaire.

## 3. Résultats

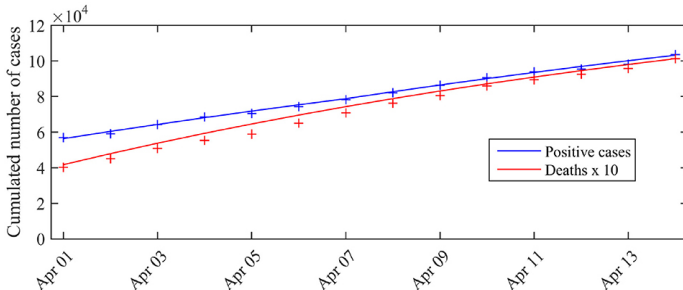
### 3.1. Ajustement du modèle

Notons  $(\alpha^*, \gamma^*, \kappa^*, I_0^*)$  le mode postérieur, et  $S^*(t)$ ,  $I^*(t)$ ,  $R^*(t)$ ,  $D^*(t)$  les solutions du système (1) associées à ces valeurs de paramètres. Le modèle d'observation (2) implique que le nombre attendu associé de cas testés positifs au jour  $t$  est  $n_t p_t^*$  (espérance d'un binôme) avec

$$p_t^* = \frac{\sigma I^*(t)}{I^*(t) + \kappa^* S^*(t)}.$$

Le modèle d'observation (3) implique que le nombre cumulé attendu de décès au jour  $t$  est  $J^*(t)$ .

Pour évaluer l'ajustement du modèle, nous avons comparé ces attentes et les observations, c'est-à-dire le nombre cumulé de cas testés positifs,  $\Sigma_t := C_0 + \sum_{s=t_0, \dots, t_0+13} \hat{\delta}_s$  avec  $C_0$  le nombre de cas testés positifs au 31 mars ( $C_0 = 52\,128$ ) et le nombre cumulé de décès  $M_t := M_0 + \sum_{s=t_0, \dots, t_0+13} \hat{\mu}_s$ , avec  $M_0$  le nombre de décès signalés (à l'hôpital) au 31 mars ( $M_0 = 3123$ ). Les résultats sont présentés dans la figure 1. Nous observons une bonne correspondance avec les données.

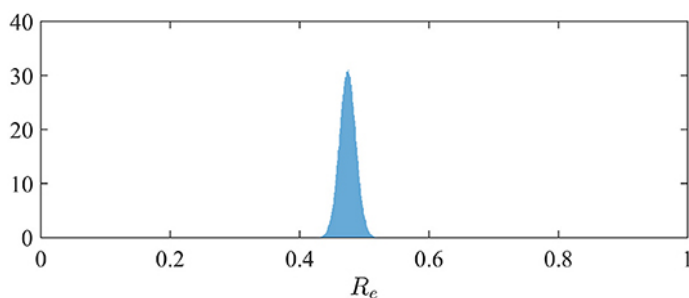


Graphique 1. Nombre attendu de cas observés et de décès associés au mode postérieur vs nombre de cas réellement détectés (nombre total de cas). La courbe bleue correspond au nombre attendu de cas testés positifs  $C_0 + \sum_{s=1, \dots, t} n_s p_s^*$  donné par le modèle, la courbe rouge correspond au nombre cumulé attendu de décès  $D^*(t)$  (hors maisons de retraite). Les croix correspondent aux observations (croix bleues: nombre cumulé de cas positifs, croix rouges: nombre cumulé de décès).  $C_0$  est le nombre de cas testés positifs au 31 mars ( $C_0 = 52\,128$ ).

Les distributions postérieures par paire des paramètres  $(\alpha, I_0)$ ,  $(\alpha, \gamma)$ ,  $(\alpha, \kappa)$ ,  $(\gamma, I_0)$ ,  $(\gamma, \kappa)$ ,  $(\kappa, I_0)$  sont représentées dans la figure S2. À l'exception du paramètre  $\gamma$  (figure S1), pour lequel nous avons choisi un a priori informatif, la distribution postérieure est clairement différente de la distribution antérieure, montrant que de nouvelles informations étaient bien contenues dans les données.

