

DREES MÉTHODES

N° 20 • février 2025

L'appariement de l'EIR de 2016 et des RI-APA-ASH de 2017

Patrick Aubert (IPP)

L'appariement de l'EIR de 2016 et des RI-APA-ASH de 2017

Patrick Aubert (IPP)

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

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

SOMMAIRE

■ INTRODUCTION	2
■ L'APPARIEMENT DES DONNÉES.....	3
Contenu des bases sources	3
Extraction des données	3
Champ de l'appariement	4
Gestion des doublons dans la table de passage	5
Gestion des non-appariés.....	5
Taux de retrouvés EIR parmi les individus des RI-APA-ASH.....	6
Taux de retrouvés RI-APA-ASH parmi les individus du champ EIR.....	12
Définition des prévalences de référence.....	13
Comparaison des prévalences	14
Ajout d'informations issues des données fiscales.....	16
Sélection des informations fiscales.....	16
Proportion de résidents en établissement et de personnes déclarant des dépenses d'aide à domicile	17
■ IMPUTATIONS ET REDRESSEMENTS COMPLÉMENTAIRES DES VARIABLES	19
Variables relatives à l'APA.....	19
Variables complémentaires sur les retraites.....	22
Résidence en établissement pour personnes âgées.....	23
Redressement du repérage des "communautés" dans l'EDP	23
Couverture de la résidence en établissement par les divers repérages disponibles	25
Comparaison des divers repérages de la résidence en établissement	26
Imputation de l'APA en établissement.....	28
■ PONDÉRATION	35
Pondérations initiales, pour tenir compte de l'échantillonnage de l'EIR.....	35
Examen de la pertinence d'un calage sur marge sur des variables d'intérêt.....	36
■ LISTE DES VARIABLES.....	40
Identifiants et pondérations	40
Variables issues de RI-APA-ASH.....	40
Variables issues de l'EIR.....	42
Variables issues de l'enrichissement de l'EIR avec certaines données fiscales	44
Variables issues de l'EDP	44
■ PREMIERS RÉSULTATS.....	46
Prévalences des prestations d'autonomie selon les caractéristiques des retraités	46
Durées de retraite passées en perte d'autonomie	48
■ PRISE EN MAIN DE LA BASE : PRINCIPAUX POINTS D'ATTENTION.....	53
Accès aux données	53
Informations disponibles	53
Champ et pondérations à utiliser	54
Repérage des populations d'intérêt	55
Repérage des informations imputées	55

■ INTRODUCTION

Ce *DREES Méthodes* documente l'appariement – c'est-à-dire le croisement des données – de la vague 2016 de l'[échantillon interrégimes de retraités \(EIR\)](#) et des [remontées individuelles sur l'APA et l'ASH \(RI-APA-ASH\)](#) de 2017, toutes deux produites par la DREES. Il présente également leur enrichissement avec certaines informations complémentaires issues de l'appariement entre l'EIR de 2016 et les données fiscales et de l'[échantillon démographique permanent \(EDP\)](#) de l'Insee. Le document s'adresse en premier lieu aux chargés d'étude et aux chercheurs qui seront amenés à utiliser les données pour leurs travaux et analyses.

L'appariement de l'EIR et des RI-APA-ASH a pour objectif, de façon générale, de répondre aux questions de recherche qui se situent à l'intersection des problématiques de retraite et d'autonomie des personnes âgées. Il vise ainsi, entre autres, à permettre l'analyse de certains déterminants de long terme de la perte d'autonomie (au sens du recours aux prestations d'aide), notamment ceux liés aux carrières et aux circonstances du départ à la retraite. Il permet, symétriquement, d'étudier les inégalités de durée de retraite en tenant compte de la part de cette durée qui est passée en perte d'autonomie.

De nombreux autres usages sont également possibles, compte tenu de la richesse de chacune des bases. L'appariement permet par exemple de disposer, pour les bénéficiaires de l'APA ou de l'ASH observés dans la base appariée, du suivi de mortalité réalisé pour les individus de l'EIR : il permet donc d'étudier la mortalité de ces bénéficiaires en fonction de leurs caractéristiques. Dans la mesure où les données de l'EIR sont appariées aux données fiscales, il permet aussi de disposer, pour les bénéficiaires des prestations d'autonomie, d'une description des revenus beaucoup plus fine que celle disponible dans les RI-APA-ASH, qui se restreint à l'information dont disposent les conseils départementaux ; cette connaissance plus détaillée des ressources permet donc de simuler des réformes éventuelles des conditions de revenus des bénéficiaires, ou bien de simuler l'impact sur les restes à charge des aides à l'autonomie d'évolutions qui impactent les revenus des personnes âgées.

Ce document détaille la méthode de croisement des données-sources, les taux d'appariement et de couverture de la base appariée par rapport à ces données initiales, les imputations réalisées (notamment celles relatives à la résidence en établissement et au fait d'y percevoir l'APA), les pondérations, ainsi que quelques premiers résultats établis à partir de la base appariée. L'utilisateur pressé pourra se reporter à la dernière partie ("*Prise en main de la base*") pour une présentation plus rapide et une synthèse des principaux points d'attention pour une bonne utilisation des données. Pour une information plus précise, à l'inverse, sur les redressements réalisés, il est également possible de consulter le code-source du programme ayant permis de produire la base et le présent *DREES Méthodes* sur la [page internet d'ouverture des codes de la DREES](#).

Ce travail a été réalisé par l'[institut des politiques publiques \(IPP\)](#), dans le cadre d'un partenariat et d'un financement de la DREES. Il s'inscrit également dans le cadre du [projet KAPPA](#), lauréat du programme prioritaire de recherche (PPR) « Autonomie » de 2022. Il a à ce titre bénéficié d'une aide de l'État gérée par l'agence Nationale de la Recherche au titre de France 2030 portant la référence ANR-22-PAVH-0004.

■ L'APPARIEMENT DES DONNÉES

Cette partie documente le croisement des données issues des deux bases-sources. En particulier, on calcule et discute ici le taux d'appariement entre la vague 2016 de l'EIR et la vague 2017 des RI-APA-ASH.

Contenu des bases sources

Les deux sources appariées constituent les deux bases statistiques de référence de la DREES pour l'analyse des retraites, d'une part, et des prestations d'aide sociale départementale aux personnes âgées, d'autre part.

L'**échantillon interrégimes de retraités (EIR)** est constitué par le rapprochement et l'harmonisation de données administratives issues des systèmes d'information de la quasi-totalité des organismes de retraite obligatoire (régimes de base, régimes spéciaux et régimes complémentaires légalement obligatoires) sur les individus qui perçoivent une pension de droit direct ou de droit dérivé d'une caisse de retraite, ainsi que sur les bénéficiaires d'une allocation du minimum vieillesse. Depuis sa vague de 2016, il inclut également des informations sur les bénéficiaires d'une pension d'invalidité dans l'un des régimes obligatoires. L'EIR contient des informations sur les montants de pension et leurs composantes (avantage principal et éventuelles majorations) dans chaque régime, mais aussi sur les déterminants de ces montants : âges et circonstances de liquidation des droits, durées et points validés pour la retraite, salaire de référence, etc. Les individus retenus dans l'échantillon sont sélectionnés en fonction de leur date de naissance : ils sont nés parmi les premiers jours d'octobre (pour les vagues de l'EIR les plus anciennes) ou de janvier, avril et juillet (pour les vagues plus récentes) des générations ayant dépassé un certain âge. Toutes les générations de retraités ne sont toutefois pas retenues dans l'échantillon : parmi les plus âgées, seules une sur deux ou une sur trois (selon la tranche d'âge) sont observées. La première vague de l'EIR a porté sur la situation des retraités au 31 décembre 1988, et de nouvelles vagues ont ensuite été collectées à un rythme quadriennal. L'EIR de 2016 contient les informations sur un échantillon d'environ 650 000 retraités ou bénéficiaires du minimum vieillesse, et d'un peu moins de 40 000 bénéficiaires d'une pension d'invalidité.

Les **remontées individuelles sur l'APA et l'ASH (RI-APA-ASH)** correspondent de même à un recueil de données administratives individuelles (ou « remontées individuelles »), mené auprès des conseils départementaux sur les deux principales aides à l'autonomie ou à l'hébergement qu'ils versent aux personnes âgées : l'allocation personnalisée d'autonomie (APA) et l'aide sociale à l'hébergement (ASH). Les informations recueillies portent sur les caractéristiques sociodémographiques des bénéficiaires, leur niveau de dépendance détaillé ainsi que l'historique des évolutions du niveau dépendance depuis la première demande d'APA, leurs ressources, les montants de leurs plans d'aide et le contenu de ces plans. Les données portant sur la situation en 2017 correspondent à la troisième vague de recueil réalisée par la DREES, après celles portant sur la situation en 2007, puis en 2011. Toutes prestations confondues, les RI-APA-ASH 2017 concernent près de 1,3 million de bénéficiaires. Toutefois, l'information pour une partie des allocataires reste manquante, notamment en ce qui concerne l'APA en établissement, du fait d'une couverture incomplète des systèmes d'information de certains départements. Cette couverture incomplète nécessite la réalisation de traitements statistiques pour rendre la base de données représentative de l'ensemble des bénéficiaires d'aides départementales à l'autonomie, qui seront détaillées dans la suite de ce document.

