

Rapport : Limites et délais dans l'estimation du nombre de reproduction

 alizon.ouvaton.org/Rapport5_R.html

Groupe de modélisation de l'équipe ETE
(Laboratoire MIVEGEC, CNRS, IRD, Université de Montpellier)
7 avril 2020

Préambule

Le contenu de ce rapport n'est en aucun cas officiel et est avant tout à destination pédagogique. De plus, nos calculs sont souvent réalisés avec des hypothèses volontairement très simplificatrices. Ainsi, les moyennes nationales ne reflètent pas des situations locales.

Contexte

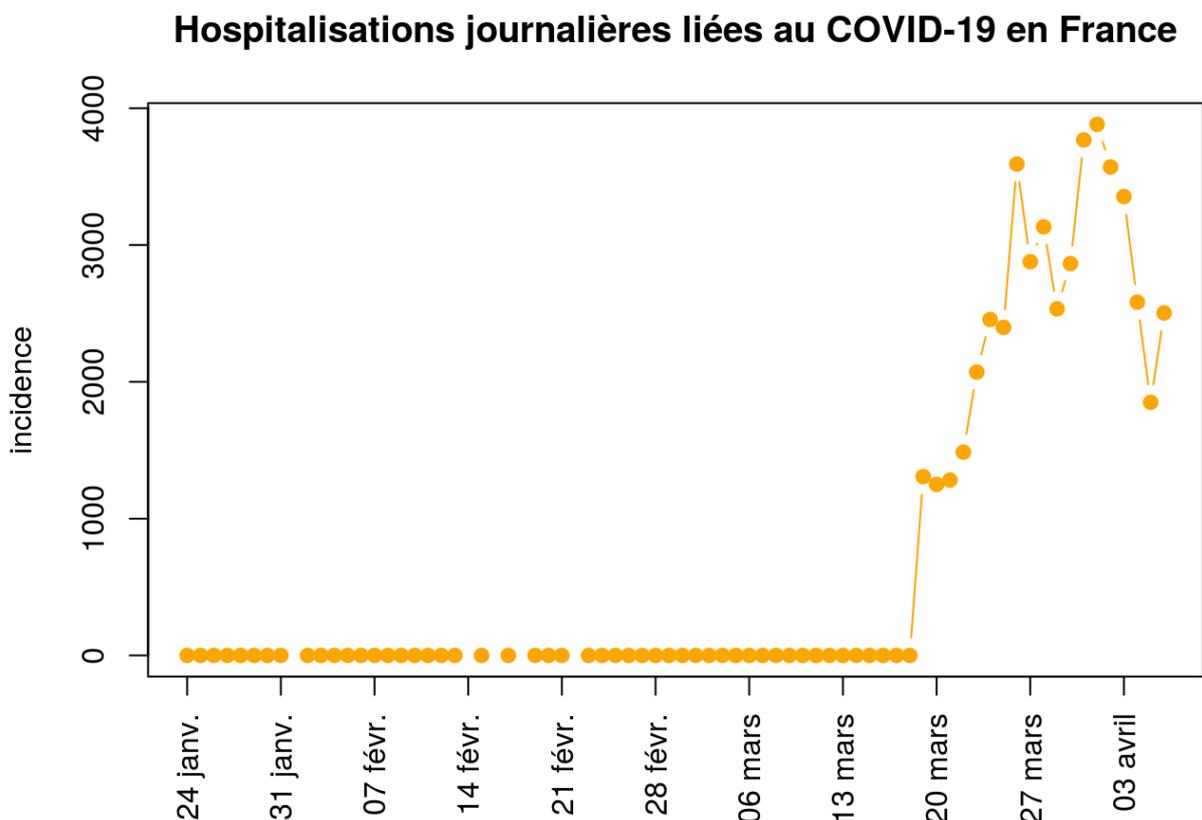
Les *délais temporels* représentent un défi majeur en santé public. En effet, il faut en général plusieurs jours, voire semaines, avant de pouvoir détecter les effets d'une intervention telle que le confinement dans les courbes d'*incidence*.

