

DREES MÉTHODES

N° 5 • octobre 2022

Les taux de personnes vaccinées et non vaccinées contre le Covid-19 en France

**Comparaison des indicateurs en open data
et tentative d'estimation**

Charlotte Geay, Tristan Haugomat, Hamid Khaoua, Milena Suarez Castillo, en collaboration avec
Mathilde Gaini (DREES)

Les taux de personnes vaccinées et non vaccinées contre le Covid-19 en France

Comparaison des indicateurs en open data et tentative d'estimation

Charlotte Geay, Tristan Haugomat, Hamid Khaoua, Milena Suarez Castillo, en collaboration avec Mathilde Gaini (DREES)

Remerciements : Les auteurs remercient Albert Vuagnat pour les exploitations du SNDS présentées dans ce dossier, la CNAM et Santé publique France pour les échanges méthodologiques, et ainsi que Benoît Ourliac, Frédéric Tallet et Javier Nicolau pour leurs commentaires et suggestions.

Retrouvez toutes nos publications sur : drees.solidarites-sante.gouv.fr

Retrouvez toutes nos données sur : data.drees.solidarites-sante.gouv.fr

DREES MÉTHODES

N° 5 • octobre 2022

Synthèse Les taux de personnes vaccinées et non vaccinées contre le Covid-19 en France

**Comparaison des indicateurs en open data
et tentative d'estimation**

Charlotte Geay, Tristan Haugomat, Hamid Khaoua, Milena Suarez Castillo, en collaboration avec
Mathilde Gaini (DREES)

Retrouvez toutes nos publications sur : drees.solidarites-sante.gouv.fr

Retrouvez toutes nos données sur : data.drees.solidarites-sante.gouv.fr

SYNTHÈSE

En France, comme dans de nombreux pays, des exploitations statistiques des systèmes d'information liés à la vaccination ont permis de suivre à une fréquence quotidienne le déploiement de la campagne de vaccination contre le Covid-19 et de mettre à disposition des estimations du taux de vaccination, en particulier par âge, un critère déterminant en termes de santé publique et de définition de la stratégie vaccinale.

Il existe aujourd'hui quatre sources de données publiées en *open data* par trois institutions différentes sur la couverture vaccinale contre le Covid-19 en France : en rapportant la population vaccinée aux estimations de population de l'Institut national de la statistique et des études économiques (Insee), la part de personnes non-vaccinées au 14 août 2022 serait de 6,5 % chez les 18 ans ou plus selon les chiffres de Santé publique France et 6,6 % selon la Caisse nationale de l'Assurance maladie (CNAM) ; 7,0 % des 20 ans ou plus seraient non vaccinés selon la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) ; enfin, la CNAM estime que 13,9 % de la population des 20 ans ou plus ayant consommé des soins remboursés en France en 2021 serait non vaccinée à la même date. Les trois premières sources donnent ainsi des estimations nationales convergentes au global, même si on constate des écarts plus importants par groupes d'âges. Les profils par âge de la part de personnes non-vaccinées sont par ailleurs très différents entre les trois premières sources et la quatrième.

L'objet de ce dossier est d'explicitier les différences méthodologiques qui expliquent les écarts plus ou moins importants entre ces quatre sources, afin d'identifier des améliorations possibles des indicateurs publiés en *open data*, mais aussi de proposer des estimations alternatives. Les raisons expliquant les écarts sont multiples. Elles tiennent à la définition du concept d'âge retenu, à la prise en compte ou non de la mortalité des personnes vaccinées, au choix de la population de référence à laquelle est rapportée la population vaccinée, ainsi qu'à la qualité perfectible de l'appariement de la base des personnes vaccinées VAC-SI avec d'autres sources permettant d'envisager des estimations du taux de vaccination parmi des populations larges (la base des « consommateurs » de soins de la CNAM, ou encore SI-DEP le système d'information sur les tests de dépistage Covid-19).

Les choix méthodologiques qui ont présidé à l'élaboration de ces statistiques ont été contraints par la nécessité de produire très rapidement et à haute fréquence des indicateurs pour le suivi de la campagne de vaccination, dans un contexte d'urgence sanitaire, et en mobilisant des données issues d'un système d'information (VAC-SI) lui-même mis en place dans des délais extrêmement courts. Disposant d'un plus grand recul sur ces données et indicateurs, il apparaît aujourd'hui important d'en explorer les limites respectives et de mener des analyses complémentaires.

En prenant en compte la mortalité et en appliquant plusieurs redressements sur les données, l'appariement des bases de données sur la vaccination (VAC-SI) et sur le dépistage (SI-DEP) conduit ainsi à proposer plusieurs alternatives aux estimations existantes, visant à assurer une plus grande homogénéité entre population vaccinée et population de référence. Bien que ce travail ne permette pas de conclure sur un chiffre de façon certaine, il conduit à estimer in fine que le taux de personnes non vaccinées de 20 ans ou plus résidant en France se situe probablement entre 8 % et 12 %, soit un niveau un peu plus élevé que celui vers lequel pointent les chiffres qui bénéficient de la plus forte couverture médiatique. Le taux de personnes non-vaccinées par âge serait le plus élevé autour de 30 ans.

L'exploitation de la quatrième vague de l'enquête EpiCov, en cours de collecte actuellement, devrait fournir une estimation complémentaire du taux de personnes vaccinées en population générale qui ne soit pas sujette aux mêmes limites que celles inhérentes aux sources et aux méthodes utilisées aujourd'hui. En revanche, elle ne fournira qu'une estimation ponctuelle ; la base VAC-SI demeure et demeurera l'outil central pour assurer le suivi de la couverture vaccinale en France.

SOMMAIRE

■ INTRODUCTION	2
■ LES DONNÉES MOBILISÉES	3
VAC-SI et les données sur la vaccination contre le Covid-19	3
SI-DEP, les données sur le dépistage du Covid-19	4
■ PLUSIEURS SOURCES MISES À DISPOSITION EN <i>OPEN DATA</i> DÉLIVRENT DES STATISTIQUES SUR LE TAUX DE NON-VACCINÉS	5
■ L'ABSENCE DE DONNÉES DISPONIBLES HOMOGENES ENTRE NUMÉRATEUR ET DÉNOMINATEUR CONDUIT À UNE PROBABLE SURESTIMATION DU NOMBRE DE VACCINÉS	9
Prendre en compte les individus décédés nécessite des traitements statistiques complexes	11
Identification des personnes vaccinées hors champ de la population résidente au sens de l'Insee	12
Estimations alternatives des taux de vaccinés, à partir d'hypothèses visant notamment à améliorer la cohérence entre la population vaccinée et la population de référence	13
Conclusion	17
Annexe 1.....	18
Annexe 2. Résultats complémentaires	19

■ INTRODUCTION

En France, comme dans de nombreux pays, des exploitations statistiques des systèmes d'information liés à la vaccination ont permis de suivre à une fréquence quotidienne le déploiement de la campagne de vaccination contre le Covid-19 et de mettre à disposition des estimations du taux de vaccination, en particulier par âge, un critère déterminant en termes de santé publique et de définition de la stratégie vaccinale.

Santé publique France (SpF), la Caisse nationale de l'Assurance maladie (CNAM) et la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) publient régulièrement en *open data* des informations sur le nombre et le taux de vaccinés, qui diffèrent sur certains points méthodologiques. Le présent dossier a pour but de dresser un état des lieux de ces écarts et de proposer une analyse détaillée de la cohérence entre les sources de données mises en place pour le suivi de l'épidémie de Covid-19 (en particulier SI-DEP et VAC-SI) et la population résidente de l'Institut national de la statistique et des études économiques (Insee).

« Vaccin-Covid », ou VAC-SI, est le système d'information permettant la mise en œuvre, le suivi et le pilotage de la campagne de vaccination en France¹. Il répond à plusieurs finalités, parmi lesquelles le suivi statistique de la campagne de vaccination, qui permet de dénombrer les personnes vaccinées en France. Ainsi, le champ de population couvert n'est qu'imparfaitement comparable à celui de la population résidente au sens de l'Insee. Le taux de vaccination correspond au ratio entre un numérateur (le nombre de personnes vaccinées) et un dénominateur (la population à laquelle on rapporte le numérateur) : en l'absence de données disponibles homogènes entre les deux, tout l'enjeu consiste à faire en sorte que ce numérateur et ce dénominateur concernent un périmètre de population homogène, le plus représentatif possible de la population résidant en France.

Il existe aujourd'hui quatre sources de données adossées à VAC-SI, publiées en *open data*, par trois institutions différentes sur la couverture vaccinale contre le Covid-19 en France : en rapportant la population vaccinée aux estimations de population de l'Insee, la part de personnes non vaccinées² serait, au 14 août 2022, de 6,5 % chez les 18 ans ou plus selon les chiffres de Santé publique France et de 6,6 % selon la CNAM ; 7,0 % des 20 ans ou plus seraient non vaccinés selon la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES) ; enfin, parmi la population des personnes ayant consommé des soins remboursés en France en 2021, la CNAM estime que 13,9 % de la population des 20 ans ou plus serait non vaccinée à la même date. Les trois premières sources donnent ainsi des estimations nationales convergentes au global, même si on constate des écarts plus importants par groupes d'âges. Les profils par âge de la part de personnes non vaccinées sont par ailleurs très différents entre les trois premières sources et la quatrième.

Les choix méthodologiques qui ont présidé à l'élaboration de ces statistiques ont été contraints par la nécessité de produire très rapidement et à haute fréquence des indicateurs pour le suivi de la campagne de vaccination, dans un contexte d'urgence sanitaire, et en mobilisant des données issues d'un système d'information (VAC-SI) lui-même mis en place dans des délais extrêmement courts. Disposant d'un plus grand recul sur ces données et indicateurs, il apparaît aujourd'hui important d'en explorer les limites respectives et de mener des analyses complémentaires.