Extraction des données

On charge d'abord les diverses bases :

- les données de l'EIR, qui détaillent les caractéristiques de toutes les personnes bénéficiaires d'un droit direct de retraite, d'une réversion, ou d'une allocation du minimum vieillesse, sur le champ des personnes dont la date de naissance correspond à l'un des jours de naissance retenus dans l'échantillon,
- les deux "bases statistiques" des RI-APA à domicile et des RI-APA-ASH en établissement, qui portent sur tous les bénéficiaires de l'APA (base domicile) ou d'au moins une prestation parmi l'APA ou l'ASH (base établissement) ayant un droit ouvert fin 2017,
- la "base brute complète" des RI-APA-ASH 2017, plus large que les deux bases statistiques précédentes car elle inclut tous les bénéficiaires de l'APA et/ou de l'ASH au moins un jour en 2017, y compris s'ils n'avaient plus de droit ouvert au titre de ces prestations au 31 décembre. La base inclut également les personnes ayant fait l'objet d'une évaluation en 2017 pour obtenir l'APA et/ou l'ASH mais n'ayant pas été éligibles,

- la table de passage entre les identifiants de l'EIR et ceux des RI-APA-ASH, construite par appariement sur le code statistique non signifiant (CSNS)¹, qui contient *a priori* l'ensemble des identifiants des personnes observées dans l'une et/ou l'autre des bases, ainsi que des indicateurs de qualité du CSNS.

Pour l'EIR, les CSNS ont été fournis pour toutes les personnes échantillonnées dans la vague 2016 (qu'elles soient déjà à la retraite ou non), mais aussi pour les personnes échantillonnées dans l'EIR 2012 et non resélectionnées dans l'EIR 2016 (pour l'essentiel, les personnes nées du 1^{er} au 10 octobre des années 1912, 1943 et 1945, ainsi que des personnes nées à l'étranger dont le mois de naissance est inconnu). On conserve ces dernières, en récupérant les données sur les retraites dans l'EIR 2012 si les personnes sont repérées comme encore en vie au 31 décembre 2016, et en actualisant les montants de pension par application des coefficients de revalorisation par régime (concrètement, on mobilise la base "générationnelle" de l'EIR, qui empile toutes les vagues pour conserver tous les retraités observés au moins une fois dans l'EIR).

Pour les RI-APA-ASH, on conserve tous les bénéficiaires et demandeurs en cours d'année 2017, disponibles dans les "bases brutes" (c'est-à-dire les bases n'ayant fait l'objet d'aucun redressement de la part de la DREES), et pas seulement ceux qui sont encore bénéficiaires fin 2017 (retenus dans les bases statistiques de diffusion construites par la DREES, qui ont fait l'objet de plusieurs étapes de redressements). En effet, la date de référence au 31/12/2016 de l'EIR fait qu'il est davantage pertinent de s'intéresser, lorsqu'on utilise les données EIR-RI APA appariées, aux bénéficiaires de l'APA en janvier 2017 plutôt qu'en décembre (la mortalité élevée des personnes âgées bénéficiaires de l'APA faisant que l'évolution de cette population entre le début et la fin de l'année ne peut pas être négligée). Toutefois, pour les personnes encore en vie et encore bénéficiaires de prestation fin 2017, il reste préférable d'utiliser les données redressées, par nature plus robuste : on fusionne donc les diverses tables, en ne conservant les observations de la "base brute" que pour les personnes qui sont absentes des bases redressées, et on construit une variable *source_ri* qui prend les valeurs "base brute" ou "base redressée" selon la version de la table dont les informations ont été tirées.

On construit par ailleurs une table *RISYNTH* contenant les informations socio-démographiques (sexe et date de naissance) pour les 1 466 502 individus observés dans les RI-APA-ASH. Dans chaque cas, on a systématiquement retenu, si elles existent, les versions des variables après correction par le prestataire de collecte (par exemple, la variable *c_DATNAISS* si elle est différente de la variable *DATNAISS*, etc.).

Les deux bases statistiques relatives aux bénéficiaires fin 2017 contiennent respectivement 617 607 observations (base "domicile") et 326 438 observations (base "établissement"). 3 029 individus sont présents dans ces deux bases, avec une date de naissance identique². Le fait que les deux bases ne soient pas totalement disjointes est lié à la définition du champ, différent selon le type d'APA perçu : personnes payées pour l'APA à domicile au titre de décembre 2017 dans un cas, personnes ayant un droit ouvert au 31 décembre 2017 dans l'autre. Par ailleurs, 510 062 individus ne sont observés ni dans la base statistique des bénéficiaires à domicile, ni dans celle des bénéficiaires en établissement ayant un droit ouvert fin décembre 2017 : on récupère alors les informations démographiques dans la "base brute complète". On fait de même pour la variable *SEXE*.

Signalons que 5 902 observations de la base brute sur les bénéficiaires et demandeurs de l'APA ne sont pas repérées dans la table (brute) contenant les données socio-démographiques des personnes (notamment le sexe et la date de naissance). Environ 90 % d'entre eux résident dans le département 17, et 9 % dans le département 66. Ces observations sont donc supprimées au début des traitements.

On construit ensuite, à partir des dates de naissance des individus observés dans les RI-APA-ASH, des indicatrices *champ_edp* et *champ_eir* d'appartenance aux champs respectifs de l'EDP ou de l'EIR (vagues 2012 ou 2016 pour ce dernier).

Champ de l'appariement

L'EIR et les RI-APA-ASH ne sont pas de même nature en termes de champ et de couverture :

- l'EIR est un échantillon de la population totale, sélectionné sur la base de jours de naissance (critère de sélection que, par la suite, on considérera aléatoire) ; sur ce champ, il est cependant censé couvrir de façon proche de l'exhaustivité les personnes âgées, car il est rare qu'une telle personne ne perçoive ni retraite (de droit direct ou de réversion) ni minimum vieillesse ;

¹ Le CSNS est obtenu par cryptage du numéro d'inscription au répertoire (ou NIR, communément appelé « numéro de Sécurité sociale »). Il est désormais utilisé pour réaliser des [appariements de données au sein du service statistique public](#).

² Notons que dans la base établissement en R, les dates de naissance sont sous un format numérique, qu'il a fallu d'abord convertir en format date. On a initialement pensé que le format numérique s'expliquait par une erreur de transcodage de la date en format SAS, auquel cas la valeur numérique aurait correspondu au nombre de jours depuis le 1/1/1960. Mais il est apparu, en se basant sur la comparaison des variables pour les individus présents dans les deux bases, que c'est par rapport à la date-origine du 1/1/1970 que les dates étaient codées.

- les RI-APA-ASH sont par définition exhaustives (même si, en pratique, certains départements ont pu ne transmettre que des données incomplètes), mais elles ne portent que sur un sous-ensemble de la population des personnes âgées, bénéficiaires de l'APA et/ou de l'ASH.

Pour cette raison, l'absence de l'une ou l'autre base (en laissant de côté à ce stade les absences liées à des problèmes de qualité des données) n'a pas la même interprétation. Une absence de l'EIR est non-signifiante, au sens où elle indique généralement juste qu'une personne est hors champ de l'échantillon compte tenu de sa date de naissance (en négligeant les quelques cas des personnes âgées ne touchant aucune retraite), tandis qu'une absence des RI-APA-ASH constitue en soi une information d'intérêt, puisqu'elle signale que la personne ne perçoit pas l'APA ni l'ASH. Bien entendu, les absences de l'une ou l'autre des bases peuvent aussi signaler un défaut de couverture lors de la constitution de chacune des bases : un tel défaut est *a priori* rare pour l'EIR, mais il est plus fréquent pour les RI-APA-ASH, dans la mesure où certains départements n'ont pas été en mesure de fournir des informations individuelles sur l'ensemble des bénéficiaires des allocations sur leur territoire (cas par exemple des bénéficiaires de l'APA en établissement dans les établissements sous dotation globale, ou des bénéficiaires de l'APA à domicile accompagnés par certains services à domicile couverts par un CPOM).

Le champ pertinent pour l'appariement EIR / RI-APA-ASH sera donc défini comme l'ensemble des personnes de plus de 60 ans résidant en France dont le jour de naissance fait partie des jours EIR, qu'elles soient bénéficiaires d'une prestation d'autonomie ou non. Les non bénéficiaires de prestations d'autonomie sont conservés car ils permettent de comparer les caractéristiques des bénéficiaires à ceux des non-bénéficiaires du même sexe et âge. Le champ inclut aussi les individus de l'échantillon EIR qui ne sont pas encore retraités en 2016 (mais peuvent l'être devenus après cette date). Ces non-retraités peuvent inclure des bénéficiaires d'une pension d'invalidité, que l'on peut retrouver dans le volet "invalidité" de l'EIR 2016.

À ce stade, une difficulté de repérage du champ subsiste, relative à l'identification des personnes résidant en France. La restriction du champ aux résidents est en effet nécessaire, puisque l'APA et l'ASH ne sont versées qu'à des personnes vivant en France. L'EIR concerne en revanche toutes les personnes connues dans les fichiers d'état-civil de l'Insee, y compris des personnes qui ont vécu quelques années en France, mais qui n'y résident plus après 60 ans. Si la personne perçoit une prestation (retraite ou minimum vieillesse) d'au moins une caisse, le lieu de résidence (France ou étranger) est alors connu, et l'inclusion ou non dans le champ est facilement déterminée. Le problème se pose en revanche pour les personnes résidant en France mais ne percevant aucune prestation de l'EIR, dont la résidence en France n'est connue que si elles sont repérées dans les RI-APA-ASH. Les diverses approches les plus évidentes posent chacune problème. En effet, si on restreint le champ aux personnes ayant une prestation dans l'EIR, on exclut tous les résidents en France ne percevant aucune retraite ni aucun minimum vieillesse, et la prévalence de l'APA risque alors d'être sous-estimée si ces derniers ont un risque plus grand d'être dépendants au sens de l'APA (hypothèse vraisemblable, puisque les personnes n'ayant acquis aucun droit propre de retraite incluent entre autres des personnes handicapées n'ayant jamais travaillé). Symétriquement, si on retient dans le champ toutes les personnes ayant une prestation dans l'EIR *OU* percevant l'APA ou l'ASH, les prévalences sont surestimées, car le numérateur du taux de prévalence inclut bien tous les bénéficiaires, mais le dénominateur de ce taux n'inclut pas les personnes sans retraite, sans APA et sans ASH. On reviendra sur ce point par la suite, dans les sections consacrées à la comparaison des prévalences de l'APA et à la pondération de la table appariée.

