

C'est probable, mais la question de contextes spécifiques limitant le bénéfice des avancées pour certains reste posée : les groupes moins favorisés semblent toujours plus exposés à des pathologies létales et invalidantes, avec des probabilités plus faibles d'en guérir ou de faire avec.

On comprend que les systèmes et politiques auront des effets différents sur les inégalités sociales d'espérance de vie en bonne et en mauvaise santé selon qu'ils développent plutôt des protections vis-à-vis de risques létaux (par exemple la lutte contre le tabagisme), des risques sociaux (i.e. lutte contre la pauvreté, accès universels aux soins), des pathologies invalidantes (i.e. réduction des troubles musculosquelettiques) ou des

facteurs qui causent ces pathologies (i.e. conditions de travail). Les systèmes peuvent aussi jouer sur les chances de réduire les risques d'incapacité, parmi les personnes ayant des pathologies invalidantes, en développant et en généralisant les stratégies visant à en réduire l'impact sur l'autonomie, ainsi que la participation sociale (accessibilité des aides techniques, adaptation des environnements professionnels, domestiques et résidentiels...). Un défi sera de mieux appréhender, notamment grâce aux espérances de vie en santé selon le niveau socio-économique, ces liens directs et indirects entre l'évolution des systèmes, l'évolution des années avec et sans problèmes de santé et l'évolution des inégalités sociales en la matière. ●

Dispositif statistique permettant la mesure et l'étude des inégalités sociales de santé : où en est-on en 2021 ?

Le niveau socio-économique est un concept multidimensionnel. Ses composantes individuelles peuvent globalement se résumer autour de l'éducation, de l'emploi, du logement, des revenus et des biens. L'analyse et le suivi des inégalités sociales de santé nécessitent l'utilisation de données socio-économiques individuelles, dont les principales sources sont issues de la statistique publique et des enquêtes représentatives de la population générale.

Afin de pallier les coûts qu'imposent le recueil, l'accès et l'exploitation des données individuelles socio-économiques, la recherche s'est fréquemment tournée vers les données socio-économiques agrégées à l'échelle géographique. Ces mesures écologiques du niveau socio-économique en population générale sont principalement diffusées en routine par l'Insee. Elles sont librement mises à disposition en ligne sans contrepartie. L'utilisation de scores synthétiques mesurant le niveau socio-économique du lieu de résidence comme « proxy-individuel », bien que contestable, est également assez répandue. Nommés historiquement « indices de désavantage social » (*social deprivation indices*), leurs fonctions initiales au Royaume-Uni étaient de mesurer et d'effectuer une veille des inégalités socioterritoriales à l'intention des décideurs de politique publique. Il existe aujourd'hui plusieurs indices de désavantages adaptés aux études en population générale, dont les méthodes de construction diffèrent en fonction de leurs objectifs initiaux [64]. Finalement, de par leurs avantages que sont l'accessibilité, la diffusion en routine et la reproductibilité dans le temps, les mesures écologiques sont principalement utiles à la veille et à l'exploration de

nouvelles hypothèses sur l'étiologie des maladies. Leur utilisation dans une étude individuelle en tant que « proxy » du niveau socio-économique expose les estimations à l'erreur écologique. Ces données sont cependant particulièrement adaptées à l'analyse des effets du contexte de résidence sur la santé, les inégalités sociospatiales pouvant alors aussi être analysées en interrelation avec les expositions environnementales et l'accessibilité aux services suivant des designs multiniveaux. À défaut de recueillir des données socio-économiques individuelles, de nombreux jeux de données de santé ne peuvent être analysés que sous l'angle des caractéristiques du lieu de résidence. C'est le cas notamment du Système national des données de santé (SNDS), qui regroupe l'ensemble des données médico-administratives recueillies à titre obligatoire en France.