L'objet de ce dossier est ainsi d'explicitier les différences méthodologiques qui expliquent les écarts plus ou moins importants entre ces quatre sources, afin d'identifier des améliorations possibles des indicateurs publiés en *open data*, mais aussi de proposer des estimations alternatives.

¹ Ce système d'information est créé par le décret n°2020-1690 du 25 décembre 2020 autorisant la création d'un traitement de données à caractère personnel relatif aux vaccinations contre le Covid-19

² Étant donné que la population française est majoritairement vaccinée, on met plutôt en avant les taux de non-vaccinés qui sont proportionnellement plus sensibles aux changements de méthode. Les taux de vaccinés s'en déduisent directement.

LES DONNÉES MOBILISÉES

VAC-SI et les données sur la vaccination contre le Covid-19

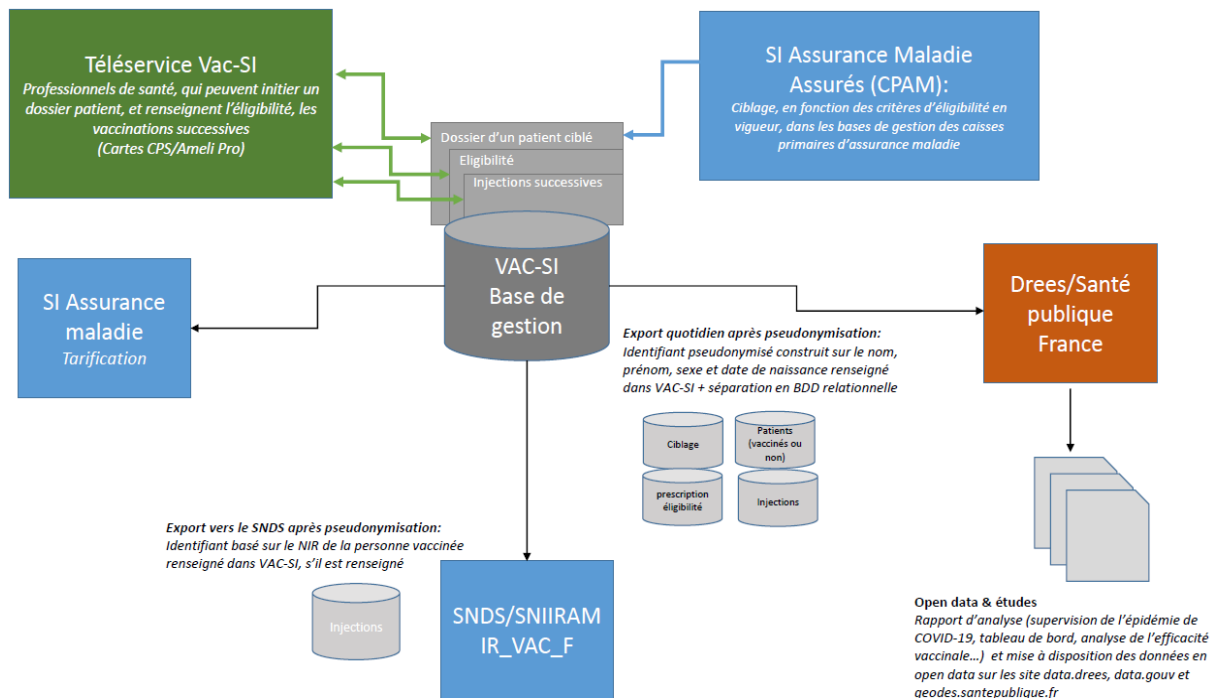
VAC-SI est le système d'information qui recense l'ensemble des personnes ayant reçu au moins une dose de vaccin en France³, ainsi que les personnes ciblées au préalable par les caisses primaires d'assurance maladie (CPAM) comme éligibles à la vaccination. Ainsi, l'ensemble des personnes éligibles à la vaccination sur la base de la stratégie vaccinale définie par le gouvernement sont potentiellement concernées.

Dès lors que l'ensemble de la population adulte sans restriction est *in fine* devenue éligible à la vaccination, la population adulte couverte par ce référentiel doit en théorie recouvrir toute la population adhérente à une caisse primaire d'assurance maladie, et même au-delà en cas de délai de mise à jour des bases de bénéficiaires (personne partie à l'étranger ou décédée depuis son inscription à une CPAM, par exemple). Cette base de données exhaustive, alimentée quotidiennement et constituée pour assurer le déploiement et le suivi en temps réel de la campagne de vaccination, a été mise en place dans des délais particulièrement courts. Il reste néanmoins fort probable que le ciblage ait été imparfait et que certaines personnes ayant droit à l'assurance maladie n'aient pas eu de dossier patient créé en amont de leur vaccination.

La DREES et Santé publique France sont destinataires des bases pseudonymisées recensant les patients ciblés lors de la campagne de vaccination (campagnes successives, avec des populations cibles de tailles croissantes) et/ou identifiés comme éligibles par un professionnel de santé, ainsi que les caractéristiques de leur parcours vaccinal (dès lors que le patient a bénéficié d'une première injection en France) [graphique 1].

Dans la suite, nous appellerons référentiel VAC-SI (ou juste VAC-SI) l'ensemble des pseudonymes issus du ciblage ou de la vaccination de la base de données VAC-SI.

Graphique 1 • VAC-SI, la base de gestion et les exports à finalité statistique



Par ailleurs, la CNAM alimente le Système national des données de santé (SNDS) d'une table, IR_VAC_F, recensant l'ensemble des injections issues de VAC-SI, et appairable aux autres données du SNDS via un pseudonyme construit à partir du NIR du bénéficiaire lorsque ce dernier existe dans la base VAC-SI. Cette table est mise à jour toutes les trois semaines.

³ Qu'elles aient ou non des droits à l'assurance maladie française.

SI-DEP, les données sur le dépistage du Covid-19

SI-DEP est le système d'information recensant l'ensemble des tests virologiques et sérologiques de dépistage du Covid-19 réalisés en France, qu'ils aient ou non été éligibles à un remboursement par l'Assurance maladie. Ainsi, toute personne ayant été dépistée en France depuis la mise en place de ce système d'information est a priori incluse.

La DREES et Santé publique France sont destinataires de la base SI-DEP, avec un identifiant pseudonymisé de la même façon que la base VAC-SI, permettant un chaînage. Au-delà de la couverture vaccinale, l'appariement VAC-SI/SI-DEP permet d'étudier la couverture immunitaire de la population, désormais hybride, c'est-à-dire de dénombrer les personnes avec un antécédent d'infection et/ou un parcours vaccinal. Bien que n'abordant pas directement cette question, la plupart des limites évoquées dans ce dossier s'étendent à cette problématique.

L'appariement VAC-SI/SI-DEP est ici utilisé pour l'information précieuse qu'il apporte afin de caractériser la population présente dans VAC-SI : une personne dépistée à une date donnée est à la fois présente sur le territoire français et en vie.

■ PLUSIEURS SOURCES MISES À DISPOSITION EN OPEN DATA DÉLIVRENT DES STATISTIQUES SUR LE TAUX DE NON-VACCINÉS

Quatre sources de données publiées par trois institutions différentes permettent d'estimer la part des personnes n'ayant reçu aucune dose de vaccin contre le Covid-19 en France :

- Santé publique France (SpF) [publie quotidiennement](#) une batterie d'indicateurs, dont une série journalière sur le taux de vaccination par âge au niveau national, régional et départemental ;
- la CNAM met à disposition mensuellement trois principaux jeux de données, présentés à la semaine et en cumulé depuis le démarrage de la campagne de vaccination :
 - le [premier](#) présente les statistiques sur le nombre de vaccinés au niveau national, régional et départemental, rapporté à la population Insee ;
 - le [deuxième](#) présente les statistiques au niveau des établissements publics de coopération intercommunale (EPCI), rapportées à la population des « consommateurs » constituée en décembre 2021⁴, c'est-à-dire les personnes ayant eu au moins une dépense de santé remboursée ou un séjour dans un établissement de santé en 2021 ;
 - le [troisième](#) présente des statistiques de vaccination par pathologie au niveau national, régional et départemental, il ne sera pas considéré par la suite car il concerne des populations spécifiques ;
- enfin la DREES [publie bi-mensuellement \(après avoir publié toutes les semaines jusqu'à fin mai 2022\)](#) les statuts vaccinaux des personnes testées et des personnes hospitalisées au niveau national et régional, à partir du croisement entre SI-DEP (tests), SI-VIC (hospitalisations) et VAC-SI (vaccination) en présentant également l'effectif de ces statuts en population générale.

Les trois sources rapportant la population vaccinée à la population Insee (SpF, DREES et données départementales de la CNAM) donnent des taux de non-vaccinés proches (tableau 1) et plus faibles que le taux estimé par la CNAM sur la population des « consommateurs ». Les taux par âge ne sont pas facilement comparables, dans la mesure où les tranches d'âge proposées diffèrent, mais on peut cependant constater des écarts entre les sources, qui peuvent d'ailleurs ne pas aller dans le même sens selon les classes d'âge (graphique 1).