Intuitivement, on se tourne le plus souvent vers le nombre de nouveaux cas détectés chaque jour pour voir les effets des interventions. Comme nous le notions dans notre [Rapport 1](#), les délais avant d'observer un effet dans les données est de plusieurs jours puisqu'il faut compter en moyenne 5 jours avant l'apparition des symptômes, temps auquel il faut rajouter celui avant la réalisation d'un test de dépistage et enfin le délai de notification. Dans le cas de la France, un autre problème que nous mentionnons est lié aux *variations des intensités de dépistage*. Clairement, le gouvernement a augmenté le nombre de dépistages, ce qui se traduit mécaniquement par une augmentation du nombre de cas détectés. Au final, **les données d'incidence de nouveaux cas détectés en France sont à peu près inexploitable sur le long terme.**

Pour résoudre ce problème, nous avons considéré l'incidence journalière des **nouvelles hospitalisations** ainsi que celle des **nouveaux décès**. Toutes deux proviennent de [Santé Publique France](#).

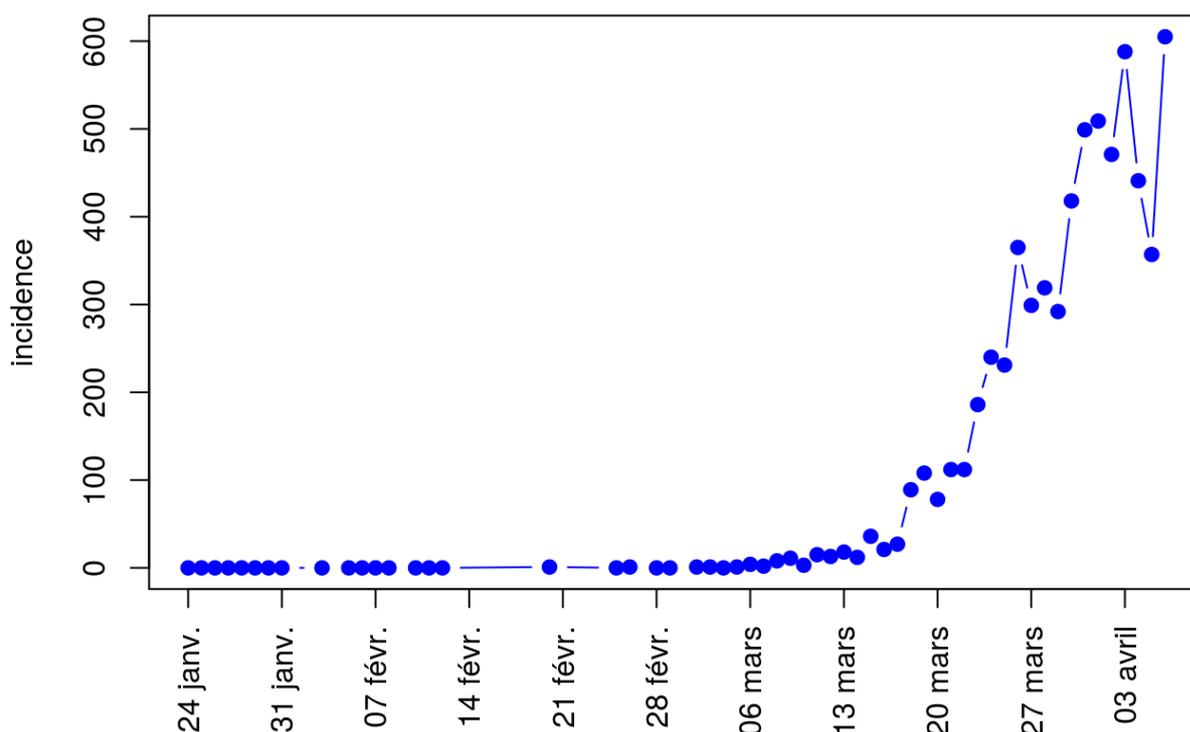
Pour les hospitalisations, les données ne sont malheureusement disponibles que depuis le **24 janv. 2020**. Toutefois, leur qualité est bien plus élevée que celles des données de nouveaux cas dépistés. De plus, le délai entre une mesure de contrôle (fermeture des écoles, confinement,...) et sa manifestation au niveau des nouvelles hospitalisation est

d'environ 2 semaines (on estime à 5 jours le temps d'apparition des symptômes et à 8 jours le temps entre l'apparition des symptômes et l'hospitalisation pour les cas sévères). Les données sont représentées sur le graphique suivant.