Deux types de sources d'informations individuelles permettent d'évaluer les inégalités sociales de santé : les mesures socio-économiques recueillies pour alimenter le même jeu de données que la mesure de santé d'intérêt (comme la profession déclarée sur le bulletin d'état civil de décès), qui sont des données non chaînées, et les données socio-économiques provenant d'une source externe indépendante, qui sont les données chaînées (ou appariées). Si les données non chaînées ont initié au cours de l'histoire la mesure des inégalités sociales de santé, elles sont associées à des biais, notamment lorsque les nombres d'événements de santé doivent être rapportés à des estimations de population pour lesquelles le niveau socio-économique n'est pas recueilli de façon comparable [67]. Une source de biais systématique est induite par la sensibilité du niveau et

Walid Ghosn
Grégoire Rey
Inserm, Centre d'épidémiologie sur les causes médicales de décès (CépiDc), Le Kremlin-Bicêtre

Les références entre crochets renvoient à la Bibliographie générale p. 54.



L'évolution des inégalités sociales relatives au tabagisme en France entre 2000

Les références entre crochets renvoient à la Bibliographie générale p. 54.

Contexte et méthodes

Chaque année, en France, le tabagisme est à l'origine de 75 000 décès [6]. Depuis 2014, la lutte antitabac a été renforcée avec la mise en place de plans nationaux. La prévalence du tabagisme est un indicateur essentiel d'évaluation des politiques menées. Cet indicateur peut être obtenu à partir des données des Baromètres de Santé publique France, conduits annuellement depuis 2016. Ces enquêtes aléatoires représentatives de la population des 18-75 ans résidant en France métropolitaine sont menées par téléphone auprès d'échantillons variant entre 9 000 et 25 000 personnes.

Au début des années 2000, la prévalence du tabagisme quotidien parmi les adultes s'élevait à 30,0 %, niveau autour duquel elle s'est maintenue jusqu'en 2016. Puis, une baisse d'une ampleur inédite a été observée, la prévalence du tabagisme quotidien atteignant 24,0 % en 2019 [60]. Ces évolutions globales masquent de grandes différences selon le niveau socio-économique des personnes.

Entre 2000 et 2016 : accroissement des inégalités sociales en matière de tabagisme

Au début des années 2000, les écarts de prévalence du tabagisme étaient relativement limités selon les caractéristiques socio-économiques. Selon le

niveau d'études, entre 28 % et 34 % des 18-75 ans déclaraient fumer quotidiennement ; selon le niveau de revenu, entre 28 % et 31 % ; et selon la situation professionnelle, entre 34 % pour les actifs occupés et 41 % pour les personnes au chômage et les étudiants.

Entre 2000 et 2016, les inégalités sociales en matière de tabagisme se sont creusées. Parmi le tiers de la population ayant les revenus les plus bas, le tabagisme quotidien a augmenté de 31,4 % à 38,8 %, alors qu'il a diminué parmi le tiers de la population ayant les revenus les plus élevés de 28,3 % à 21,0 % (figure 1). Les mêmes constats sont observés selon le niveau d'études et la situation professionnelle.

Entre 2016 et 2019 : stabilisation des inégalités

Entre 2016 et 2019, la diminution de la prévalence du tabagisme observée est d'une ampleur inédite depuis le début des années 2000, et cette baisse s'observe pour la première fois parmi les populations les plus défavorisées.

Ainsi le tabagisme quotidien a baissé de 38,8 % à 29,8 % parmi les personnes aux plus bas revenus. L'écart avec le tiers de la population ayant les revenus les plus élevés s'élève à 12 points en 2019 (prévalence de 18,2 %) (figure 1). Cette tendance est également observée selon le niveau d'études et la situation professionnelle.

Perspectives

Plusieurs facteurs peuvent expliquer une prévalence du tabagisme plus élevée parmi les populations socialement défavorisées : l'utilisation de la cigarette pour gérer le stress, la difficulté à se projeter dans l'avenir, la méfiance à l'égard des messages de prévention, le déni du risque, une dépendance nicotinique plus importante, une norme sociale en faveur du tabagisme... [61]. De plus, les fumeurs des catégories sociales moins favorisées sont aussi nombreux que les autres à vouloir et à tenter d'arrêter de fumer mais ils y arrivent moins souvent [44].