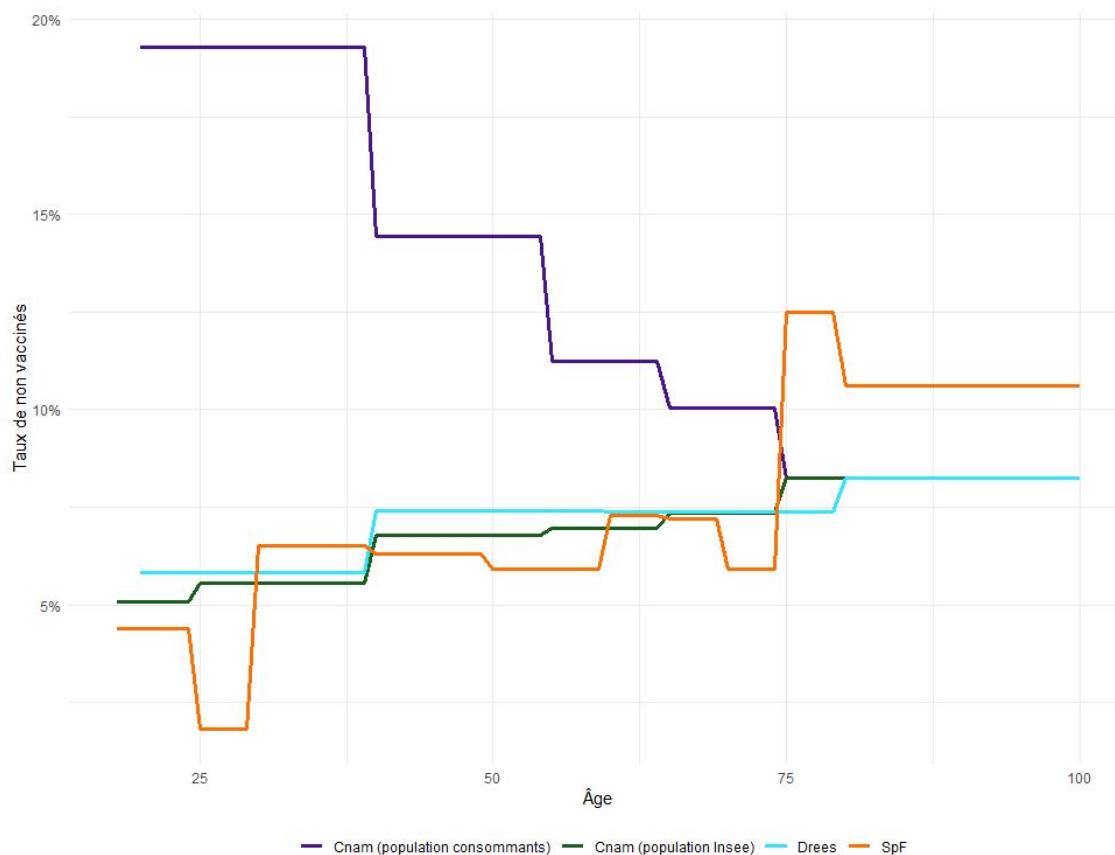
Tableau 1 • Taux de non-vaccinés (en %) selon différentes sources disponibles en open data

Classe d'âge	SpF	CNAM (population Insee)	CNAM (population consommateurs)	DREES
Ensemble	19,6 %	20,2 %	26,0 %	21,0 %
18 ans ou plus	6,5 %	6,6 %		
20 ans ou plus			13,9 %	7,0 %

Source > SpF (data.gouv.fr), CNAM (datavaccin-covid.ameli.fr) et DREES (data.drees.solidarites-sante.gouv.fr), données en date du 14 août 2022, extraites le 26 septembre 2022.

⁴ Ces données portaient initialement sur la population « consommateurs 2020 », elles ont été mises à jour en population « consommateurs 2021 » le 23 septembre 2022.

Graphique 1 • Taux de non-vaccinés par tranches d'âge selon différentes sources disponibles en open data



Source > SpF (data.gouv.fr), CNAM (datavaccin-covid.ameli.fr) et DREES (data.drees.solidarites-sante.gouv.fr), données en date du 14 août 2022, extraites le 26 septembre 2022.

Ces écarts s'expliquent par des choix méthodologiques différents opérés par chaque institution sur le numérateur (nombre de vaccinés – ou de non vaccinés) et le dénominateur (la population de référence), choix qui ont été contraints par les données disponibles, la fréquence attendue de publication des indicateurs et leurs niveaux de déclinaison territoriale (cf *infra*).

Concernant le numérateur :

- d'une part, l'âge retenu peut être soit l'âge au moment de la première injection de vaccin, soit l'âge à une date fixe (1^{er} janvier d'une année par exemple), soit l'âge courant. Dans VAC-SI, sont connus l'année de naissance du patient et son âge lors de la première injection. Dans le SNDS, on dispose du mois et de l'année de naissance. Ainsi, les données produites par SpF reposent sur l'âge au moment de la première injection, les données départementales de la CNAM sur l'âge au 1^{er} janvier 2022⁵, tandis que les données par EPCI de la CNAM considèrent l'âge courant des individus. Pour la DREES, la donnée source utilisée est l'âge construit à partir de l'année de naissance puis une interpolation permet de présenter les résultats en âge courant ;
- d'autre part, le nombre de personnes comptées comme vaccinées peut être « brut » (décompte issu du système d'information) ou tenir compte d'une correction pour la mortalité. Celle-ci peut être appliquée individuellement à partir de la date de décès (c'est le cas pour les données par EPCI de la CNAM fondées sur les consommateurs identifiés dans le SNDS) ou de façon agrégée à partir des taux de mortalité de l'Insee et en exploitant l'information sur les dates de vaccination et de test (c'est la méthode retenue par la DREES). Il n'est pas possible de traiter l'ensemble de la mortalité au niveau individuel à

⁵ La CNAM a effectué fin septembre 2022 une actualisation de la population et du calcul des âges sur l'ensemble des données, c'est-à-dire sur l'ensemble des indicateurs actualisés depuis fin décembre 2020. Pour le calcul des taux de vaccination, le dénominateur est la population recensée par l'Insee au 1^{er} janvier 2022 (et non plus au 1^{er} janvier 2020) et le numérateur correspond à l'ensemble des personnes ayant reçu une injection après exclusion des décédés de 2021 tels que signalés par les CPAM. L'âge pour le numérateur est également calculé au 1^{er} janvier 2022.

partir de la base VAC-SI car l'information sur les décès des personnes vaccinées de VAC-SI ne remonte pas « en continu », et n'est pas transmise automatiquement à SpF et à la DREES. Ainsi, les données publiées par SpF ne peuvent pas prendre en compte la mortalité et conservent donc de fait des personnes décédées depuis qu'elles ont été vaccinées. Cependant des travaux sont en cours pour intégrer les informations sur le statut vital des personnes identifiées dans VAC-SI dont dispose la CNAM. Pour les données départementales de la CNAM les décès ne sont pris en compte que jusqu'au 31 décembre 2021 (par cohérence avec le dénominateur, voir infra) ;

- enfin, on peut choisir (ce qui n'est possible qu'à partir du SNDS, et donc uniquement dans les données de la CNAM présentées par EPCI) de se restreindre aux individus ayant consommé des soins en 2021 (par cohérence avec le dénominateur, cf. ci-après).

Concernant le dénominateur :

- conceptuellement, il peut s'agir de la population résidente au sens de l'Insee ou de la population des « consommateurs » 2021 (pour les données par EPCI de la CNAM) ;
- lorsqu'il s'agit de la population de l'Insee, elle peut être prise à date fixe (1^{er} janvier d'une année donnée, mise à jour ou non au changement d'année) ou évoluer dans le temps. Pour SpF il s'agit de la population au 1^{er} janvier de l'année en cours. Pour la CNAM, il s'agit de la population au 1^{er} janvier 2022. Pour la DREES, la population est calculée à *chaque date* à partir des populations Insee au 1^{er} janvier 2021 et 2022, par interpolation ou application de taux de mortalité, ce qui revient à estimer une population « courante »⁶.

Le tableau ci-dessous résume les sources de divergence mentionnées ci-dessus.

Tableau 2 • Différences de méthode entre les différentes sources publiées en open data

	SpF	CNAM (données départementales)	CNAM (données par EPCI)	DREES
Source	Santé publique France données de vaccination par âge, région et département (data.gouv.fr)	Données vaccination par tranche d'âge, type de vaccin et département/région (datavaccin-covid.ameli.fr)	Données vaccination par EPCI (datavaccin-co-vid.ameli.fr)	Statut des personnes vaccinées, testées ou hospitalisées par âge (Data.drees.fr)
Population des vaccinés	Ensemble des vaccinés recensés dans VAC-SI	Ensemble des vaccinés recensés dans VAC-SI	Vaccinés recensés dans VAC-SI et appariés à la population des « consommateurs » du SNDS (voir infra)	Ensemble des vaccinés recensés dans VAC-SI
Variable d'âge retenue	Âge au moment de la première injection	Âge au premier janvier 2022	Âge courant (en mois, observé et actualisé tous les mois)	Âge courant (estimé)
Dénominateur	Population Insee au 1 ^{er} janvier de l'année concernée	Population Insee au 1 ^{er} janvier 2022	Population des « consommateurs » (ayant bénéficié d'au moins une prestation en 2021 et/ou ayant séjourné au moins une fois dans un établissement de santé public ou privé en 2021, et résidant en France), à laquelle appartiennent plus de 95 % des vaccinés. Âge courant	Population à chaque date estimée à partir des populations Insee et des taux de mortalité

⁶ Plus précisément, l'interpolation entre 2021 et 2022 permet de prendre en compte tous les flux de population (entrée et sortie du territoire, décès), tandis que pour l'année 2022 seule la mortalité est prise en compte.