En ce qui concerne les décès, les données sont disponibles depuis le début de l'épidémie. Leur qualité est potentiellement plus élevée que celles des données de nouveaux cas, mais peut-être moindre que celle des données d'hospitalisation car tous les décès attribuables au COVID-19 ne sont pas forcément recensés. De plus les décès dans les EHPAD n'étaient initialement pas signalés. Le délai entre une intervention et sa manifestation au niveau des décès est évidemment plus long et est de l'ordre de 3 à 5 semaines. Les données sont représentées sur le graphique suivant.

Décès journaliers liés au COVID-19 en France



Notez que les données de Santé Publique France ne signalent pas directement le nombre de nouvelles hospitalisations. Nous avons obtenu ce nombre ajoutant le nombre de nouvelles hospitalisations entre une date t et une date $t-1$, le nombre de décès à l'hôpital entre $t-1$ et t et le nombre de nouvelles guérisons à l'hôpital entre $t-1$ et t à partir des données de Santé Publique France. Il est possible que ces chiffres ne correspondent pas exactement au nombre de nouvelles hospitalisations. D'ailleurs, notre incidence cumulée est de **48759** hospitalisations et le nombre de Santé Publique France est au 7 avril était **54036**, ce qui illustre les limites de nos résultats.

Enfin, les données d'incidence de patient·e·s en réanimation seraient a priori plus fiables que celle d'hospitalisation mais il est pour le moment impossible d'avoir accès à leur incidence (on sait juste le nombre de personnes actuellement en réanimation).

Méthode

Nous avons utilisé les méthodes décrites dans notre [Rapport 1](#) pour estimer le nombre de reproduction temporel (ou efficace), noté $R(t)$, qui correspond au nombre de personnes qu'infecte une personne infectée au cours de son infection à une date t . Cette estimation est effectuée grâce au package [R0](#).

Pour mémoire, si à une date t , $R(t)$ est inférieur à 1, l'épidémie décroît et un pic épidémique est donc atteint. Tant que $R(t) > 1$ l'épidémie est en phase de croissance. Le but des mesures de confinement et de toutes les mesures de contrôle est de diminuer $R(t)$.

Il faut préciser que notre approche n'est pas conventionnelle. En effet, ces calculs sont conçus pour être effectués sur des séries temporelles d'incidence de nouveaux cas.

Les données les plus récentes d'hospitalisation datent du **06 avril 2020** et celles de décès du **06 avril 2020**.

L'intervalle sériel et ses limites

Une des raisons pour lesquelles notre approche est peu orthodoxe est que le calcul du nombre de reproduction repose sur la durée moyenne pendant laquelle une personne infectée est contagieuse. Cette information n'est pas présente dans les séries temporelles d'incidence et pour l'obtenir, il faut faire du suivi de contact entre personnes infectées. Pour un exemple, on peut voir l'étude de Li *et alii* (2020, [NEJM](#)).

En première approche (voire Britton & Scalia Tomba 2020 [J R Soc Interface](#) pour une revue pointue en anglais), le nombre de reproduction peut être estimé à partir des données de la manière suivante :

$R = \text{taux de croissance exponentielle} \times \text{durée de contagiosité} + 1$
 $R = \text{taux de croissance exponentielle} \times \text{durée de contagiosité} + 1$

Si le taux de croissance, calculé à partir des courbes d'incidence, est propre à la France, la durée de contagiosité provient quant à elle de données collectées en Asie.

En pratique, la durée de contagiosité est difficile à estimer et ne capture pas la variation du potentiel de transmission des individus infectés au cours du temps. C'est pourquoi les approches statistiques lui préfèrent l'**intervalle sériel**, défini comme la durée qui sépare l'apparition des symptômes chez un infecteur et son infecté. L'intervalle sériel est un bon reflet du **temps de génération**, à savoir la durée qui sépare la contamination de l'infecteur et la contamination de l'infecté, qui est le paramètre clé de la cinétique épidémiologique, mais qui lui aussi difficile à déterminer directement. Schématiquement, à taux de croissance exponentiel constant, plus le temps de génération est élevé, plus l'intervalle sériel et la durée de contagiosité sont élevés, et donc plus le nombre de reproduction est élevé.