Malgré une baisse de la prévalence du tabagisme quotidien depuis 2016 parmi les populations moins favorisées, les écarts entre groupes socio-économiques restent marqués. Ces résultats sont cohérents avec ceux retrouvés chez les jeunes de 17 ans : malgré une baisse importante observée, la prévalence du tabagisme reste nettement plus élevée parmi les jeunes en apprentissage ou sortis du système scolaire et parmi les adolescents issus de milieu défavorisé [48]. Même si des effets de désirabilité sociale, potentiellement plus importants ces dernières années, ne peuvent être exclus, les résultats vont néanmoins dans le sens d'une stabilisation des inégalités sociales liées au tabagisme.

Ces résultats incitent à poursuivre et renforcer les actions menées en direction des populations défavorisées. Il est ainsi

Anne Pasquereau
Raphaël Andler
Romain Guignard
Viêt Nguyen-Thanh
Santé publique
France, Direction
de la prévention
et de la promotion
de la santé, unité
Addictions

de la mesure socio-économiques aux événements de santé eux-mêmes, comme une perte de revenus suite à une longue maladie ou l'incapacité du sujet concerné à déclarer lui-même des informations le concernant (en cas de décès notamment). Ces biais ont progressivement poussé les acteurs de la santé publique vers des sources de données socio-économiques externes recueillies indépendamment de l'événement de santé, les bases de données en population générale documentant le niveau socio-économique individuel.

Les bases de données en population générale couvrant les dimensions du niveau socio-économique qui sont produites en routine sont principalement issues de la statistique publique. Dispersées au sein d'institutions et d'organismes divers, certaines informations se retrouvent toutefois centralisées par la Direction

générale des finances publiques (DGFIP) et la Caisse nationale d'assurance vieillesse (Cnav). Contenant une large quantité de données individuelles, elles sont principalement utilisées à des fins de gestion à l'échelle de l'individu, locale et macroéconomique, ce qui justifie l'échange systématique de certaines informations. Parallèlement à ces données, une source d'informations très riche de nature déclarative se retrouve également à l'Insee. En effet, les bulletins individuels issus du recensement de la population sont tirés annuellement pour alimenter l'Échantillon démographique permanent (EDP), qui est construit exclusivement à des fins d'études populationnelles.

Les données de la DGFIP centralisent les revenus et le patrimoine des individus depuis les différentes déclarations d'impôts annuelles obligatoires pour toutes

et 2019

nécessaire de concevoir des actions de prévention adaptées, efficaces et acceptables pour les personnes ayant un plus faible niveau socio-économique. Une offre de services d'aide au sevrage tabagique de proximité (implantation prioritaire dans les zones défavorisées), une démarche proactive des organisations promouvant l'arrêt du tabac (appels ou SMS en direction des populations cibles par exemple) et une co-construc-

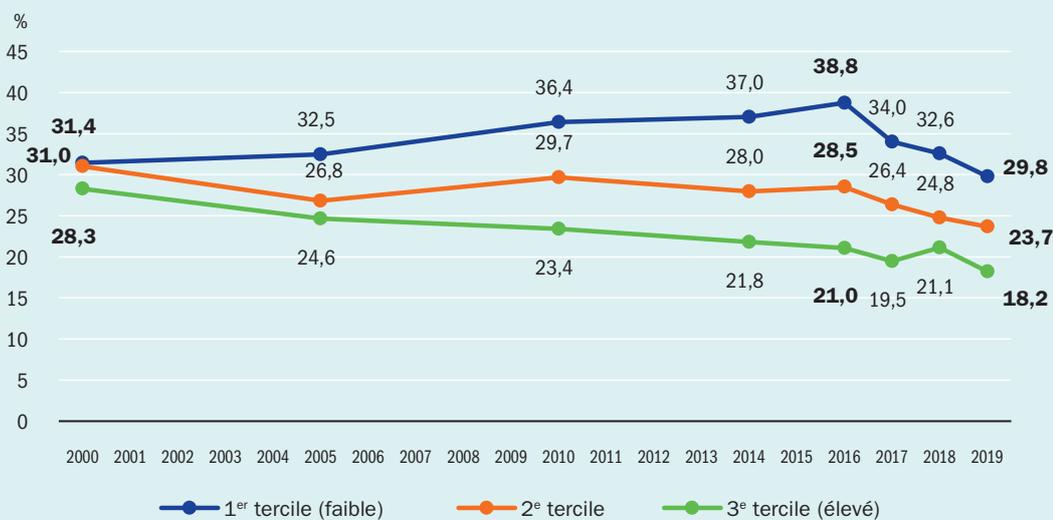
tion des interventions avec les publics cibles pourraient constituer des leviers permettant d'atteindre et d'accompagner davantage les populations précaires dans la réduction du tabagisme. L'efficacité de certaines interventions adaptées a été montrée [37].