Gestion des doublons dans la table de passage

La table de passage construite par rapprochement des bases grâce au CSNS contient quelques doublons (très peu nombreux). Il y a 0 individu des RI (identifiant *id_ApaAsh2017*) pour lesquels on retrouve plusieurs identifiants distincts de l'EIR, mais il y a en revanche 140 individus de l'EIR (identifiant *id_EIR2016* dans la table de passage, correspondant à l'identifiant *NEWNORDRE* de l'EIR) pour lesquels on retrouve plusieurs identifiants distincts des RI-APA-ASH.

Dans de tels cas, une première sélection consiste à retenir uniquement les paires d'identifiants pour lesquelles les données socio-démographiques (sexe, année et mois de naissance) sont cohérentes dans les deux bases EIR et RI-APA-ASH. Une seconde sélection consiste à retenir les paires d'identifiants pour lesquelles la qualité du CSNS (dans les RI) est la meilleure. Après cette double restriction, il reste cependant encore 22 individus de l'EIR associés à plusieurs identifiants différents des RI, ayant le même niveau de qualité du CSNS et le même statut de cohérence entre données socio-démographiques de l'une et l'autre base. Vu le faible nombre de cas, on met à valeur manquante les identifiants EIR et on considère ces cas comme non-appariés.

Gestion des non-appariés

L'utilisation du code statistique non signifiant (CSNS) permet d'obtenir un très bon taux d'appariement entre les bases de données, mais celui-ci n'est pas pour autant égal à 100 %, car le CSNS n'a parfois pas pu être codé, ou bien a pu être mal codé, du fait d'informations d'identification (nom, prénom, etc.) erronées dans les bases. Afin de maximiser le taux d'appariement entre les deux bases, on tente donc de compléter l'appariement réalisé avec le

CSNS par un appariement statistique, sur la base d'informations socio-démographiques : sexe, département de résidence, date de naissance et, le cas échéant, année et mois de décès.

En pratique, dans les RI-APA-ASH, on connaît la date de décès entre le 1^{er} janvier 2017 et la date d'extraction des données par les conseils départementaux, qui peut varier d'un département à l'autre (entre 2018 et 2019). Dans l'EIR, on connaît l'année et le mois de décès (grâce aux données d'état-civil de l'Insee), mais on ne connaît le jour de naissance que pour les personnes encore en vie en 2020 ; pour celles décédées avant cette date, le jour de naissance n'est connu que de façon regroupée (par exemple, pour les nés en octobre, et selon la génération : du 1 au 4 octobre, du 5 au 10, ou bien du 5 au 6 et du 7 au 10, et du 11 au 24 octobre). On réalise donc l'appariement statistique en trois temps, avec une partie de la clé d'appariement commune aux trois étapes (sexe, département de résidence, année et mois de naissance) et une partie qui change :

1. Pour les individus repérés comme décédés dans les deux bases : jour de naissance regroupé, année et mois de décès.
2. Pour les individus non repérés comme décédés dans les RI-APA-ASH, mais décédés avant 2020 d'après l'EIR : jour de naissance regroupé.
3. Pour les individus encore en vie en 2020 : jour de naissance au jour près.

Dans chaque base, on ne retient que les observations pour lesquelles il y a une seule personne par clé d'appariement.

L'appariement statistique sur la base des caractéristiques socio-démographiques permet de rapprocher 4 345 personnes entre les deux bases, parmi les 25 859 personnes observées dans les RI-APA-ASH sur le champ de l'EIR et les 1 923 457 personnes observées dans l'EIR (figure 1). Cependant, il n'améliore que de façon marginale l'appariement, car la plupart avaient déjà été rapprochées grâce au CSNS. Seules 83 personnes sont appariées uniquement d'après l'appariement statistique ; cela permet malgré tout de diminuer d'un peu plus de 10 % le nombre d'individus des RI-APA-ASH non retrouvés dans l'EIR. Dans les autres cas, les deux méthodes d'appariement (CSNS et statistique) conduisent bien à rapprocher les mêmes identifiants dans près de 99 % des cas : seuls 57 paires d'identifiants ne sont pas cohérentes selon la méthode d'appariement.

Figure 1 Nombre d'observations selon la situation d'appariement

Situation	Effectif	Proportion
Appariement CSNS uniquement	20 965	1,1
Appariement statistique uniquement	83	0,0
Appariements CSNS et statistique, cohérents	4 205	0,2
Appariements CSNS et statistique, incohérents	57	0,0
Individus EIR non appariés	1 898 147	98,7
Individus RI-APA-ASH non appariés	549	0,0
Total	1 924 006	100,0

Champ > Personnes dans le champ de l'EIR d'après leur date de naissance.

Source > Appariement EIR + RI-APA-ASH, DREES.

Dans toute la suite, l'appariement portera sur la table de passage construite grâce au CSNS et augmentée des paires d'individus rapprochées par l'appariement statistique.

Taux de retrouvés EIR parmi les individus des RI-APA-ASH

On s'intéresse ici aux individus pour lesquels un identifiant RI-APA-ASH est connu, et on calcule la proportion, parmi eux, d'individus retrouvés dans l'échantillon de l'EIR. Ces proportions sont ventilées selon une indicatrice d'appartenance au champ de l'EIR (construite dans les sous-sections précédentes à partir de la date de naissance) et selon l'indicateur de qualité du CSNS dans les RI.

Pour les individus dont la date de naissance est inconnue ou est en dehors du champ de l'EIR, on ne retrouve aucun individu de l'EIR dans la très grande majorité des cas (figure 2). Un identifiant de l'EIR est toutefois retrouvé pour quelques centaines de personnes : il pourrait s'agir d'individus pour lesquelles la qualification comme "hors champ EIR" est en réalité erronée, car la date de naissance d'après les RI serait elle-même erronée. Pour 46 % d'entre eux, la date de naissance dans les RI est en effet le 1^{er} janvier ou le 31 décembre, ce qui pourrait signaler une valeur indiquée par défaut plutôt que la vraie date de naissance.

Parmi les individus qui, sur la base de leur date de naissance dans les RI, font bien partie du champ de l'EIR, 97,5 % sont effectivement retrouvés dans cet échantillon, mais 2,2 % (549 individus) ne le sont pas. Lorsqu'un identifiant EIR est retrouvé, les informations sociodémographiques (sexe, année et mois de naissance) sont cohérentes entre les deux bases dans la quasi-totalité des cas (seuls 70 cas environ sont appariés mais avec des informations sociodémographiques différentes d'une base à l'autre). Par ailleurs, le taux d'individus non-retrouvés est d'autant plus grand que la qualité du CSNS (associé à l'identifiant RI) est mauvaise : il est de moins de 1 % pour les CSNS de meilleure qualité (*qualiteCSNS_ApaAsh* = 1) mais de 32 % pour le CSNS de moins bonne qualité (*qualiteCSNS_ApaAsh* = 7).

Les personnes dont la date de naissance est inconnue (que ce soit dans l'EIR ou dans les RI-APA-ASH) ont généralement des profils particuliers : il s'agit souvent de personnes nées à l'étranger et très précaires. Les exclure de la base pourrait donc induire un biais (même si celui-ci serait *a priori* d'ampleur très limitée vu le faible nombre de cas). **Dans la base finale, on retiendra donc tous les individus appariés, même si leurs informations sociodémographiques ne sont pas cohérentes entre les deux bases et si ces individus ne devraient pas être dans l'EIR d'après leur date de naissance renseignée dans les RI-APA-ASH.** En pratique, pour ces individus, ce seront les sexe et année de naissance de l'EIR qui seront retenus, cette information (issue des données d'état civil de l'Insee, dans lesquelles l'EIR est échantillonné) étant jugée plus fiable.