Ce rapport a donc plusieurs limites liées à l'intervalle sériel :

1. En travaillant sur les hospitalisations et les décès, nous faisons l'hypothèse que cet intervalle reste en moyenne inchangé, ce qui peut se révéler trop simplificateur. En outre, plus l'événement de référence pour le calcul de l'intervalle est éloigné de la contamination (l'hospitalisation et le décès arrivent après l'apparition des symptômes), plus la variance augmente, et donc l'incertitude des estimations ainsi obtenues.

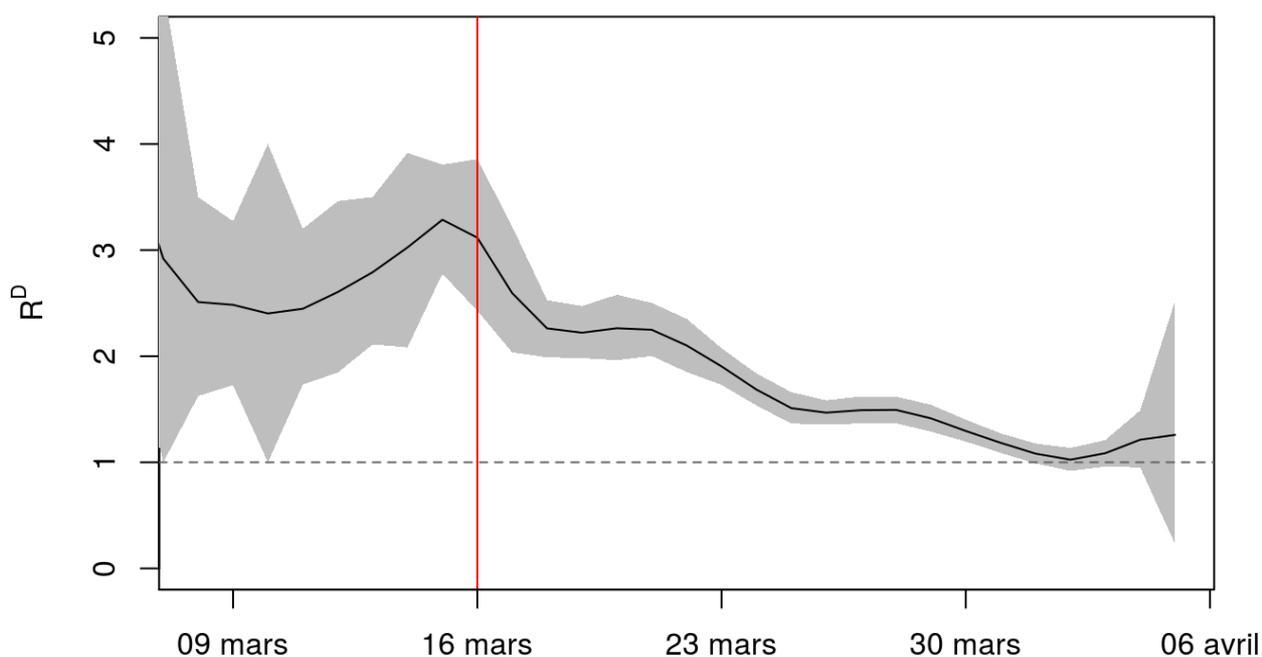
2. Les données servant à estimer l'intervalle sériel proviennent de l'épidémie en Asie et nous avons utilisé les données brutes compilées par Nishiura *et alii* (2020, [In J Inf Dis](#)). À ce jour, il n'existe pas de données sur l'intervalle sériel en France, ce qui pourrait grandement affecter les valeurs exactes de $R(t)$ si la durée moyenne de l'intervalle est supérieure à 4,2 jours (durée estimée par Nishiura *et alii* 2020). Ainsi, certaines études ont trouvé un intervalle moyen de 7 jours.
3. On pourrait même envisager que l'intervalle sériel soit différent dans différents départements français, ce qui engendrerait des différences locales.