			(en mois, observé et actualisé tous les mois)	
Cohérence âge numérateur / dénominateur	Non	Oui	Oui (les populations au numérateur et au dénominateur sont cohérentes car toutes deux issues de la population des consommateurs)	Oui
Prise en compte de la mortalité	Non	Partielle : prise en compte des décès jusqu'au 31 décembre 2021	Oui (au numérateur et au dénominateur)	Oui (au numérateur et au dénominateur)

Certains choix méthodologiques ont été contraints : accès ou non à certaines données, périodicité de mise à jour des données d'entrée, possibilités de croisement avec d'autres sources, périodicité de mise à jour des données publiées. Certains choix méthodologiques peuvent également être contraints par la nécessité de produire des indicateurs territoriaux plus fins : distinction entre le niveau national ou régional et le niveau départemental (ex : correction de la mortalité à un niveau agrégé par la DREES pour publier des résultats nationaux et régionaux).

De plus, si les données du SNDS sont potentiellement plus riches (lieu de résidence infra-départemental, âge courant, décès du patient), elles ne permettent pas de réaliser un suivi quotidien car elles sont disponibles sous plusieurs semaines. Seules les données de VAC-SI peuvent être mobilisées dans ce cas. Santé publique France, qui a procédé à des productions quotidiennes pendant les premières années de la pandémie dans l'objectif d'un pilotage fin et réactif de la stratégie de vaccination, s'appuie ainsi *de facto* pour la vaccination sur les données VAC-SI.

Les actualisations des méthodes utilisées attestent par ailleurs de l'engagement des institutions pour améliorer des statistiques mises en place dans un contexte d'urgence.

■ L'ABSENCE DE DONNÉES DISPONIBLES HOMOGENES ENTRE NUMERATEUR ET DENOMINATEUR CONDUIT A UNE PROBABLE SURESTIMATION DU NOMBRE DE VACCINES

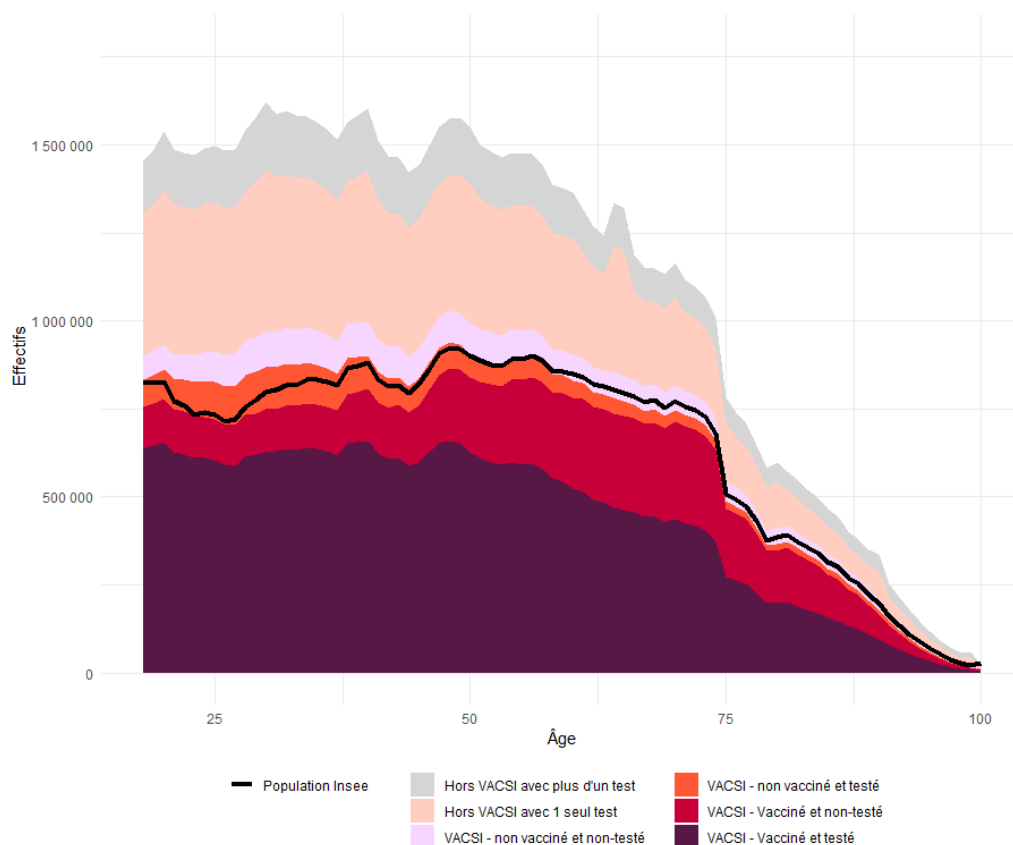
Hormis pour la méthode des « consommateurs » (données de la CNAM présentées par EPCI), les taux de vaccinés publiés rapportent un décompte de population vaccinée estimée à partir de VAC-SI à une population résidente estimée par l'Insee. Or, il n'y a aucune garantie *ex ante* de cohérence entre les deux concepts. En particulier, l'ensemble des dossiers de patients figurant dans VAC-SI ne correspond pas à un référentiel de personnes résidentes en France. Le ciblage effectué dans VAC-SI à partir des bases de bénéficiaires a en effet été large et inclus probablement des doublons, des individus résidant à l'étranger, des individus décédés. L'appariement de VAC-SI et de SI-DEP permet de mettre au jour ces problèmes de cohérence, mais il convient d'abord de présenter les différentes populations qui y figurent⁷. Ces dernières peuvent être classées en quatre groupes, en croisant le fait d'être vacciné et le fait d'avoir réalisé un test de dépistage :

- personnes vaccinées non testées ;
- personnes vaccinées testées ;
- personnes non vaccinées non testées ;
- personnes non vaccinées testées.

À partir des systèmes VAC-SI et SI-DEP, il est en théorie possible de dénombrer les effectifs de population correspondant, peu ou prou, à cette partition. Plus précisément, les deux premières catégories concernant les personnes vaccinées pourraient être estimées par simple croisement entre VAC-SI et SI-DEP. Puis, en assimilant les non-vaccinés aux personnes identifiées dans VAC-SI comme ciblées mais sans injection renseignée, l'appariement VAC-SI-SI-DEP serait également suffisant pour dénombrer les deux dernières catégories.

⁷ Les systèmes d'information à partir desquels ces travaux sont réalisés ne reposent en aucun cas sur un référentiel de population qui correspondrait à la population résidant en France au sens de l'Insee. SI-DEP recense l'ensemble des tests de dépistage réalisés sur le territoire français, qu'ils aient été ou non éligibles à un remboursement de la Sécurité sociale et que les personnes résident ou non en France. De même, VAC-SI recense l'ensemble des doses de vaccin injectées en France, que les personnes concernées aient, ou non, des droits à la Sécurité sociale et qu'elles résident, ou non, en France. Par ailleurs, ce système d'information comporte également des dossiers pour les personnes éligibles (sur la base de listings établis par les CPAM), dites « ciblées », qu'elles aient ou non reçu une dose de vaccin. L'objectif de cette inclusion était de faciliter l'incitation à se faire vacciner : elle avait une finalité purement opérationnelle et nullement statistique (doublons possibles, par exemple).

Graphique 2 • Dénombrement des différents statuts possibles suite à l'appariement de VAC-SI et SI-DEP pour les 18 ans ou plus



Note > Source : Insee. VAC-SI, SI-DEP, exploitations DREES. Comparaison, au 28 juin 2022, des effectifs décomptés via les appariements VAC-SI-SI-DEP avec la population résidente estimée par l'INSEE au 1^{er} janvier 2021.

Cependant, l'appariement entre les bases est imparfait. Il est en effet fondé sur un pseudonyme construit à partir de quatre traits d'identité (nom, prénom, sexe et date de naissance) et non pas sur le numéro d'identification au répertoire (NIR), ou numéro de sécurité sociale, identifiant unique utilisé notamment par l'Assurance maladie pour la délivrance des cartes vitales, qui permettrait un appariement exact. Une simple erreur de saisie dans le nom, ou la saisie du nom marital au lieu du nom de naissance, voire une date de naissance manquante, conduisent donc à des pseudonymes différents dans les deux bases pour une même personne. Par ailleurs, l'algorithme de pseudonymisation à partir des quatre traits d'identité a évolué en mars 2021, pour améliorer la normalisation de certains caractères spéciaux. Enfin, depuis décembre 2021, chaque test saisi dans SI-DEP est transmis au « téléservice INS » qui vise notamment à attribuer un NIR à partir des traits saisis (environ 85 % de succès), et des traits d'identités normalisés en cas de succès. Il résulte donc de ces caractéristiques de l'algorithme de pseudonymisation que l'appariement entre SI-DEP et VAC-SI est imparfait⁸. Dès lors :

- il est probable qu'une partie de la population identifiée comme vaccinée sans test ait en fait réalisé un ou plusieurs tests, mais pour lesquels l'appariement n'a pas fonctionné ;
- il en est de même pour les personnes identifiées comme non vaccinées sans test ;
- il est probable qu'une partie de la population ayant réalisé un ou plusieurs tests non appariés à VAC-SI corresponde à des doublons de pseudonyme ;

Ainsi, à partir de la population ciblée de VAC-SI, complétée de l'information issue de SI-DEP, on dénombre en réalité les catégories suivantes :

- pseudonymes vaccinés et testés (présents dans SI-DEP),
 - chaque pseudonyme correspondrait à un unique individu vacciné et testé ;
- pseudonymes vaccinés et non testés (absents de SI-DEP),
 - chaque pseudonyme correspondrait à un unique individu vacciné, mais à cause des défauts d'appariement une partie a en réalité été testée sans qu'il soit à ce stade possible d'en déterminer l'ampleur ;

⁸ C'est le cas aussi pour l'appariement de ces bases avec SI-VIC pour les données hospitalières.