Figure 2 Taux d'appariement, selon le champ et la qualité du CSNS

champ_eir	qualiteCSNS_ApaAsh	Non retrouvés EIR	Retrouvés EIR, info sociodémographiques incohérentes	Retrouvés EIR, info sociodémographiques cohérentes
Hors champ EIR	Ensemble	1425483 (100%)	351 (0%)	154 (0%)
Champ EIR	1 (meilleure qualité)	22 (0.5%)	4 (0.1%)	4419 (99.4%)
	2	207 (1.1%)	2 (0%)	18533 (98.9%)
	3	83 (10%)	51 (6%)	715 (84%)
	4	45 (21%)	3 (1.4%)	162 (77%)
	5	53 (15%)	5 (1.5%)	286 (83%)
	6	17 (14%)	1 (0.8%)	100 (85%)
	7 (faible qualité)	122 (32%)	7 (1.8%)	251 (66%)
	Ensemble	549 (2.2%)	73 (0.3%)	24466 (97.5%)
Champ inconnu	Ensemble	15228 (98.3%)	266 (1.7%)	

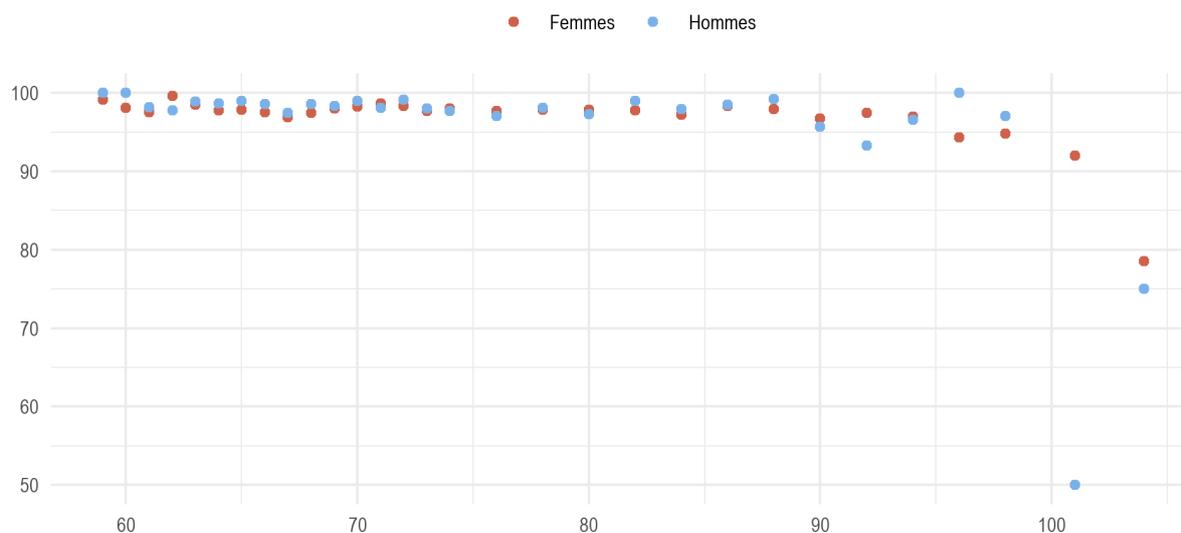
Note > Les individus finalement conservés dans la base appariée correspondent à la totalité de ceux comptabilisés dans les deux dernières colonnes du tableau.

Champ > Personnes pour lesquelles un identifiant RI est retrouvé dans la table de passage.

Source > Appariement EIR + RI-APA-ASH, DREES.

Le graphique suivant détaille, sur le champ des individus dont le jour de naissance d'après les RI-APA-ASH est bien dans le champ de l'EIR, les parts de personnes effectivement retrouvées dans l'EIR, en les ventilant selon l'année de naissance (figure 3). Le taux de retrouvés est généralement supérieur à 95 %, mais il apparaît un peu plus bas pour les générations les plus anciennes (et même nettement plus bas pour celles ayant plus de 100 ans), qui sont, malheureusement, justement celles où les prévalences des prestations d'autonomie sont les plus élevées.

Figure 3 Proportion de retrouvés dans l'EIR parmi les individus observés dans les RI-APA-ASH, selon l'âge (en %)



Note > Âges au 31/12/2016.

Champ > Personnes pour lesquelles un identifiant RI est retrouvé dans la table de passage et dont le jour de naissance est dans le champ de l'EIR.

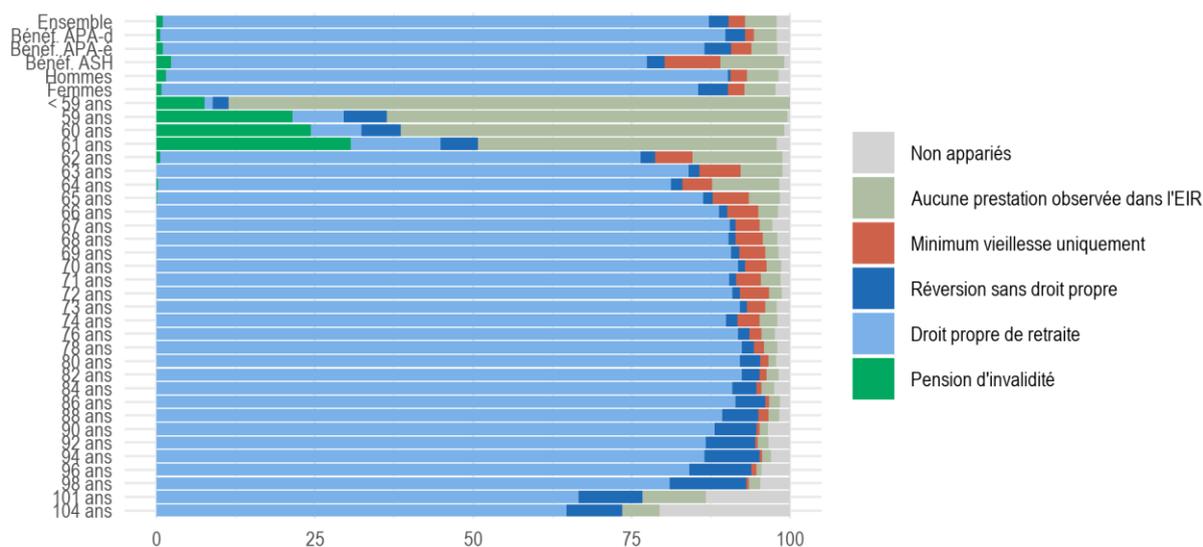
Source > Appariement EIR + RI-APA-ASH, DREES.

Le graphique suivant complète l'analyse en ventilant selon le type de prestation observée dans l'EIR (figure 4). Le fait qu'un identifiant de l'EIR soit retrouvé dans la table de passage signifie en effet que l'individu est retrouvé dans la base de sondage de l'EIR, mais pas forcément qu'une prestation (pension d'invalidité, retraite ou minimum vieillesse) a effectivement été observée dans l'une au moins des caisses participant à l'exercice.

En pratique, une partie importante des personnes appariées qui ont moins de 62 ans fin 2016 n'ont aucune prestation observée dans l'EIR. Il s'agit donc de personnes ayant commencé à percevoir l'APA alors qu'elles n'ont pas encore pu liquider un droit de retraite (direct ou réversion), et qui ne bénéficient pas non plus d'une pension d'invalidité. Le fait que la proportion diminue fortement à partir de l'âge d'ouverture des droits à la retraite (62 ans) et que la part de bénéficiaires du minimum vieillesse uniquement devienne aussi non négligeable à cet âge indique que les bénéficiaires "jeunes" de l'APA sont pour l'essentiel des assurés reconnus inaptes du point de vue des caisses de retraite. Les bénéficiaires de l'APA ayant uniquement le minimum vieillesse à 62 ans sont vraisemblablement d'anciens bénéficiaires de l'AAH.

La part de bénéficiaires de l'APA (ou de l'ASH) n'ayant aucune prestation dans l'EIR, même si elle diminue très fortement après 62 ans, reste toutefois non-négligeable après cet âge. Elle est en particulier plus élevée avant 65 ans qu'elle ne l'est après cet âge : cela correspondrait *a priori* à des personnes handicapées (au sens où le recours à l'APA peut être vu comme la conséquence d'une situation de handicap) qui ne seraient pas reconnues comme inaptes par les caisses de retraite (ce qui les conduirait à attendre l'âge d'annulation de la décote pour liquider leur droit à retraite). La part de bénéficiaires de l'APA ou de l'ASH percevant uniquement le minimum vieillesse (mais n'ayant aucune pension de retraite) apparaît maximale aux âges jeunes (à partir de 62 ans), puis décroît progressivement avec l'âge, pour devenir quasiment nulle après 85 ans. Cette diminution peut s'expliquer par le veuvage des personnes qui n'ont aucun droit propre de retraite, le veuvage étant plus fréquent aux grands âges, mais aussi (et sans doute davantage) par la mortalité nettement plus élevée des personnes qui n'ont pas pu acquérir de droit propre de retraite.

Figure 4 Type de prestations dans l'EIR perçues par les individus observés dans les RI-APA-ASH (en %)



Note > Âges au 31/12/2016. Données non pondérées.

Champ > Personnes observées dans les RI-APA-ASH, et qui soit sont dans le champ de l'EIR d'après leur date de naissance, soit ont un identifiant EIR retrouvé dans la table de passage. Lorsque la date de naissance ou le sexe sont inconnus dans les RI-APA-ASH, on a retenu les valeurs connues dans l'EIR.

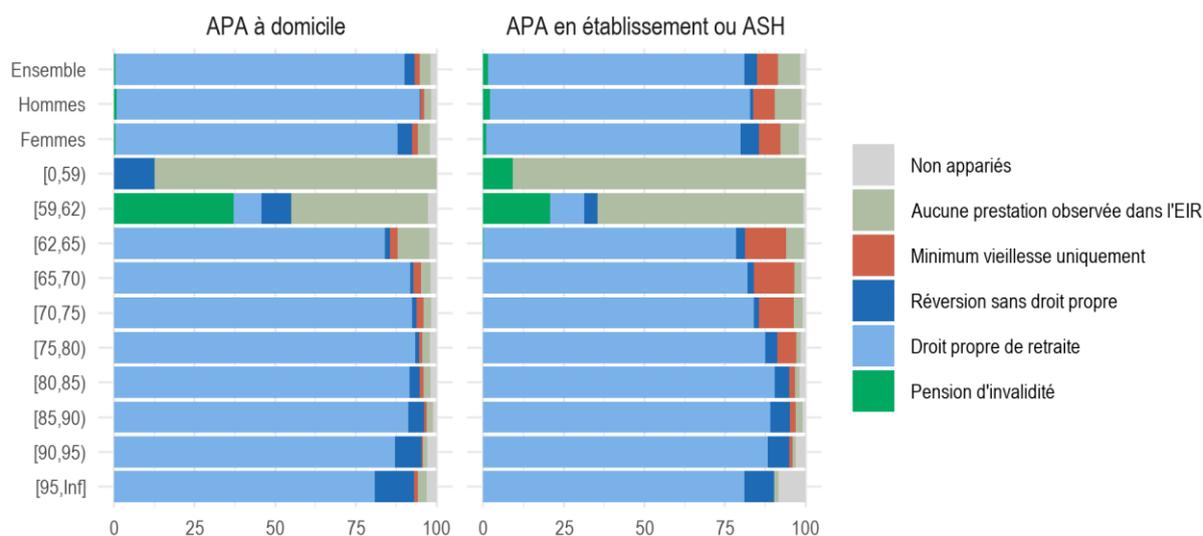
Source > Appariement EIR (2012 et 2016) + RI-APA-ASH (2017), DREES.