Le programme national de réduction du tabagisme 2014-2019 a souligné l'importance de lutter contre les inégalités sociales en matière de tabagisme, puis

le programme national de lutte contre le tabac 2018-2022 a amplifié cette dynamique en créant des objectifs spécifiques pour les populations précaires : actions d'accompagnement spécifiques, accessibilité des aides à l'arrêt, hausses de prix du tabac (intervention qui peut être plus efficace auprès des personnes défavorisées), outils de prévention accessibles à tous dans le cadre des campagnes de marketing social. ●

figure 1

Figure 1 : Prévalence du tabagisme quotidien selon le revenu par unité de consommation parmi les 18-75 ans entre 2000 et 2019 (en pourcentages)



Sources : Baromètres de Santé publique France 2000, 2005, 2010, 2014, 2016, 2017, 2018 et 2019.

personnes résidant sur le territoire national. Actuellement, deux dispositifs, nommés Fideli et FiLoSoFi, produisent en routine des données géographiques sur la distribution des revenus à des échelles infracommunales. Des données individuelles sont également reconstruites pour les individus de l'EDP et leurs proches déclarés dans le même logement. Elles permettent d'obtenir des informations précises sur les caractéristiques de leur logement, le statut d'occupation de celui-ci, ainsi que sur la composition du ménage fiscal. De plus, le rapprochement entre les revenus fiscaux et les prestations sociales transmises par la Cnav et les organismes de prestations sociales (OPS) permet notamment depuis 2011 d'enrichir l'Échantillon démographique permanent d'un revenu disponible monétaire du ménage estimé pour chaque individu. Cependant, les structures collectives

(maisons de retraite, communautés religieuses, cités universitaires, prisons, etc.) ainsi que les personnes les plus précaires (en foyers de travailleurs, etc.), les sans domicile fixe ou sans-abri en sont exclues. Bien que ces données soient largement reconstruites selon des règles de décision pouvant aboutir à des erreurs, les dimensions revenus et logement y sont presque exhaustivement couvertes.

Les données de la Cnav, plus particulièrement celles issues du Système national de gestion des carrières (SNGC), sont également une source d'informations très riche en données socio-économiques. Pour chaque individu sont chaînées à l'aide du Système national de gestion des identités (SNGI) les données nécessaires à l'acquisition de droits et de calcul de pensions, soient des flux d'informations provenant des diffé-



rents OPS (régimes de retraites, Caisse nationale de l'assurance maladie [Cnam], Caisse nationale des allocations familiales [Cnaf], Pôle emploi) et de la Déclaration sociale nominative (DSN). Ces bases sont donc construites avec pour finalité la centralisation des mesures longitudinales caractérisant l'emploi et le revenu individuel de manière très détaillée : nature du contrat de travail, statut, quotité travaillée, salaire, période de chômage et d'inactivité, domaine d'activité de l'entreprise, qualification du poste, arrêts maladie et de travail, toutes prestations sociales concernant l'emploi et la santé, etc.

Les données issues de l'Éducation nationale et de l'Enseignement supérieur sont également une source d'information pertinente pour l'étude des inégalités de santé pour plusieurs raisons. Le niveau d'études atteint entre 25 et 30 ans reste stable au cours de la vie d'un individu, ce qui permet un suivi temporel aisé des inégalités de santé par génération ainsi que leur comparaison entre pays [52]. Par ailleurs, les évaluations des acquis scolaires des élèves permettent d'en dégager des mesures objectives sur des notions clés pour la santé, comme le niveau de lecture à certains âges de la vie en lien avec la littératie en santé.