- pseudonymes dans le référentiel VAC-SI, non vaccinés, et testés (présents dans SI-DEP),
 - chaque pseudonyme correspondrait à un unique individu non vacciné et testé ;
- pseudonymes dans le référentiel VAC-SI, non vaccinés et non testés (absents de SI-DEP),
 - à cause des défauts d'appariement une partie a en réalité été testé, sans qu'il soit possible à ce stade d'en déterminer l'ampleur ;
 - à cause du ciblage potentiellement plus large que la population résidente, une partie correspondrait à des individus non résidents.

On ajoute la catégorie des pseudonymes testés mais absents du référentiel VAC-SI, dont on ne connaît donc pas le statut vaccinal, au découpage précédent, séparée en deux selon qu'un ou plusieurs tests sont observés pour les individus concernés :

- pseudonymes associés à un seul test (resp. plusieurs tests) dans SI-DEP mais non retrouvés dans VAC-SI (« testé non apparié – un seul test » et « testé non apparié – plusieurs tests »).

Ces deux dernières catégories proviennent soit d'un défaut d'appariement entre VAC-SI et SI-DEP, soit du fait que des personnes hors du référentiel VAC-SI peuvent avoir réalisé un ou plusieurs tests en France (voyageurs à l'aéroport par exemple).

Le décompte ainsi proposé conduit au graphique 2, où le trait noir correspond à la population résidente au sens de l'Insee (au 1^{er} janvier 2021). Si l'ensemble des tests et vaccins étaient réalisés par des personnes incluses dans la population résidente de l'Insee, s'il n'y avait pas de doublons et si toute personne résidant en France avait été soit testée au moins une fois, soit vaccinée, soit identifiée comme éligible à la vaccination dans les remontées des CPAM à la CNAM, le cumul des différentes catégories devrait être au plus égal à la population résidente de l'Insee.

Que le nombre total de pseudonymes par âge excède très nettement la population Insee révèle l'ampleur des défauts d'appariement sur la population des testés non appariés. À l'exception des pseudonymes correspondant à des personnes testées et non résidentes, les catégories « testé non apparié – un seul test » et « testé non apparié – plusieurs tests » correspondent à des doublons, qui ont donc vocation à être réaffectés dans les autres catégories, selon le vrai statut vaccinal des personnes concernées, faisant grossir les catégories « vacciné/non-vacciné avec ou sans antécédent » et diminuer les catégories « vacciné/non-vacciné non testé ».

Pour autant, même en faisant abstraction de l'ensemble des personnes testées absentes du référentiel VAC-SI (les « testés non appariés »), la population décomptée dans VAC-SI excède encore la population résidente de l'Insee. Ce phénomène est particulièrement important pour les moins de 50 ans. Il est même flagrant pour les plus jeunes, aux alentours de 25 ans, pour lesquels le seul décompte des vaccinés frôle la population Insee (et la dépasse si on lui ajoute les testés positifs non vaccinés).

De ces constats, on peut déduire une probable surestimation du nombre de personnes résidentes en France et vaccinées à partir du décompte brut dans les systèmes d'information, ainsi que du taux de personnes vaccinées dès lors que l'on rapporte ce nombre à la population résidente de l'Insee (SpF, données départementales de la CNAM et de la DREES). Déduire par différence à la population Insee le nombre de personnes non vaccinées revient à obtenir des effectifs nettement inférieurs à ceux décomptés dans la catégorie « VAC-SI, non vaccinés et testés ». Ces personnes ayant un dossier dans une CPAM et étant dépistées dans SI-DEP, il est dès lors difficile d'ignorer leur effectif relativement important.

Prendre en compte les individus décédés nécessite des traitements statistiques complexes

Les bases VAC-SI ne comprennent pas d'information sur les décès éventuels survenus en cours d'année de personnes dénombrées en début d'année ou au moment de leur vaccination. Il serait possible de disposer de cette information additionnelle (mois et année de décès – éventuel) *via* la CNAM, mais outre des difficultés juridiques et techniques en cours de traitement pour la DREES et SpF, se posent des questions sur la fréquence de mise à jour de cette information, ainsi que sur son exhaustivité et les délais de remontée afférents. C'est pourquoi il est sans doute préférable de corriger statistiquement les décomptes de personnes vaccinées en leur appliquant un taux de décès et un vieillissement de population (de source Insee), afin de disposer d'une estimation de la population courante - au moins au niveau national et régional, mais cette méthode pose plus de difficultés pour produire des estimations départementales. Ce n'est pas, à ce jour, le cas des données de Santé publique France (publiées quotidiennement et jusqu'au niveau départemental), ni de celles de la CNAM au niveau départemental fondées sur la population résidente de l'Insee (du moins pour les décès de 2022). En revanche, les données par EPCI produites mensuellement par la CNAM, qui reposent sur la population des « consommateurs », tiennent compte de la mortalité renseignée dans les bases de l'Assurance maladie (données SNDS). Pour leur part, les données issues des appariements publiées par la DREES (à un rythme hebdomadaire jusqu'à mai 2022 et jusqu'au niveau régional) comportent une correction

statistique pour mortalité et vieillissement ; cette correction exploite également l'information issue de la date de vaccination et des tests (qui garantit que la personne concernée est au moins vivante jusqu'à la date de sa vaccination ou d'un dépistage).

Identification des personnes vaccinées hors champ de la population résidente au sens de l'Insee

Par ailleurs, il apparaît dans les bases de vaccination que certaines personnes ont reçu en France une ou plusieurs injections alors qu'elles pourraient ne pas répondre au critère de la population résidente en France. Cela pose, en termes de statistiques, un problème pour estimer le taux de vaccinés parmi les résidents (tableau 3). De fait, on peut isoler parmi les personnes vaccinées (ou présentes dans le SI) celles qui satisfont au moins un des critères suivants :

- lieu de résidence à l'étranger ou dans les COM (Collectivité d'Outre-mer⁹);
- lieu de vaccination à l'étranger (l'accès à la vaccination ayant été large en France, l'hypothèse que seuls les résidents français à l'étranger ont fait le choix de se vacciner à l'étranger paraît raisonnable) ou dans les COM;
- « ciblage manuel dégradé d'urgence » : ces cas correspondent aux personnes non présentes dans VAC-SI avant leur injection et pour lesquelles aucun NIR n'a été retrouvé suite à une procédure manuelle effectuée lors de la création du dossier au moment de la vaccination.

Tableau 3 • Décomposition des différentes populations de vaccinés par tranches d'âges avant et après filtres

Type de population	0 à 19 ans				20 à 39 ans				40 à 59 ans			
	Vacciné	Vacciné ou positif apparié VACSI	Vacciné ou testé apparié VACSI	VACSI	Vacciné	Vacciné ou positif apparié VACSI	Vacciné ou testé apparié VACSI	VACSI	Vacciné	Vacciné ou positif apparié VACSI	Vacciné ou testé apparié VACSI	VACSI
Ensemble	6 514 661	9 516 460	12 293 488	14 270 102	14 990 723	16 162 792	17 141 834	18 928 374	16 208 026	16 934 044	17 620 227	19 289 385
Ciblage manuel dégradé d'urgence	193 986	194 213	195 015	203 734	597 088	597 966	601 730	635 556	333 799	334 228	336 531	357 521
Département de vaccination étranger	19 493	19 493	19 493	19 493	98 529	98 529	98 529	98 529	65 097	65 097	65 097	65 097
Département de résidence étranger	8 560	12 150	26 739	131 567	25 359	29 024	46 733	187 570	23 630	26 666	44 404	219 168
Part filtrée	3.1%	2.2%	1.8%	2.4%	4.3%	4.0%	3.9%	4.5%	2.3%	2.2%	2.2%	3.0%

Type de population	60 à 79 ans				80 ans ou plus				Tous âges			
	Vacciné	Vacciné ou positif apparié VACSI	Vacciné ou testé apparié VACSI	VACSI	Vacciné	Vacciné ou positif apparié VACSI	Vacciné ou testé apparié VACSI	VACSI	Vacciné	Vacciné ou positif apparié VACSI	Vacciné ou testé apparié VACSI	VACSI
Ensemble	12 886 718	13 143 611	13 563 158	14 837 662	3 609 582	3 693 831	3 849 386	4 429 822	54 209 710	59 450 738	64 468 093	71 755 345
Ciblage manuel dégradé d'urgence	128 339	128 506	129 517	137 703	13 993	14 013	14 113	15 143	1 267 205	1 268 926	1 276 906	1 349 657
Département de vaccination étranger	44 824	44 824	44 824	44 824	7 145	7 145	7 145	7 145	235 088	235 088	235 088	235 088
Département de résidence étranger	18 256	20 037	34 892	256 482	2 485	2 719	4 523	97 025	78 290	90 596	157 291	891 812
Part filtrée	1.3%	1.3%	1.3%	2.8%	0.6%	0.6%	0.6%	2.6%	2.6%	2.4%	2.3%	3.2%

Source > VAC-SI, exploitations DREES. Les dossiers des ciblés manuels dégradés d'urgence devraient tous correspondre à des personnes vaccinées, or ce n'est pas le cas. Une hypothèse serait que certaines vaccinations ont pu être annulées, ou des dossiers initialisés (identification du patient, éligibilité) sans aboutir à une vaccination.

Actuellement, ces personnes sont incluses dans les taux de vaccination diffusés (à l'exception des personnes avec département de vaccination à l'étranger ou dans les COM et des « ciblages manuels dégradés d'urgence » pour les données par EPCI de la CNAM).