### 3.2. Taux de contact et numéro de reproduction effectif

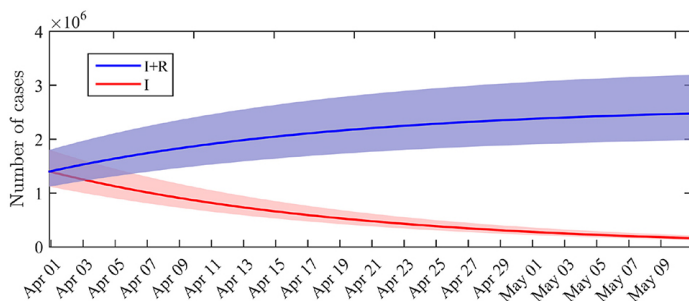
Le nombre effectif de reproduction peut être simplement dérivé de la relation  $R_e = \alpha / (\beta + \gamma)$  lorsque S est proche de N (3). La distribution de  $R_e$  est donc facilement dérivée de la distribution marginale postérieure du taux de contact  $\alpha$  (puisque nous avons supposé  $\beta = 1/10$ ; voir section 2.2). Il est illustré à la figure 2. Nous observons une valeur moyenne de  $R_e$  de 0,47 (95% -CI: 0,45-0,50).



Graphique 2 . Distribution postérieure du nombre effectif de reproduction  $R_e$  en France.

### 3.3. Dynamique de la classe infectieuse

La distribution marginale postérieure de  $I_0$  indique que le nombre d'individus infectieux au début de la période considérée (c'est-à-dire au 1er avril) est de  $1,4 \cdot 10^6$  (IC à 95%:  $1,1 \cdot 10^6 - 1,8 \cdot 10^6$ ). Le calcul de la solution de (1) avec la distribution postérieure des paramètres conduit à un nombre d'infectieux  $I(t_f) = 7,0 \cdot 10^5$  et un nombre total de cas infectés (y compris guéris)  $(I+R)(t_f) = 2,0 \cdot 10^6$  à la fin de la période d'observation (14 avril). D'ici le 10 mai, si les politiques de restriction restent inchangées, nous obtenons une prévision de  $I(T) = 1,6 \cdot 10^5$  cas infectieux (IC 95%:  $1,3 \cdot 10^5 - 2,1 \cdot 10^5$ ) et  $(I+R)(T) = 2,5 \cdot 10^6$  cas infectés, y compris guéris (IC à 95%:  $2,0 \cdot 10^6 - 3,2 \cdot 10^6$ ). La dynamique des distributions de I et I + R est illustrée à la figure 3. Au 10 mai, le nombre total de cas infectés (y compris guéris) correspond donc à une fraction de 3,7% de la population totale française. Cette valeur n'inclut pas les cas guéris avant le 1er avril.



Graphique 3 . Distribution du nombre de cas infectieux I (t) et du nombre cumulé de cas infectés I (t) + R (t) dans le temps. Lignes pleines: valeur moyenne obtenue à partir de la distribution postérieure des paramètres. Zones ombrées: 0,025–0,975 plages interquartiles.

## 4. Discussion

De nombreuses études se sont concentrées sur l'estimation du nombre de reproduction de base  $R_0$  de l'épidémie de COVID-19, basée sur des méthodes basées sur les données et des modèles mathématiques [par exemple, (4 , 27)] décrivant l'épidémie depuis son début. En moyenne, la valeur estimée de  $R_0$  était d'environ 3,3. Nous nous sommes concentrés ici sur une période d'observation qui a débuté après la mise en place du confinement en France.

Nous avons obtenu un nombre effectif de reproduction divisé par un facteur 7, comparé à l'estimation du  $R_0$  réalisée en France au début de l'épidémie, avant que le pays ne passe en confinement [une valeur  $R_0 = 3,2$  a été obtenue en (15)]. Cela indique que les politiques de restriction ont été très efficaces pour réduire le taux de contact et donc le nombre de cas infectieux. En particulier, la valeur  $R_e = 0,47$  est nettement inférieure à la valeur seuil 1 où l'épidémie commence à s'éteindre.