Le graphique suivant sépare les bénéficiaires selon le lieu de bénéfice de la prestation APA ou ASH (domicile ou établissement). On se restreint ici aux seuls bénéficiaires ayant un droit ouvert au 31 décembre 2016, soit un champ un peu plus restreint que pour le graphique précédent. Les âges sont regroupés par tranche d'âge quinquennal, sauf aux âges les plus jeunes, où on définit les bornes des tranches d'âges d'après les âges légaux. En toute rigueur, les RI-APA-ASH de 2017 ne permettent pas de repérer les personnes dont le droit aurait été refermé *exactement* le 31/12/2016 : on négligera l'impact de cette restriction de champ dans l'analyse.

Les résultats apparaissent assez similaires, si ce n'est que la proportion de personnes ne percevant que le minimum vieillesse sans aucune pension de retraite apparaît plus élevée parmi les bénéficiaires de l'APA ou de l'ASH en établissement (figure 5). Aux grands âges (à partir de 95 ans), la proportion de non-appariés semble un peu plus élevée parmi les bénéficiaires en établissement qu'à domicile.

Remarquons par ailleurs que certains bénéficiaires de l'APA ou de l'ASH d'après les RI-APA-ASH ont moins de 59 ans fin 2016, donc moins de 60 ans fin 2017. Ces cas posent question, car le bénéfice de l'APA n'est normalement possible qu'après 60 ans : il s'agit soit d'appariements erronés (l'individu EIR avec qui le bénéficiaire de l'APA a été rapproché n'est en réalité pas la bonne personne), soit de cas où la prestation aurait été, par exception, ouverte à une personne en dessous de la limite d'âge (dans le cadre de l'action extralégale des départements, ou parce que le département aurait octroyé la prestation sans attendre de connaître précisément l'âge de la personne).

Figure 5 Type de prestations dans l'EIR perçues par les bénéficiaires de l'APA ou l'ASH au 31 décembre 2016 (en %), selon le type de prestation perçue



Note > Âges au 31/12/2016. Données non pondérées.

Champ > Bénéficiaires ayant un droit ouvert au 31/12/2016 dans les RI-APA-ASH, et qui soit sont dans le champ de l'EIR d'après leur date de naissance, soit ont un identifiant EIR retrouvé dans la table de passage.

Source > Appariement EIR (2012 et 2016) + RI-APA-ASH (2017), DREES.

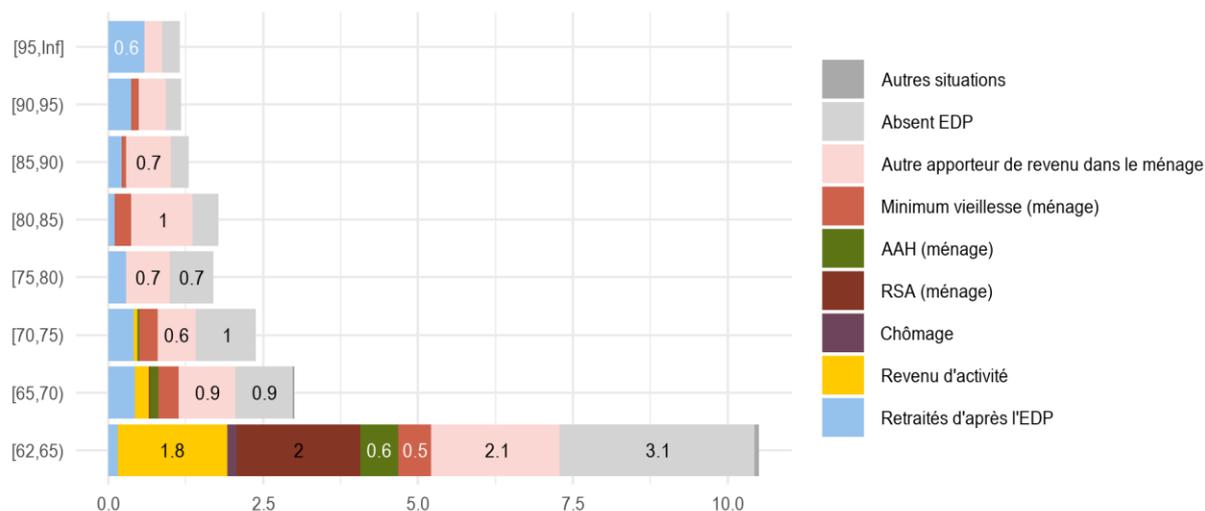
Afin de mieux comprendre la situation des bénéficiaires de l'APA ou de l'ASH qui ne perçoivent aucune prestation d'après l'EIR, on se restreint au champ des personnes dont le jour de naissance fait partie de ceux retenus pour l'échantillon démographique permanent (EDP), et on observe les revenus connus dans cette dernière base (figure 6). L'avantage de l'EDP est en effet de disposer d'une information au niveau du ménage, à partir des sources fiscales. Comme pour l'EIR, on retient dans l'EDP les informations sur les revenus de l'année 2016.

Une partie des personnes pour lesquelles aucune prestation personnelle n'est observée dans l'EIR sont également absentes des données fiscales de l'EDP. Leur proportion est cependant faible : ces personnes représentent moins de 1 % des bénéficiaires de l'APA ou de l'ASH, quelle que soit la tranche d'âge à partir de 65 ans. Il pourrait s'agir de personnes résidant en établissement, puisque l'information fiscale sur les revenus de ces personnes n'est pas systématiquement remontée dans l'EDP sur ce sous-champ.

Une autre partie des personnes sans prestation dans l'EIR perçoit par ailleurs une retraite personnelle ou bien le minimum vieillesse d'après l'EDP. Il pourrait s'agir d'un problème de défaut de couverture dans l'EIR, mais d'autres explications sont également possibles. Par exemple, une personne qui perçoit uniquement des pensions de retraite versées par des régimes étrangers n'est pas observée dans l'EIR (qui ne porte que sur les régimes de retraite français) mais doit bien déclarer ses pensions au fisc. Les personnes identifiées comme bénéficiant du minimum vieillesse pourraient être des conjoints d'allocataires de cette prestation : l'EIR ne repère en effet que les allocataires eux-mêmes, tandis que l'EDP repère à la fois les allocataires et leur conjoint (et toutes les personnes qui vivent avec elles dans le même ménage), puisque le bénéfice du minimum vieillesse est défini au niveau du ménage.

Pour le reste, les bénéficiaires de l'APA ou l'ASH qui ne perçoivent pas de prestation connue dans l'EIR ne perçoivent pas non plus de retraite ni de minimum vieillesse d'après l'EDP, mais vivent dans un ménage comptant au moins un apporteur de ressource. Aux âges les plus jeunes (mais après 62 ans — les moins de 62 ans n'étant pas représentés sur le graphique par souci de lisibilité), certains bénéficiaires ont également encore des revenus du travail, ou perçoivent des minima sociaux d'âge actif (RSA ou AAH).

Figure 6 Situation dans l'EDP des individus observés dans les RI-APA-ASH et appariés mais sans prestation retrouvée dans l'EIR (en %)



Note > Âges au 31/12/2016. Données non pondérées. Les revenus d'activité, allocations chômage et retraites sont mesurées pour l'individu concerné seulement. Les minimum sociaux (RSA, AAH, minimum vieillesse) sont mesurés au niveau du ménage : l'allocataire peut donc être le conjoint de l'individu concerné.

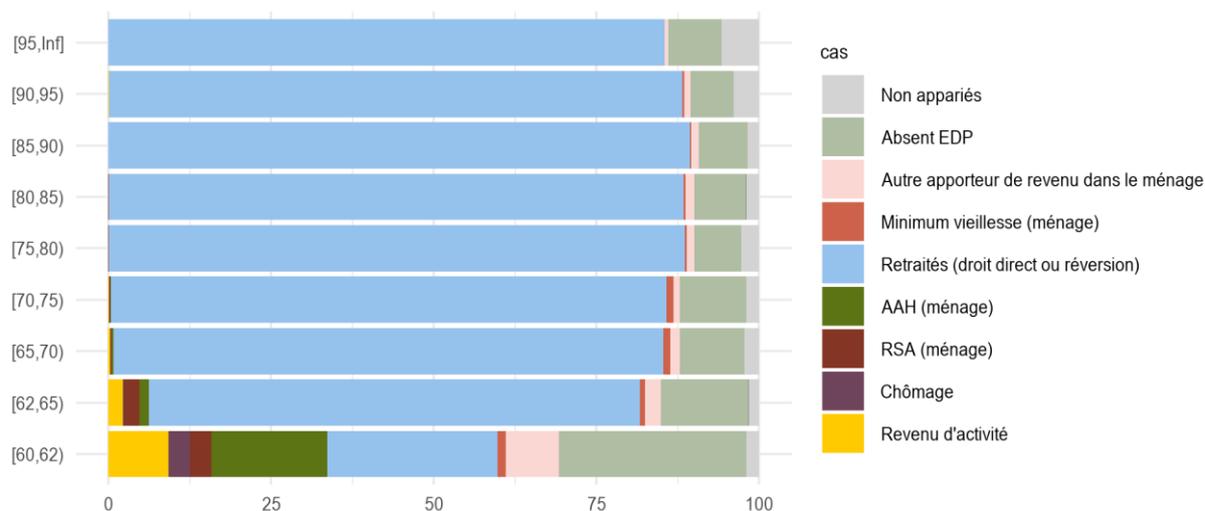
Champ > Personnes de 62 ans ou plus observées dans les RI-APA-ASH, et qui soit sont dans le champ de l'EIR d'après leur date de naissance, soit ont un identifiant EIR retrouvé dans la table de passage, et qui sont en outre dans l'EDP.