Dans la suite de la note, on considère que ces trois populations ne sont pas résidentes en France métropolitaine ou dans les DROM. Il s'agit cependant d'une hypothèse forte pour les « ciblages manuels dégradés d'urgence », car des personnes peuvent être résidentes sur le territoire sans pour autant avoir de droits à l'Assurance maladie, ou disposer des droits ouverts sans qu'une procédure manuelle d'identification au moment de l'enregistrement du dossier ait pu réussir (par exemple : cas d'une personne se présentant dans un centre de vaccination sans connaître son numéro de sécurité sociale, sans présenter sa carte Vitale et dont les traits d'identité communiqués ne donnent aucun écho dans le référentiel CNAM).

⁹ La population résidente en France au sens de l'Insee ici utilisée étant celle de France métropolitaine et des départements d'outre-mer. Les Collectivités d'Outre-Mer comprennent les territoires de : Saint-Barthélemy, Saint-Martin, Saint-Pierre-et-Miquelon, la Polynésie française et les îles Wallis et Futuna.

Estimations alternatives des taux de vaccinés, à partir d'hypothèses visant notamment à améliorer la cohérence entre la population vaccinée et la population de référence

On cherche à savoir si le décompte de personnes vaccinées, une fois apportées les deux corrections sur la mortalité et les non-résidents, est pertinent. Pour cela, il s'agit de vérifier si la population totale ainsi obtenue se rapproche de celle mesurée par l'Insee.

Au départ, on considère la population du référentiel VAC-SI, soit l'ensemble des catégories de population présentées sur le graphique 2, hormis les personnes testées pour lesquelles on n'a pas d'information sur le statut vaccinal (ce qui revient à considérer que ces dernières sont déjà comptées dans les autres catégories, mais non repérées comme telles du fait d'un défaut d'appariement, cf. ci-dessus) : c'est le « total brut ».

Ensuite, la catégorie « non vacciné non testé » est dans un premier temps elle aussi exclue : c'est le « total net ». Sans correction, cette population est de taille équivalente à celle des « non vaccinés testés », ce qui semble peu vraisemblable. On peut penser que, dans une proportion importante, les personnes de cette catégorie se trouvent sur les « listings » des CPAM mais ne sont pas ou peu présentes sur le territoire national au cours de l'année, donc ne font pas partie du champ des personnes recensées par l'Insee¹⁰. Pour appuyer cette hypothèse, on utilise l'appariement des données VAC-SI sur les injections avec les données du SNDS, pour chercher à mieux caractériser la population des non-vaccinés non testés. 60 % des 5,7 millions de bénéficiaires de 20 ans ou plus non vaccinés et sans test remboursé par l'Assurance maladie dans le SNDS¹¹ n'ont aucune autre consommation de soins renseignée entre 2020 et août 2022 (graphique 3). Les 40 % restant, c'est-à-dire les 2,4 millions de personnes sans vaccin ni test retrouvé dans le SNDS mais qui ont eu une autre consommation de soins, représentent environ 30 % de l'ensemble des non-vaccinés avec consommation de soins dans le référentiel des bénéficiaires du SNDS. Il est possible qu'une partie de la population non vaccinée n'ait jamais effectué de test, mais il est probable que cette population, après deux ans et demi d'épidémie, compte tenu de l'instauration du passe sanitaire et de la multiplicité des cas contacts lors de la vague Omicron, soit en réalité de taille plus limitée, ce d'autant que tous les tests n'ont vraisemblablement pas été remontés dans le SNDS¹² (voir graphique en annexe).

On peut donc faire deux hypothèses, donnant des bornes inférieures et supérieures du taux de non-vaccinés :

- une hypothèse « stricte » (correspondant au « total net ») en excluant tous les non-vaccinés non testés, qui conduit à une sous-estimation du taux de non-vaccinés ;
- une hypothèse « large » où on conserve 40 % des non-vaccinés non testés, en s'appuyant sur les estimations issues du SNDS (graphique 3).

¹⁰ Une personne est comptabilisée dans le recensement de la population comme résidente dans une commune si elle y a sa résidence habituelle depuis plus d'un an ou si elle a l'intention d'y demeurer au moins un an.

¹¹ On utilise ici la table IR_VAC_F (une copie de VAC-SI mais comportant un autre pseudonyme) appariée au SNDS, et on récupère les informations de dépistage via les facturations remontées dans le SNIIRAM et non via la table SI-DEP.

¹² En effet, le nombre de tests dans le SNDS est nettement inférieur au nombre de tests comptabilisés dans SI-DEP. Rappelons que l'utilisation de la carte Vitale du patient n'était pas obligatoire pour renseigner SI-DEP, ni pour que le professionnel de santé soit rémunéré. De plus, pour qu'un test soit remonté dans le SNDS, il doit faire l'objet d'un remboursement par l'Assurance maladie. Or, pour la population non vaccinée, les tests de dépistage du Covid-19 ne sont plus remboursés depuis le 15 octobre 2021 sauf exception (par exemple pour les mineurs ou les cas contacts, ou dans certains territoires d'outre-mer). Enfin, il n'est pas possible d'identifier dans le SNDS les tests réalisés dans les établissements publics de santé en cours d'hospitalisation.

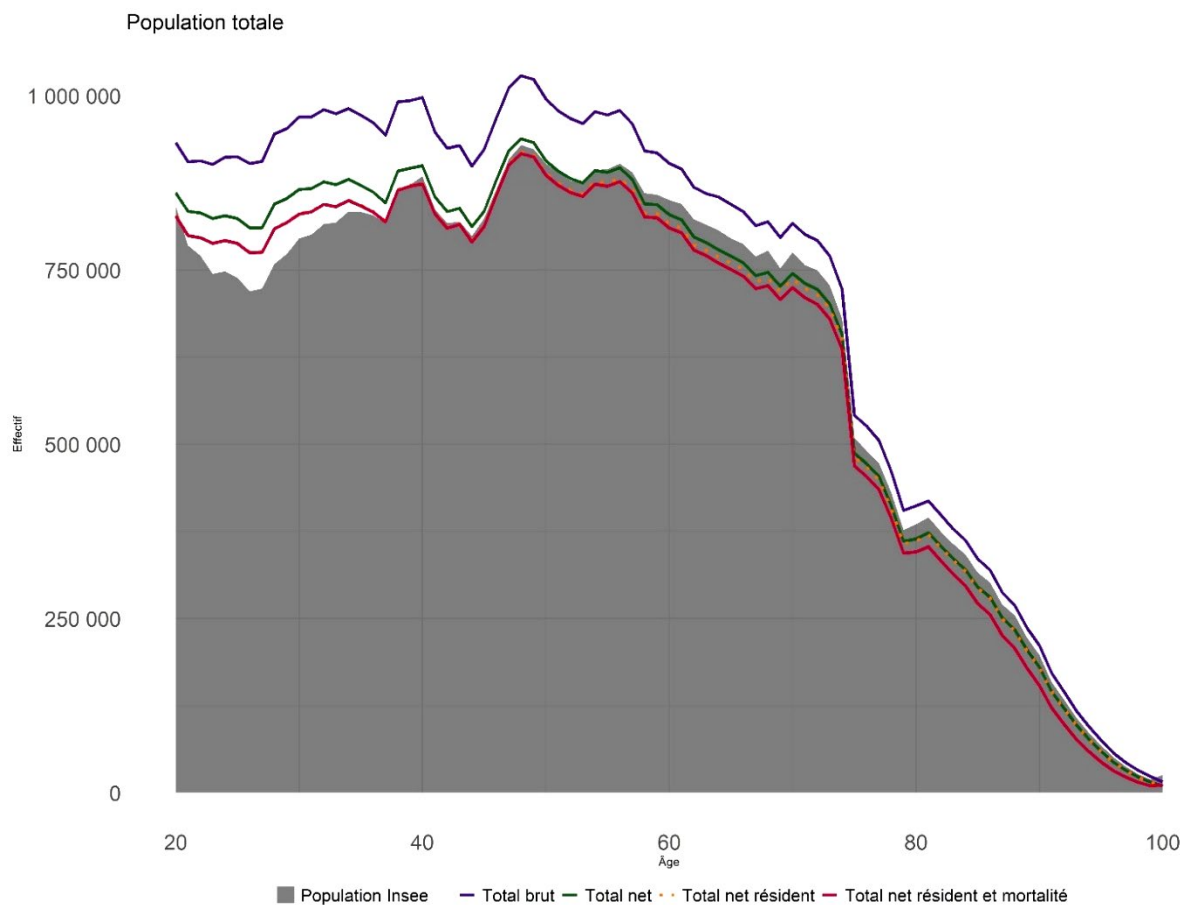
Graphique 3 • Comparaison du nombre de patients non vaccinés et non testés entre l'appariement SI-DEP-VAC-SI et le SNDS



Note > Les données « SNDS » sont issues de l'exploitation de l'appariement de VAC-SI et du SNDS, les consommateurs sont définis ici comme les personnes qui ont eu au moins une consommation entre le 1^{er} janvier 2020 et le 31 août 2022, hors test de dépistage contre le Covid-19.

Les filtres sur la population résidente (« total net résident ») puis la simulation de la mortalité (« total net résident et mortalité ») sont ensuite appliqués au « total net » : la correction sur les résidents joue surtout sur les plus jeunes, tandis que la correction sur la mortalité joue surtout sur les âges plus élevés (graphique 4).

Graphique 4 • Décomptes de population par âge : mise en cohérence de la population « VAC-SI » avec la population Insee



Note > Source : Insee. VAC-SI, SI-DEP, exploitations DREES. Comparaison, au 28 juin 2022, des effectifs décomptés via les appariements VAC-SI-SI-DEP avec la population résidente estimée par l'INSEE au 1^{er} janvier 2021.