L'Échantillon démographique permanent présente un statut centralisateur particulier dans le paysage des données permettant de mesurer les inégalités de santé en France. Échantillonné à 1 % de la population générale entre 1968 et 1999 puis à 4 % depuis 2004, il est produit initialement par rapprochement entre les données d'état civil (naissance, mariage, décès) et les bulletins individuels de recensement. Il comporte dans sa version actuelle des questions précises sur le mode de conjugalité, la situation professionnelle ou la recherche d'emploi le cas échéant, les caractéristiques de l'emploi actuel ou antérieur, le niveau d'études ou la formation en cours, sur la localisation du logement actuel et antérieur. Cette base de données individuelles qui comprend les bulletins des recensements de 1968, 1975, 1982, 1990, et 1999, est aujourd'hui consolidée annuellement avec les données issues de l'Enquête annuelle de recensement (EAR). Exclusivement mise à disposition à des fins de recherche, elle a récemment été enrichie par des bases de données externes, notamment par le panel « tous salariés » composé des DSN complétées des fiches de paies des agents de l'État (FPE) de manière exhaustive depuis 2002, et des données sociofiscales depuis 2011 issues des dispositifs FiLoSoFi et Fideli. Ces données chaînées concentrent à ce jour le plus large éventail d'informations socio-économiques individuelles sur un échantillon représentatif de la population.

Un besoin de chaîner les données socio-économiques individuelles avec les données de santé

S'il existe un certain nombre d'enquêtes et d'échantillons créés en routine à partir de données issues de la statistique publique, rares sont les bases de données contenant des mesures objectives du niveau

socio-économique chaînées avec des bases de données de santé de type médico-administratives, de registres ou de cohortes. On pourra citer à titre d'exemple la cohorte Constances [73] et l'Échantillon démographique permanent-Santé [30].

En effet, depuis 2016 un projet nommé EDP-Santé est porté par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees). Il vise à chaîner l'Échantillon démographique permanent au Système national des données de santé. Des problèmes d'ordre réglementaire ont induit des délais exceptionnels jusqu'à la mise à disposition de ces données chaînées. La demande d'autorisation de l'appariement ayant été acceptée en mars 2019, ces données sont maintenant à disposition de la Drees pour évaluer la stratégie nationale de santé 2018-2022. Leur mise à disposition à la recherche est depuis 2020 la priorité annoncée.

Conclusion

Le chaînage de ces informations socio-économiques individuelles, très riches mais réidentifiantes avec des données de santé, reste suspendu à la réglementation et aux procédures institutionnelles associées, à la conformité des infrastructures de mise à disposition et d'exploitation, à la capacité des acteurs à appliquer des méthodes d'appariement efficaces ou *via* l'accès à un référentiel et un identifiant communs.

La mise en place de procédures industrialisées de chaînage entre des données socio-économiques individuelles et des données de santé (SNDS, registres, cohortes) permettrait un gain significatif dans la finesse d'analyse des inégalités interindividuelles. En ce qui concerne le SNDS, une étape préalable consiste néanmoins à identifier l'incidence des maladies à travers des indicateurs, ou « proxys », fondés sur des algorithmes combinant des données de consommation de soins, en veillant à ce que ces proxys ne soient pas biaisés socialement ou spatialement. Cela ne peut se faire aujourd'hui sans le recours à un *gold standard* que représente un recueil dédié associé à un contrôle qualité rigoureux, comme peut l'exercer un registre ou une cohorte. Cela souligne à quel point la mesure des inégalités de santé est dépendante d'un écosystème des données produites par la puissance publique, à la fois au sein des données de santé, mais également avec d'autres sources produites par la statistique publique. Le Health Data Hub (HDH) est une initiative qui va dans ce sens en ce qui concerne les données de santé, même si la dimension santé publique et les données en vie réelle autres que de santé sont aujourd'hui peu présentes. La construction de partenariats forts et structurés entre l'Insee, le Health Data Hub, et les infrastructures de la recherche et de la veille en santé publique, est une nécessité pour avancer sur les procédures et les outils d'accompagnement (formation, documentation) encadrant ces sujets. La communauté de la recherche et de la veille en santé publique est porteuse de ce besoin. ●