La décroissance du nombre de cas infectieux peut également être observée à partir de nos simulations. Il faut noter que, bien que le nombre de cas infectieux soit un processus latent, ou «non observé», le cadre mécanistique-statistique nous a permis d'estimer sa valeur (figure 3). Le nombre cumulé de cas infectés que nous avons obtenu au 10 mai (I + R) correspond à une fraction de 3,7% (95% -CI: 3,0–4,8%) de la population totale française, sans tenir compte du nombre d'individus guéris avant le 1er avril, ce qui n'est pas connu. Sur la base d'une valeur  $R_0 = 3,2$ , le seuil d'immunité du groupe, correspondant à la fraction minimale de la population qui doit être immunisée pour arrêter l'épidémie, serait de  $1 - 1 / R_0 \approx 69\%$  [un seuil de 80% a été proposé dans (28)]. Cette proportion ne sera probablement pas atteinte d'ici le 10 mai. Comme le souligne Angot (29), un assouplissement trop rapide des restrictions liées au confinement avant que l'immunité du groupe ne soit atteinte ou qu'une prophylaxie efficace ne soit développée exposerait la population à une seconde vague incontrôlée de l'infection. Dans le pire des cas, le nombre effectif de reproduction  $R_e$  approcherait la valeur initialement estimée de  $R_0$ , et la deuxième vague commencerait avec environ  $1,6 \cdot 10^5$  individus infectieux (en comparaison avec les quelques cas qui ont déclenché la première vague de France) et environ  $64 \cdot 10^6$  individus sensibles. Garder une valeur faible de  $R_e$  est donc crucial pour éviter la saturation des installations hospitalières.

Nous avons délibérément choisi un modèle mécaniste parcimonieux avec quelques paramètres pour éviter les problèmes d'identifiabilité. Les extensions possibles incluent des modèles structurés par étapes, où la classe infectieuse I et le taux de contact  $\alpha$  dépendraient d'une autre variable:  $I = I(t, \tau)$  et  $\alpha = \alpha(t, \tau)$  avec  $\tau$  le temps depuis l'infection, à prendre en tenant compte de la dynamique de la charge virale sur l'infectiosité. Voir par exemple Murray (3) (chapitre 19.6) pour une introduction à ces approches de modélisation. Une autre extension perspicace consisterait à utiliser des modèles spatialement explicites, par exemple des modèles réaction-diffusion (30) pour décrire la propagation spatiale de l'épidémie et pour pouvoir estimer les valeurs locales du paramètre  $R_e$  et le nombre de cas sensibles. Bien que l'immunité du groupe soit loin d'être atteinte à l'échelle du pays, il est probable que la fraction d'individus immunisés varie fortement sur le territoire, avec des effets d'immunité locale possibles [par exemple, au 4 avril, la proportion de personnes atteintes du SRAS-CoV-2 confirmé l'infection basée sur la détection des anticorps était de 41% dans un lycée situé dans le Nord de la France (31)].

### Déclaration de disponibilité des données

Des ensembles de données accessibles au public ont été analysés dans cette étude. Ces données sont disponibles ici: <https://www.gouvernement.fr/info-coronavirus/carte-et-donnees> <https://geodes.santepubliquefrance.fr> et <https://ourworldindata.org/coronavirus-testing> .

## Contributions d'auteur

LR, EK, JP, AS et SS ont conçu le modèle et conçu l'analyse statistique. LR et SS ont rédigé l'article. LR a effectué les calculs numériques. Tous les auteurs ont révisé le manuscrit.

## Le financement

Ce travail a été financé par le réseau INRAE: MEDIA.

## Conflit d'intérêt

Les auteurs déclarent que la recherche a été menée en l'absence de toute relation commerciale ou financière pouvant être interprétée comme un conflit d'intérêts potentiel.

## Remerciements

Ce manuscrit a été publié en pré-impression à medRxiv (32).

## Matériel complémentaire

Le matériel supplémentaire pour cet article peut être trouvé en ligne à: <https://www.frontiersin.org/articles/10.3389/fmed.2020.00274/full#supplementary-material>