Résultats

Nous rappelons que ce travail est avant tout académique. En particulier, du fait des incertitudes autour de l'intervalle sériel, les résultats sont à interpréter de manière **qualitative**. De plus, des données nationales ne reflètent pas les situations **locales**.

À partir de l'incidence des décès, on peut estimer un nombre de reproduction temporel (noté $R^D(t)$). Comme il y a un délai d'au moins trois à quatre semaines entre le nombre de décès à une date donnée et l'état de l'épidémie à cette date, on ne s'attend pas à voir un effet du confinement sur le graphique suivant. La ligne rouge est le début du confinement et la zone grisée l'intervalle de confiance à 95 % de l'estimation de $R^D(t)$.

Nombre de reproduction efficace (données de mortalité en France)

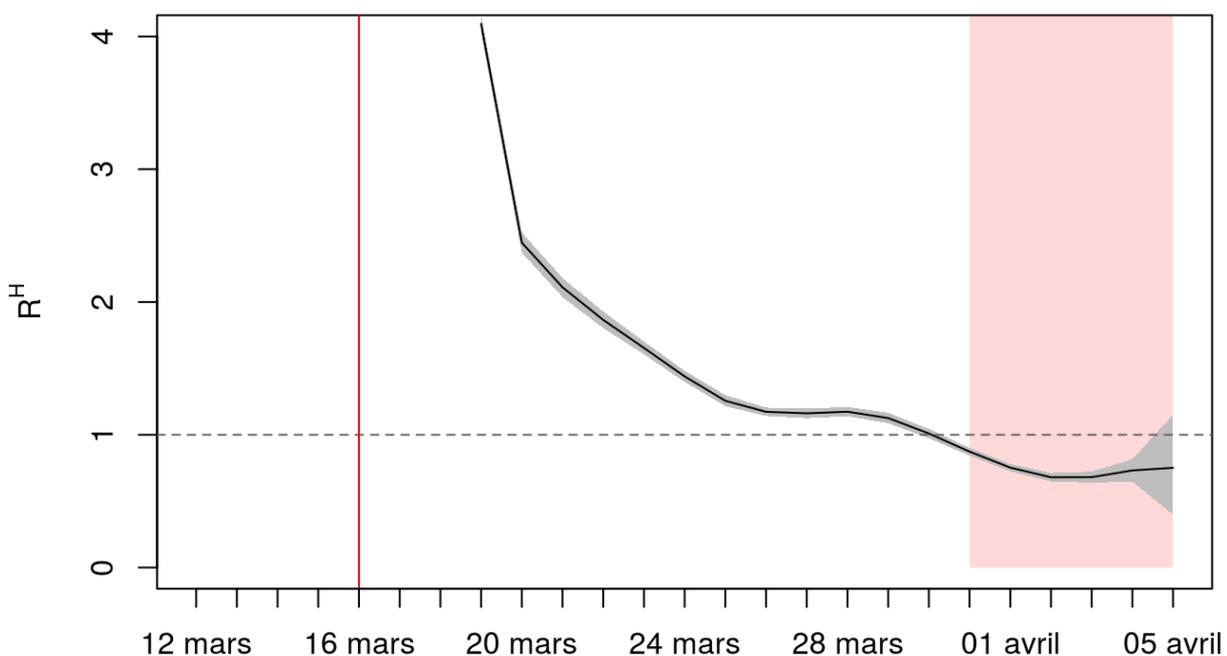


À partir de l'incidence des décès, la valeur la plus récente du nombre de reproduction de base, qui correspond à l'état de l'épidémie il y a environ 1 mois, soit environ le **07 mars 2020**, est la suivante :

```
## [1] "R^D(t) = 1.26 avec 95 % des valeurs entre 0.23 et 2.52."
```

Une approche similaire peut être réalisée pour estimer un nombre de reproduction à partir des données de nouvelles hospitalisations (noté $R^H(t)$). Dans ce cas, le délai entre le nombre d'hospitalisations et l'état de l'épidémie est d'environ 2 semaines. Sur le graphique suivant, on a représenté en rouge clair la période de temps pour laquelle on pourrait s'attendre à voir un effet du confinement.