Les corrections précédentes permettent de se rapprocher, à partir des bases VAC-SI et SI-DEP, de la population résidente de l'Insee : au total, la population obtenue (50,1 millions de personnes) est inférieure de 1,2 million de personnes à la population résidente de 20 ans ou plus de l'Insee (51,3 millions de personnes), mais il subsiste des écarts par âge (la population demeure surestimée jusqu'à 40 ans environ et devient sous-estimée au-delà de 50 ans environ).

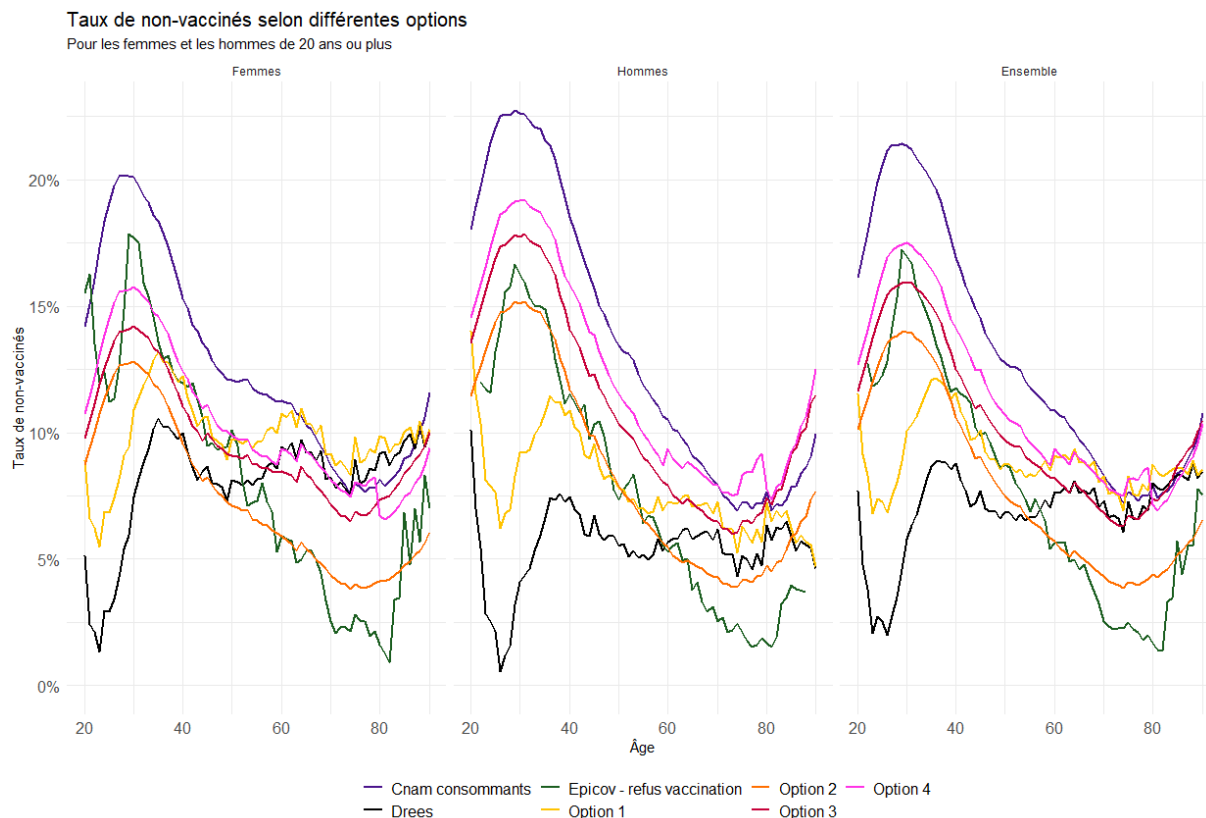
Au final, à partir des appariements des bases VAC-SI et SI-DEP plusieurs options se dégagent pour estimer le taux de vaccinés (et par différence le taux de non-vaccinés) :

- [Option 1] : estimer le taux de vaccinés comme le ratio entre les vaccinés (avec filtre de mortalité et de non-résidents) et la population Insee : ce taux sera surestimé pour les jeunes et sous-estimé pour les plus âgés (donc le taux de non vaccinés sera sous-estimé pour les jeunes et surestimé pour les plus âgés) ;
- [Option 2] : estimer le taux de vaccinés comme le ratio entre les vaccinés (avec filtre de mortalité et de non-résidents) et la population testée et/ou vaccinée (avec les mêmes filtres), en excluant donc les non-vaccinés non testés (hypothèse « stricte ») : le taux sera calculé de façon homogène mais fait disparaître le lien avec la population résidente au sens de l'Insee. Cela revient également à supposer que le taux de vaccination de la population testée et/ou vaccinée est identique à celui de la population résidente, *i.e.* que la population testée et/ou vaccinée constitue un échantillon représentatif de la population résidente ;
- [Option 3] : estimer le taux de vaccinés (avec les mêmes filtres que précédemment) sur la population des seuls individus ayant fait un test et retrouvés dans VAC-SI. Par rapport à l'option précédente, ce taux sera mécaniquement plus faible (donc le taux de non-vaccinés plus élevé), puisqu'on retire la même quantité des numérateur et dénominateur (les vaccinés non testés), mais la population sera plus « homogène ». Là encore, cela revient à supposer que le taux de vaccination de la population testée est identique à celui de la population résidente, *i.e.* que la population testée constitue un échantillon représentatif de la population résidente.

- [Option 4] : estimer le taux de vaccinés comme le ratio entre les vaccinés (avec filtre de mortalité et de non-résidents) et la population testée et/ou vaccinée (avec les mêmes filtres), à laquelle on ajouterait une partie des non-vaccinés non testés, à partir de proportions par âge issues d'exploitation du SNDS (hypothèse « large »)¹³.

Ces quatre options sont représentées sur le graphique suivant.

Graphique 5 • Taux de non-vaccinés par âge et par sexe selon différentes options



Note > Les données sur les « consommateurs 2020 » ont été produites par la CNAM en mars 2022. Les séries option 1, option 2, option 3 et option 4 ainsi que « DREES » sont basées sur les données DREES (appariement SI-DEP-VAC-SI) extraites en août 2022. La courbe « DREES » représente la série des taux de non-vaccinés par âge, plus détaillés que dans l'*open data* diffusé sur le site de la DREES (fourniture des données par âge vingtenal). La série issue de la troisième vague de l'enquête EpiCov (juin 2021) est fondée sur la question liée « au refus de vaccination ». Elle est affichée en moyenne mobile d'ordre 3 pour avoir une courbe moins bruitée.

Au total, le taux de non-vaccinés au 28 juin 2022 pour la population des 20 ans ou plus passerait de 7 % (taux diffusé en *open data* par la DREES) à :

- 9,3 % dans l'option 1 ;
- 8,5 % dans l'option 2 ;
- 11,1 % dans l'option 3 ;
- 11,8 % dans l'option 4.

Il est difficile de privilégier l'une de ces estimations. Cependant, les options 2, 3 et 4 semblent plus cohérentes avec le profil par âge du taux de vaccination issu de l'exploitation de la troisième vague de collecte de l'enquête EpiCov réalisée en juin 2021 et avec celui estimé sur le référentiel des bénéficiaires ayant eu au moins une consommation en 2020 (graphique 5). Plus précisément, dans l'enquête EpiCov il était demandé aux personnes non vaccinées si elles pensaient refuser la vaccination, permettant d'inférer un taux de non-vaccinés « à terme », estimé à 8,4 % sur les 20 ans ou plus, bien qu'il y ait toujours un décalage possible entre vaccination et intention de vaccination. La quatrième vague de cette enquête, actuellement en cours de collecte, permettra, lorsque les résultats seront disponibles, d'avoir un point de comparaison à date. Le taux de non-vaccinés parmi les « consommateurs » de 2020 est, lui, de 13,9 %, donc plus élevé que les options 2, 3 et 4. Cette différence de niveau s'explique par le fait que parmi les « consommateurs » il existe des non-vaccinés non testés (4 %), qui ne sont pas

¹³ On applique également un filtre de mortalité sur cette population des non vaccinés et non testés.

pris en compte dans les options 2 et 3 précédemment présentées, mais également par le fait que l'appariement entre VAC-SI et le SNDS est lui aussi, à ce stade, imparfait, ce qui conduit probablement à surestimer le taux de non-vaccinés.

Conclusion

La mise en place dans des délais très courts d'une base permettant le suivi de la vaccination, mise à jour quotidiennement, ainsi que la publication d'indicateurs journaliers, ont permis d'obtenir des estimations en temps réel de la couverture vaccinale contre le Covid-19.