## Références

1. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19. World Health Organization (2020).
2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* (2020) 20:533–4. doi: 10.1016/S1473-3099(20)30120-1
3. Murray JD. *Mathematical Biology*. 3rd Edition. *Interdisciplinary Applied Mathematics* 17. New York, NY: Springer-Verlag (2002).
4. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med.* (2020) 27:taaa021. doi: 10.1093/jtm/taaa021
5. De Serres G, Gay NJ, Farrington CP. Epidemiology of transmissible diseases after elimination. *Am J Epidemiol.* (2000) 151:1039–48. doi: 10.1093/oxfordjournals.aje.a10145
6. Kraemer MU, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science.* (2020) 368:493–7. doi: 10.1126/science.abb4218
7. Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science.* (2020) 368:742–6. doi: 10.1101/2020.02.18.20024414
8. Tian H, Liu Y, Li Y, Wu CH, Chen B, Kraemer MU, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science.* (2020) 368:638–42. doi: 10.1126/science.abb6105
9. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Age profile of susceptibility, mixing, and social distancing shape the dynamics of the novel coronavirus disease 2019 outbreak in China. *medRxiv.* (2020). doi: 10.1101/2020.03.19.20039107
10. Liu Z, Magal P, Seydi O, Webb G. Understanding unreported cases in the 2019-nCoV epidemic outbreak in Wuhan, China, and the importance of major public health interventions. *MPDI Biol.* (2020) 9:50. doi: 10.3390/biology9030050
11. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health.* (2020) 5:e261–70. doi: 10.1101/2020.03.09.20033050
12. Wikle CK. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology.* (2003) 84:1382–94. doi: 10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2
13. Roques L, Bonnefon O. Modelling population dynamics in realistic landscapes with linear elements: a mechanistic-statistical reaction-diffusion approach. *PLoS ONE.* (2016) 11:e0151217. doi: 10.1371/journal.pone.0151217
14. Abboud C, Bonnefon O, Parent E, Soubeyrand S. Dating and localizing an invasion from post-introduction data and a coupled reaction-diffusion-absorption model. *J Math Biol.* (2019) 79:765–89. doi: 10.1007/s00285-019-01376-x
15. Roques L, Klein E, Papaix J, Sar A, Soubeyrand S. Using early data to estimate the actual infection fatality ratio from COVID-19 in France. *MDPI Biol.* (2020) 9:97. doi: 10.3390/biology9050097
16. 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. *Lancet Infect Dis.* doi: 10.1016/S1473-3099(20)30243-7
17. Russell TW, Hellewell J, Jarvis CI, van Zandvoort K, Abbott S, Ratnayake R, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Eurosurveillance.* (2020) 25:2000256. doi: 10.2807/1560-7917.ES.2020.25.12.2000256
18. Santé Publique France. COVID-19: Point épidémiologique du 16 avril 2020. <https://github.com/owid/covid-19-data/blob/master/public/data/testing/>; <https://geodes.santepubliquefrance.fr/> (accessed 8 May, 2020).
19. DREES. 728 000 résidents en établissements d'hébergement pour personnes âgées en 2015. [https://dreessolidarites-santegouvfr/IMG/pdf/er1015pdf\(2020\).](https://dreessolidarites-santegouvfr/IMG/pdf/er1015pdf(2020).)
20. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* (2020) 395:1054–62. doi: 10.1016/S0140-6736(20)30566-3
21. He X, Lau EH, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med.* (2020) 26:672–5. doi: 10.1038/s41591-020-0869-5
22. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New Engl J Med.* (2020) 382:1199–207. doi: 10.1056/NEJMoa2001316
23. Liu Z, Magal P, Seydi O, Webb G. A COVID-19 epidemic model with latency period. *Infect Dis Modell.* (2020) 5:323–37. doi: 10.1016/j.idm.2020.03.003
24. Li L, Zhang J, Liu C, Zhang HT, Wang Y, Wang Z. Analysis of transmission dynamics for Zika virus on networks. *Appl Math Comput.* (2019) 347:566–77. doi: 10.1016/j.amc.2018.11.042
25. Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA.* (2020) 323:1843–4. doi: 10.1001/jama.2020.3786
26. Marin JM, Robert CP. *Bayesian Core*. New York, NY: Springer (2007).
27. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int J Infect Dis.* (2020) 92:214–7. doi: 10.1016/j.ijid.2020.01.050
28. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand. London: Imperial College (2020).
29. Angot P. Early Estimations of the Impact of General Lockdown to Control the Covid-19 Epidemic in France. <https://hal.archives-ouvertes.fr/hal-02545893> (2020).
30. Cantrell RS, Cosner C. *Spatial Ecology via Reaction-Diffusion Equations*. Chichester: John Wiley & Sons Ltd (2003). doi: 10.1002/0470871296

31. Fontanet A, Tondeur L, Madec Y, Grant R, Besombes C, Jolly N, et al. Cluster of COVID-19 in northern France: a retrospective closed cohort study. medRxiv. (2020). doi: 10.1101/2020.04.18.20071134. [Epub ahead of print].
32. Roques L, Klein EK, Papaix J, Sar A, Soubeyrand S. Effect of a one-month lockdown on the epidemic dynamics of COVID-19 in France. medRxiv. (2020). doi: 10.1101/2020.04.21.20074054. [Epub ahead of print].