Source > Appariement EIR + RI-APA-ASH, DREES.

À titre d'information, les deux graphiques suivants illustrent la situation des bénéficiaires de l'APA ou de l'ASH, d'après les seules informations de l'EDP. Le champ reste toutefois restreint aux personnes qui sont dans le champ des RI-APA-ASH, de l'EDP ET de l'EIR, car on ne dispose pas pour l'instant de la table de passage entre les identifiants des RI-APA-ASH et de l'EDP (qui permettrait de repérer aussi les personnes communes aux RI et à l'EDP mais qui ne sont pas dans l'EIR).

Comme pour l'EIR, la part de non-appariés est faible, mais augmente avec l'âge (figure 7). Cette similarité est quasi-mécanique, puisque ce statut de non-appariement s'explique pour l'essentiel par la qualité du CSNS retrouvé dans les RI-APA-ASH, et puisque les non-appariés de l'EIR sont par construction également non-appariés pour l'EDP.

Figure 7 Situation dans l'EDP des individus observés dans les RI-APA-ASH (en %)



Note > Revenus perçus en 2016. Âges au 31/12/2016. Données non pondérées.

Champ > Personnes de 60 ans ou plus observées dans les RI-APA-ASH, et qui soit sont dans le champ commun de l'EDP et de l'EIR d'après leur date de naissance, soit ont un identifiant EDP retrouvé grâce aux tables de passage, et qui sont en outre dans l'EDP.

Source > Appariement EIR + EDP + RI-APA-ASH, DREES et Insee.

La proportion de personnes n'ayant aucun revenu dans l'EDP apparaît un peu plus élevée que dans l'EIR. Il ne s'agit pas forcément de personnes qui n'ont *réellement* aucun revenu : cette catégorie correspond vraisemblablement davantage aux personnes vivant en ménage non ordinaire (pour lesquelles les données fiscales du ménage, issues de FILOSOFI, ne sont pas remontées dans l'EDP) ou aux personnes pour lesquelles l'appariement de

l'échantillon EDP avec les données fiscales n'a pas fonctionné. Remarquons que, du fait du premier motif, on se serait attendu à observer une forte proportion d'absents de l'EDP parmi les bénéficiaires de l'APA en établissement ou de l'ASH, et une faible proportion d'absents parmi les bénéficiaires de l'APA à domicile. Or, si les écarts entre proportions vont bien dans le sens attendu, celles-ci n'apparaissent pas si élevée que cela dans le premier cas, et pas si négligeables dans le second (figure 8). Cela traduirait donc plutôt un problème d'appariement imparfait entre l'EDP et les données fiscales.

La proportion non négligeable d'individus de l'EDP non-retrouvés dans les données fiscales, y compris parmi celles vivant en ménage ordinaire (bénéficiant de l'APA à domicile), pourrait s'expliquer, au moins pour partie, par le fait qu'une partie d'entre eux sont "personnes à charge" de leur déclaration fiscale, et non déclarant principal ou conjoint. Les données d'identification de ces personnes à charge sont en effet moins complètes dans les fichiers fiscaux, ce qui expliquerait un taux d'appariement moins élevé.

Figure 8 Situation dans l'EDP des individus observés dans les RI-APA-ASH (en %), ventilés selon le lieu de perception des prestations



Note > Revenus perçus en 2016. Âges au 31/12/2016. Données non pondérées.

Champ > Personnes de 60 ans ou plus observées dans les RI-APA-ASH, et qui soit sont dans le champ commun de l'EDP et de l'EIR d'après leur date de naissance, soit ont un identifiant EDP retrouvé grâce aux tables de passage, et qui sont en outre dans l'EDP.

Source > Appariement EIR + EDP + RI-APA-ASH, DREES et Insee.

Taux de retrouvés RI-APA-ASH parmi les individus du champ EIR

On s'intéresse maintenant aux proportions de personnes retrouvées dans les RI-APA-ASH parmi les personnes de 60 ans et plus échantillonnées dans l'EIR et résidant en France. On s'attend bien sûr ici à des taux de personnes retrouvées nettement inférieurs à 100 %, puisqu'une partie importante de la population âgée ne perçoit pas de prestation d'autonomie. C'est donc aux taux de prévalences de l'APA et de l'ASH dans la population, connues d'après les statistiques officielles diffusées par la DREES à partir de son [enquête "Aide sociale" auprès des conseils départementaux](#), qu'on se comparera pour juger de la qualité de l'appariement et de la couverture des données.

Comme on l'a déjà signalé, le repérage des personnes résidant en France parmi celles échantillonnées dans l'EIR n'est pas totalement évident. Si cette information est connue pour les bénéficiaires d'une retraite ou du minimum vieillesse (elle est dans ce cas fournie dans l'EIR par la caisse chargée de verser la prestation) ainsi que pour celles percevant l'APA ou l'ASH d'après les RI-APA-ASH (ces prestations concernant par définition des résidents), ce n'est pas le cas pour celles ne percevant aucune de ces prestations. Dans cette dernière situation, on utilise donc l'information de l'EDP, pour les personnes de l'EIR faisant partie du champ de l'EDP : la personne est repérée comme résidente en France si elle est observée dans les données fiscales de l'EDP, et supposée non-résidente dans le cas contraire. Cette sélection du champ des "résidents en France" nécessite de pondérer les données, pour tenir compte de l'échantillonnage spécifique de l'EIR, qui ne retient pas les mêmes jours de naissance pour toutes les générations : les pondérations utilisées tiennent compte de ce plan d'échantillonnage et des populations au 1^{er} janvier 2017 diffusées par l'Insee (cf. [première sous-partie de la partie "pondérations" ci-après](#) pour le détail du calcul).

Concernant le bénéfice de l'APA ou de l'ASH, on retient ici le fait d'avoir un droit ouvert au 31/12/2016, calculé d'après les dates d'ouverture et, le cas échéant, de fermeture des droits d'après les RI-APA-ASH. Pour certaines observations, ces dates ont été redressées : cf. la [partie "imputations et redressements des variables" ci-après](#). Pour l'EIR, on se restreint aux retraités encore en vie fin 2016, c'est-à-dire les personnes observées dans la vague

2016 de l'EIR, et celles observées dans la vague 2012, qui n'ont pas été rééchantillonnées dans la vague 2016 mais qui sont encore en vie à cette date.

La table analysée ici contient 741 833 observations de personnes résidentes en France et en vie fin 2016, représentant, en tenant compte des pondérations, 16,95 millions de personnes de 60 ans et plus. Elle compte 614 096 retraités de droit direct (représentant 14,56 millions de personnes), 95 425 retraités de droit dérivé (3,55 millions de personnes), 22 489 allocataires du minimum vieillesse (0,57 million de personnes), 12 638 bénéficiaires de l'APA à domicile (0,74 million de personnes), 3 514 bénéficiaires de l'APA en établissement (0,24 million de personnes) et enfin 3 391 bénéficiaires de l'ASH (0,13 million de personnes).

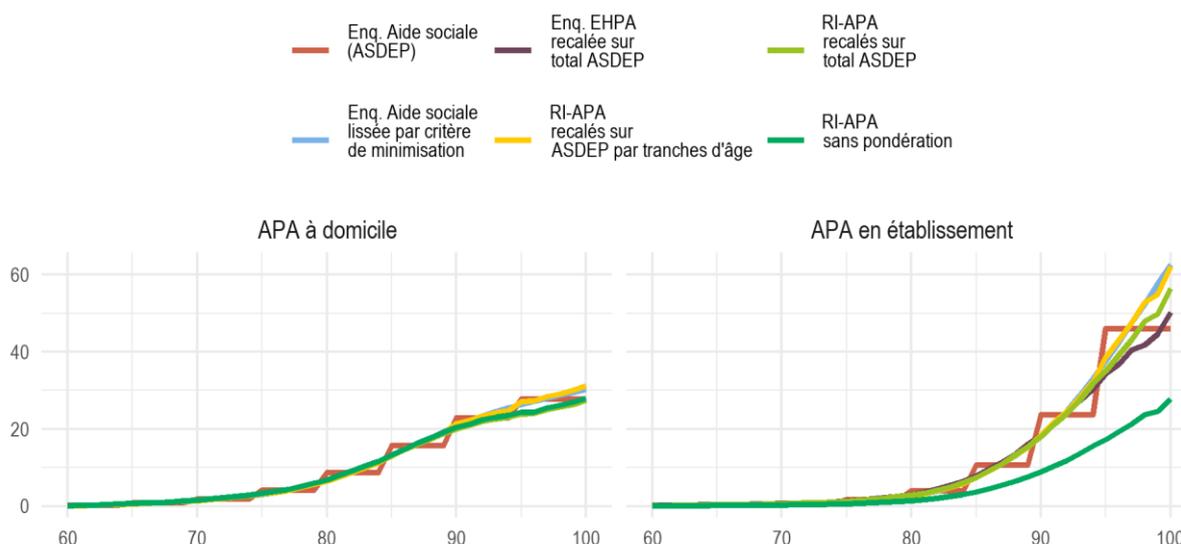
Définition des prévalences de référence

Les prévalences de l'APA dans la population de 60 ans et plus sont collectées et diffusées par la DREES avec une ventilation par tranches d'âge quinquennal. Afin de les comparer aux prévalences dans l'appariement EIR-RI APA, elles sont d'abord, ici, lissées pour disposer de prévalences par âge fin.