Nombre de reproduction efficace (données d'hospitalisation en France)



À partir de l'incidence des hospitalisations, la valeur la plus récente du taux de reproduction de base, qui correspond à l'état de l'épidémie il y a environ 2 semaines, soit environ le **23 mars 2020**, est la suivante :

```
## [1] "R^H(t) = 0.75 avec 95 % des valeurs entre 0.4 et 1.15."
```

Discussion et interprétations

Une des difficultés majeures en santé publique est que les décisions prises ne se manifestent pas tout de suite dans les données. Ainsi, la mise en place d'un confinement prend environ 2 semaines pour être détectable au niveau de l'incidence des hospitalisations et environ 4 semaines au niveau de l'incidence des décès.

Aujourd'hui, 22 jours se sont écoulés depuis la mise en place du confinement en France. On ne s'attend donc pas à détecter un effet dans la série temporelle des décès mais à potentiellement voir un signal dans celle des hospitalisations.

L'estimation d'un nombre de reproduction à partir des décès montre que la croissance de l'épidémie a commencé à diminuer avant la mise en place du confinement. On observe aussi des variations de la médiane qui peuvent être dues au bruit statistique dans les données, qui peut provenir des délais à faire remonter les données ou de l'agglomération des données locales. On voit aussi qu'il y a une grande incertitude quant à la valeur la plus récente de $RD(t)$, qui correspond à l'état de l'épidémie il y a environ 4 semaines.

L'estimation d'un nombre de reproduction temporel à partir des hospitalisations suggère une diminution plus forte que celle observée via les décès. Le délai pour la valeur médiane de $RH(t)$ à passer sous la barre de 1 (et le fait que la borne supérieure de l'intervalle de confiance soit encore bien supérieure à 1) peut avoir plusieurs origines :

- certaines données d'hospitalisation arrivent avec un délai important qui crée un décalage dans la série temporelle d'incidence,
- les mesures de confinement n'ont pas eu un effet immédiat suffisant pour faire décroître l'épidémie,
- en France, le délai entre la date d'infection et la date d'hospitalisation est supérieure à deux semaines.

Enfin, il faut garder à l'esprit que ces données sont nationales et peuvent masquer des **hétérogénéités spatiales**. Ainsi, certains départements sont peut-être déjà à leur pic épidémique, alors que d'autres en sont potentiellement encore loin. **Parler d'un nombre de reproduction de l'épidémie au niveau national aurait donc peu de sens** s'il y a trop de différences entre départements. Cela peut aussi expliquer certaines des fluctuations observées dans les courbes.

Rappelons que ces données sont à interpréter de manière **qualitative**, d'autant que l'utilisation d'incidences portant sur les hospitalisations ou les décès peut générer des incertitudes quant à l'**intervalle sériel**. De plus, ce dernier est pour le moment peu connu en France et dans les épidémies de COVID-19 hors d'Asie. Si la distribution de ses valeurs se trouve avoir une moyenne une variance différentes de celles utilisées ici, les valeurs exactes de RR pourraient être modifiées.

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 des nouvelles hospitalisations et de décès proviennent du site <https://github.com/opencovid19-fr>
- Pour approfondir les limites liées à l'estimation de l'intervalle sériel, on peut se référer à l'excellent article de Britton T & Scalia Tomba G (2019) Estimation in emerging epidemics: biases and remedies. *J R Soc Interface* 16:20180670 <https://doi.org/10.1098/rsif.2018.0670>
- Les données de l'intervalle sériel proviennent de l'analyse de Nishiura *et alii* (2020, *In J Inf Dis*) et sont issues du site <https://github.com/aakhmetz/nCoVSerialInterval2020>
- L'estimation de $R(t)$ est faite grâce au package de Obadia T, Haneef R & Boëlle PY (2012) The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Med Inform Decis Mak* 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 $R(t)$: SA, BR, MS
 - rédaction du rapport : SA et MS
 - validation du rapport : ensemble de l'équipe
- contact : mircea.sofonea@umontpellier.fr samuel.alizon@cnr.fr
-  Ce(tte) œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/).