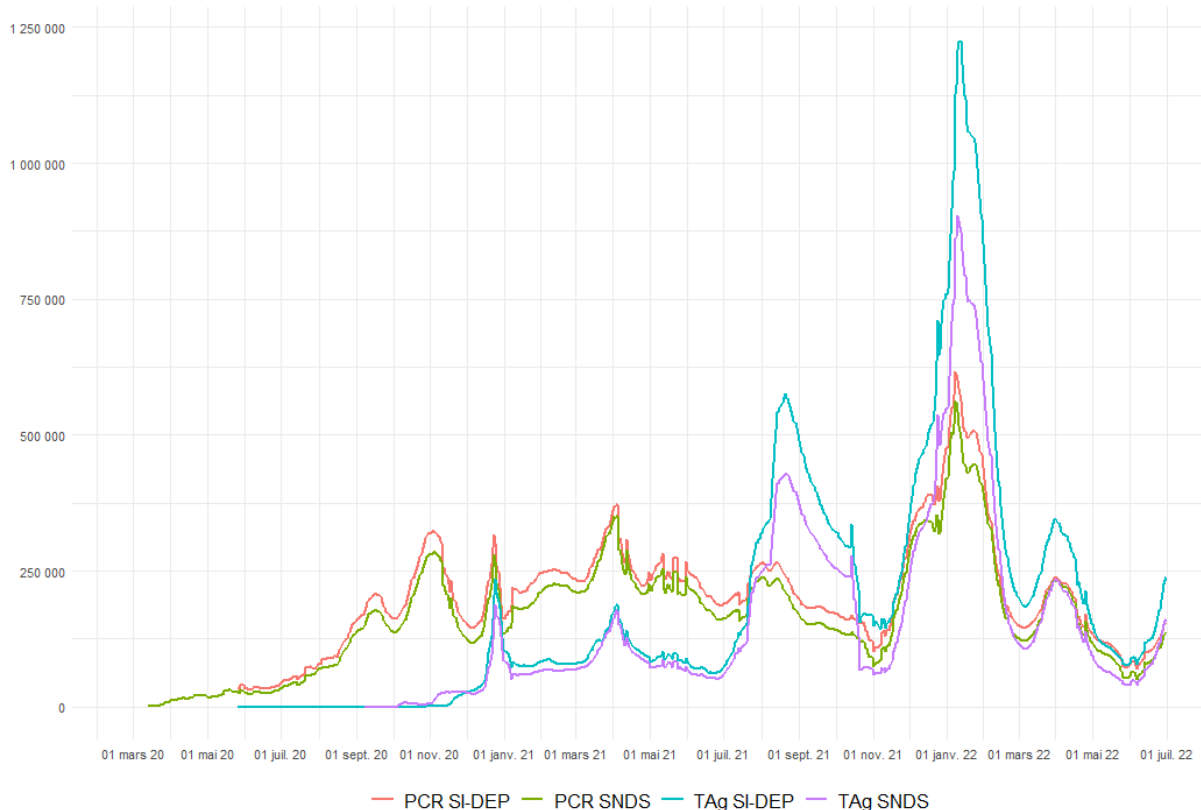
Ces estimations sont pour la plupart convergentes au niveau national sur l'ensemble de la population, mais des différences plus importantes selon les groupes d'âge soulèvent des questions méthodologiques qu'il est possible d'analyser aujourd'hui avec le recul dont on dispose sur les données et indicateurs. En particulier, l'absence d'homogénéité entre le décompte de la population vaccinée, au sens de VAC-SI, et la population de référence, au sens du recensement de l'Insee, conduit probablement à une surestimation et à un biais dans le profil par âge du taux de personnes vaccinées dans la plupart des estimations disponibles en *open data*. Selon les hypothèses retenues sur cette population, le taux de non-vaccinés parmi les personnes de 20 ans ou plus serait probablement compris entre 8 % et 12 %, et le plus élevé autour de 30 ans.

Annexe 1.

Cette annexe présente une comparaison des effectifs de personnes ayant réalisé des tests de dépistage du Covid-19, selon le nombre de tests réalisés, entre SI-DEP et le SNDS. Le nombre de tests PCR et de TAGs apparaît nettement plus faible dans le SNDS.

Graphique 6 • Comparaison journalière des nombres de tests PCR et TAG entre SNDS et SI-DEP

Calculé en moyenne sur 7 jours glissants

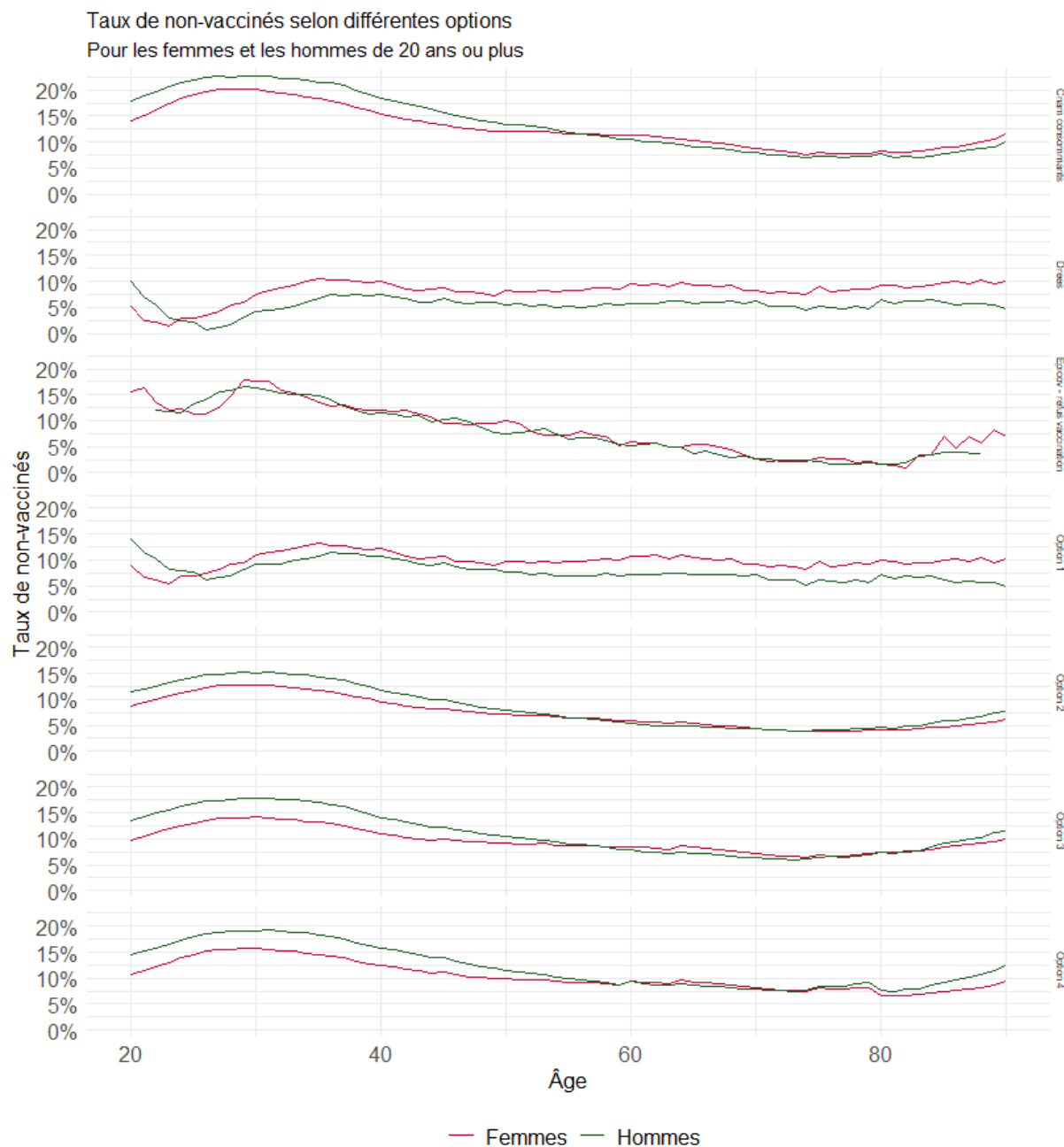


Champ > Le périmètre des tests considérés est distinct selon la source. Dans la source SI-DEP, il s'agit d'une part des tests PCR et antigéniques (TAG) enregistrés dans SI-DEP à l'exclusion des tests PCR salivaires et tests en milieu divers. Dans le DCIR, il s'agit des tests PCR et antigéniques prélevés sur la période ayant donné lieu à un remboursement par l'Assurance maladie et réalisés respectivement en laboratoire (PCR) et en pharmacie (TAG). Pour les tests antigéniques effectués en pharmacie, en l'absence de code spécifique permettant leur repérage, le code prestation utilisé est « préparation magistrale remboursable » avec un filtre sur le tarif pour isoler ces derniers. Les tests réalisés par d'autres professionnels de santé que les pharmaciens (infirmières, médecins généralistes), également remboursés, ne sont pas comptabilisés. Néanmoins, leur prise en compte ne permet pas d'expliquer la totalité de l'écart constaté entre SI-DEP et le DCIR. Enfin, il n'est pas possible d'identifier dans le SNDS les tests réalisés dans les établissements publics de santé en cours d'hospitalisation. Extraction SI-DEP : 26 septembre 2022. Extraction SNDS-DCIR : 9 septembre 2022.

Source > SI-DEP ; SNDS-DCIR ; exploitations DREES.

Annexe 2. Résultats complémentaires

Graphique 7 • Taux de non-vaccinés selon différentes options



Note > Les données sur les « consommateurs 2020 » ont été produites par la **CNAM** en mars 2022. Les séries option 1, option 2, option 3, option 4 ainsi que « DREES » sont basées sur les données Drees (appariement SI-DEP-VAC-SI) extraites en août 2022. La courbe « DREES » représente la série des taux de non-vaccinés par âge, plus détaillés que dans l'*open data* diffusé sur le site de la DREES qui fournit des données par âge vingtennal. La série issue de la troisième vague de l'enquête EpiCov (juin 2021) est basée sur la question liée « au refus de vaccination ». Elle est affichée en moyenne mobile d'ordre 3 pour avoir une courbe moins bruitée.

DREES MÉTHODES

N° 5 • octobre 2022

**Les taux de personnes vaccinées et non vaccinées
contre le Covid-19
en France**

Directeur de la publication

Fabrice Lenglard

Responsable d'édition

Valérie Bauer-Eubriet

ISSN

2495-120X

Ministère des Solidarités et de la Santé
Direction de la recherche, des études, de l'évaluation et des statistiques (DREES)

14 avenue Duquesne - 75 350 paris 07 SP
Retrouvez toutes nos publications sur drees.solidarites-sante.gouv.fr et nos données sur www.data.drees.sante.fr