Plusieurs méthodes de lissage peuvent être retenues : soit utiliser les données par âge fin dans les RI-APA pour ventiler les effectifs agrégés tirés dans l'enquête *Aide sociale*, soit lisser directement les prévalences par âge quinquennal selon un critère de minimisation (ici : la somme des différences seconde entre les prévalences par âge fin). Les résultats de ces diverses méthodes sont comparés sur les deux graphiques suivants.

Pour le calcul des prévalences de référence, on a utilisé les données sur la ventilation des effectifs de bénéficiaires payés de l'APA par tranches d'âge, diffusées sous [data.drees](#). Ces ventilations ne sont diffusées que pour l'ensemble femmes et hommes confondus, et la comparaison des méthodes de lissage a donc été réalisée à ce niveau³.

Figure 9 Prévalences de l'APA par âge fin, selon diverses méthodes de lissage (en %)



Note > Les âges au-delà de 100 ans ne sont pas représentés, pour des raisons de lisibilité du graphique (valeurs extrêmes liées aux faibles effectifs). Données non-pondérées pour le calcul de la structure par âge des bénéficiaires de l'APA à partir des RI-APA.

Champ > Bénéficiaires payés de l'APA fin 2016.

Source > DREES, enquêtes Aide sociale (2016) et EHPA (2015), RI-APA-ASH.

Pour l'APA à domicile, le taux de couverture de l'ensemble des bénéficiaires par les RI-APA est très élevé, si bien que les prévalences calculées à partir des effectifs bruts tirés de cette source sont très proches de celles calculées à partir des effectifs recalés sur le nombre total de bénéficiaires d'après l'enquête *Aide sociale* (figure 9). En revanche, la structure par âge dans les RI-APA semble conduire à un poids des bénéficiaires de 90 ans et plus moins élevé que celui observé dans l'enquête *Aide sociale* : les prévalences après cet âge apparaissent par conséquent plus basses que celles calculées en recalant les RI-APA sur l'enquête *Aide sociale* par tranche d'âge. Les prévalences avant cet âge sont, symétriquement, plus élevées – même si cela n'est pas visible sur le graphique du fait du faible écart. Par ailleurs, les prévalences par âge fin calculées à partir des RI-APA recalées sur l'enquête Aide

³ Signalons que des prévalences de l'APA par classes d'âge et sexe fin 2016 sont également diffusées, mais elles ont été calculées à partir des estimations provisoires de populations, qui peuvent être fragiles aux âges les plus élevés du fait des faibles effectifs. En pratique, ces prévalences pour les 95 ans et plus, voire pour les 90 ans et plus ne semblent pas cohérentes avec celles qu'on peut recalculer (tous sexes confondus) avec les estimations définitives de population, et pas non plus cohérentes avec les données des RI-APA-ASH, si bien qu'on a préféré ne pas les mobiliser.

sociale par tranche d'âge apparaissent très proches de celles calculées par lissage de cette dernière par un critère de minimisation, et ce à tous âges.

En établissement, les données brutes des RI-APA conduisent à des prévalences nettement plus basses que celles de l'enquête Aide sociale, du fait de la mauvaise couverture par cette source, déjà bien documentée par la DREES, des bénéficiaires de l'APA résidant dans des établissements sous dotation globale. En outre, la structure par âge dans les RI-APA traduit un poids légèrement moindre des bénéficiaires âgés de 90 ans et plus. Les prévalences calculées en recalant les RI-APA sur les effectifs totaux de bénéficiaires dans l'enquête *Aide sociale* sont en effet un peu plus basse après cet âge que celles calculées en recalant les RI-APA sur les effectifs par tranches d'âge, qui sont, elles, très proches des prévalences calculées en lissant les résultats de l'enquête *Aide sociale* selon un critère de minimisation.

À titre de comparaison, on a aussi calculé des prévalences en ventilant les effectifs de bénéficiaires de l'APA en établissement de l'enquête *Aide sociale* selon la répartition par âge fin des résidents en établissements selon l'enquête EHPA (2015) de la DREES – en ne retenant, dans cette dernière, que les résidents éligibles à l'APA en établissement, c'est-à-dire les personnes de 60 ans et plus, classés en GIR 1 à 4 et résidant en EHPAD ou en USLD⁴. Le poids des 90 ans et plus parmi les résidents selon cette enquête est encore plus faible que dans les RI-APA, ce qui conduit à des prévalences encore plus basses après 90 ans.

On sait donc déjà que, si on compare les prévalences de l'APA dans l'appariement EIR-RI APA aux données *par âge* de l'enquête *Aide sociale*, celles-ci apparaîtront sous-estimées aux plus grands âges. Cependant, le but ici n'est pas de porter un diagnostic sur la représentativité des RI-APA en général, mais plutôt sur celle de l'appariement EIR-RI APA par rapport aux RI-APA exhaustive. Dans la suite, on comparera donc les prévalences observées dans l'appariement sur celles calculées en recalant les résultats des RI-APA sur les effectifs totaux de l'enquête Aide sociale, et non sur les effectifs par tranches d'âge dans cette enquête.

Comparaison des prévalences

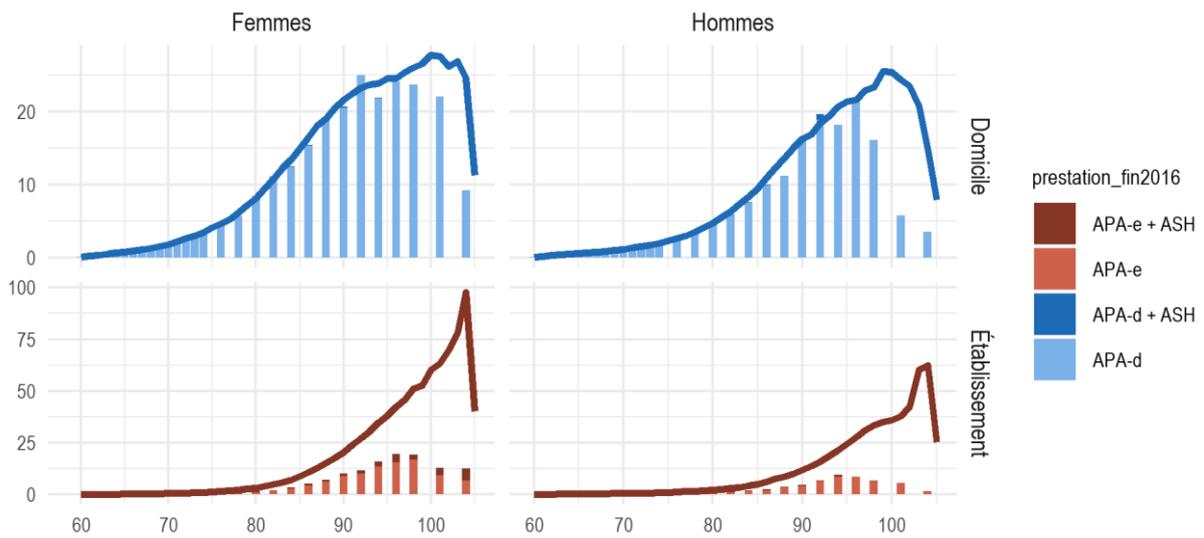
Le graphique suivant compare les proportions de personnes de l'EIR qui sont retrouvées dans les RI-APA-ASH (barres sur le graphique) avec les prévalences de l'APA à domicile et en établissement fin 2016 d'après l'enquête Aide sociale de la DREES, ventilées par âge fin selon la structure par âge dans les RI-APA-ASH (lignes en trait épais). La comparaison n'est pas tout à fait homogène : on considère les bénéficiaires payés au titre de décembre 2016 pour l'enquête *Aide sociale* (information jugée la plus robuste dans cette enquête) mais l'ensemble des bénéficiaires ayant un droit ouvert au 31 décembre 2016 dans la base appariée. On néglige cet écart de notion dans la suite⁵.

Les prévalences de l'APA dans l'EIR, telles que mesurées grâce à l'appariement EIR-RI-APA, apparaissent parfois plus basses que les données de cadrage, surtout aux âges les plus élevés (figure 10). C'était attendu pour ce qui concerne l'APA en établissement, car une partie importante des bénéficiaires, résidant en établissement sous dotation globale, n'est pas observée dans les RI-APA. Pour l'APA domicile, la sous-estimation est sensiblement moins marquée, car la couverture des RI-APA est nettement meilleure. Elle n'apparaît en réalité significative qu'à partir de 98 ans (génération née en 1918 et précédentes). Cette observation est tout à fait cohérente avec les résultats mis en lumière dans la sous-partie précédente, à savoir le fait que taux "non appariés" parmi les personnes observées dans les RI-APA-ASH et dont le jour de naissance est dans le champ de l'EIR est plus faible parmi les générations les plus âgées.

⁴ On n'a, en revanche, pas exclu les résidents en établissements comptant moins de 25 places, éligibles à l'APA à domicile. Ces derniers ne représentent qu'environ 1 % des résidents sur le champ retenu. Les critères appliqués conduisent à sélectionner environ 557 000 résidents dans l'enquête EHPA (effectifs pondérés), soit un peu plus que les 517 000 bénéficiaires de l'APA en établissement fin 2015 dans l'enquête *Aide sociale*.

⁵ D'après l'édition 2020 du panorama *Aide et actions sociales en France*, [93 % des personnes ayant un droit ouvert à l'APA au 31 décembre ont été payées au titre du mois de décembre 2020](#).

Figure 10 Proportion de bénéficiaires de l'APA ou de l'ASH en décembre 2016 (en %)



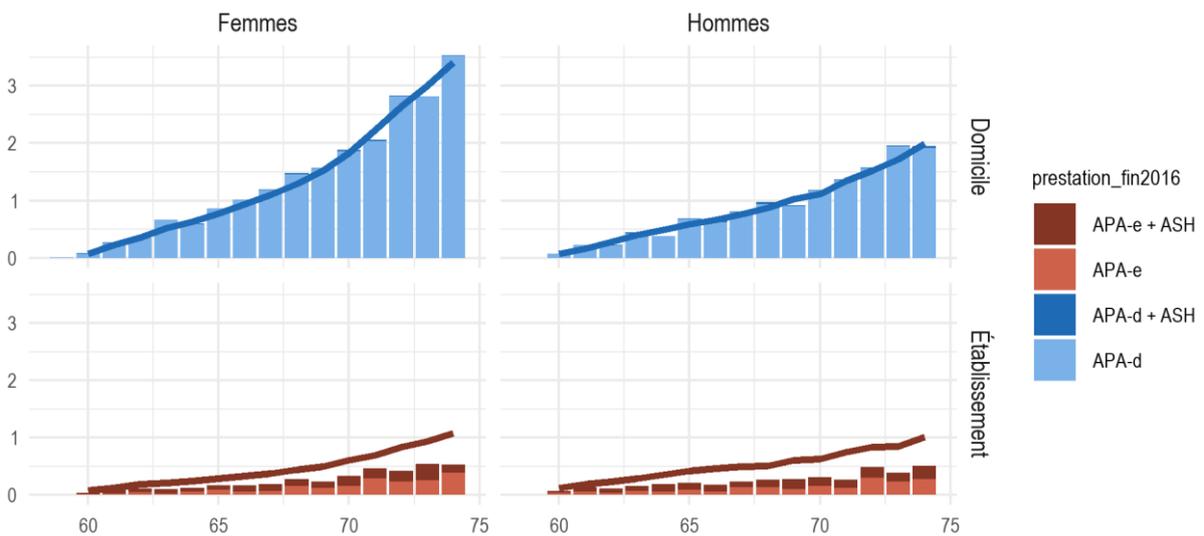
Note > Âge au 31/12/2016. Les données de l'enquête Aide sociale (courbes en trait épais) sont ventilées par âge quinquennal, mais les prévalences ont ici été lissées pour obtenir des courbes par âge fin, en utilisant la structure par âge fin observée dans l'ensemble des RI-APA.

Champ > Personnes échantillonnées dans l'EIR et en vie fin 2016, résidentes en France. Bénéficiaires payés pour l'enquête Aide sociale ; bénéficiaires en droit ouvert pour les RI-APA-ASH.

Source > Appariement EIR + RI-APA-ASH et enquête Aide sociale, DREES.

Dans le graphique suivant, on restreint le champ aux personnes les plus jeunes (moins de 75 ans). Ce *focus* permet de constater que la sous-estimation des prévalences de l'APA en établissement dans l'appariement EIR-RIAPA, par comparaison avec les données de cadrage, concerne aussi ces âges jeunes (figure 11). Pour l'APA à domicile, en revanche, les prévalences avant 75 ans apparaissent très proches. Cette proximité, bien que les bénéficiaires de l'APA soient très rares à ces âges, est permise grâce à la grande taille de l'échantillon de l'EIR.

Figure 11 Proportion de bénéficiaires de l'APA ou de l'ASH en décembre 2016 (personnes de moins de 75 ans) [en %]



Note > Âge au 31/12/2016. Les données de l'enquête Aide sociale (courbes en trait épais) sont ventilées par âge quinquennal, mais les prévalences ont ici été lissées pour obtenir des courbes par âge fin, en utilisant la structure par âge fin observée dans l'ensemble des RI-APA.

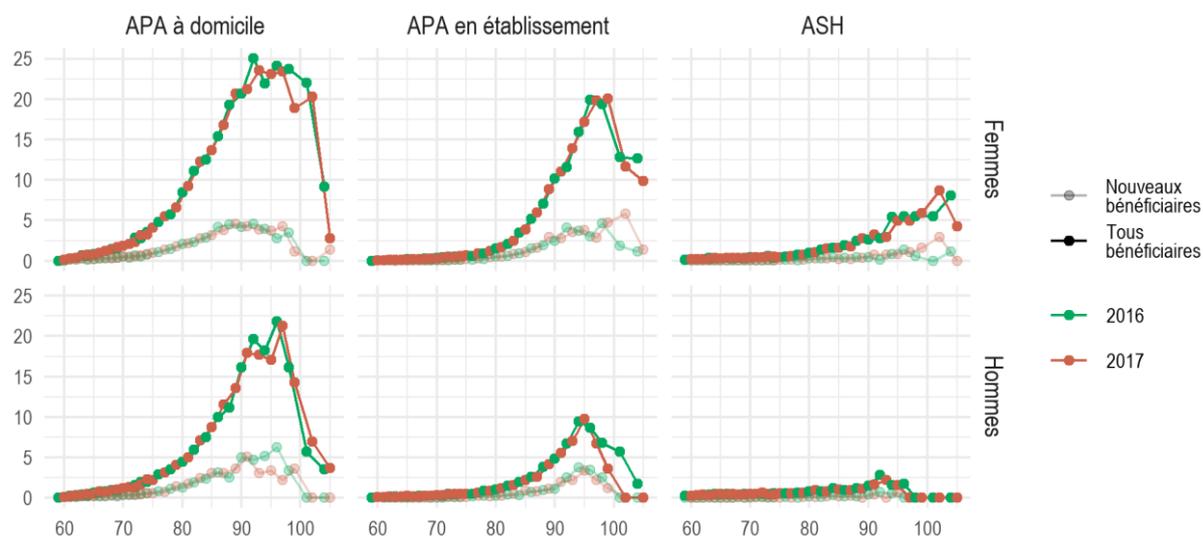
Champ > Personnes échantillonnées dans l'EIR et en vie fin 2016, résidentes en France. Bénéficiaires payés pour l'enquête Aide sociale ; bénéficiaires en droit ouvert pour les RI-APA-ASH.

Source > appariement EIR + RI-APA-ASH et enquête Aide sociale, DREES.

Les RI-APA-ASH incluant toutes les personnes ayant eu un droit ouvert sur l'année 2017, leur appariement à l'EIR permet d'estimer les proportions de bénéficiaires d'aides à la fois fin 2016 et fin 2017 (dans le premier cas, en négligeant les bénéficiaires dont le droit est fermé le 31 décembre 2016 exactement). Ces proportions s'avèrent très proches, hormis aux très grands âges (après 100 ans), où les écarts sont plus marqués du fait du "bruit" lié à l'estimation sur un nombre très restreint de personnes encore en vie.

Il est également possible de calculer les proportions de nouveaux bénéficiaires en cours d'année, définis ici comme les personnes dont le droit a été ouvert pendant l'année *et est encore au 31 décembre*. Là encore, les prévalences apparaissent très proches selon qu'elles sont calculées sur 2016 ou sur 2017, ce qui confirme la possibilité de mobiliser les données sur ces deux années (figure 12). Notons toutefois que la définition utilisée ici ne couvre qu'une partie des nouveaux bénéficiaires, puisqu'elle exclut ceux dont le droit aurait été à la fois ouvert puis fermé en cours d'année. Par construction, cette définition, plus large, ne pourra être calculée que sur l'année 2017, puisque les informations sur les bénéficiaires dont le droit a été fermé avant le 1^{er} janvier 2017 n'ont pas été collectées.

Figure 12 Prévalences des aides à l'autonomie selon la date d'observation (en %)



Note > Âges au 31/12 de l'année. Les bénéficiaires sont définis ici comme les personnes ayant un droit ouvert à la prestation au 31 décembre de l'année ; les nouveaux bénéficiaires sont définis comme les personnes dont le droit à la prestation a été ouvert en cours d'année, et est encore ouvert au 31 décembre.

Champ > Personnes échantillonnées dans l'EIR et en vie aux 31 décembre 2016 et 2017, résidentes en France.

Source > Appariement EIR + RI-APA-ASH, DREES.

Ajout d'informations issues des données fiscales

Dans l'optique d'études et d'analyses sur le financement des aides à l'autonomie des personnes âgées, on complète la base issue de l'appariement EIR + RI-APA-ASH avec quelques informations, relatives à cette thématique, extraites des données fiscales appariées à l'EIR. Ces données portent sur la situation en 2016, et sont donc par construction manquantes pour les générations présentes dans l'appariement EIR-RI APA au titre de leur présence dans l'EIR 2012, mais qui sont absentes de l'EIR 2016 (plus précisément, les générations 1912, 1943 et 1945).

Cet enrichissement se justifie d'une part par la couverture incomplète du champ des bénéficiaires de l'APA en établissement dans les RI-APA-ASH, qui pousse à rechercher d'autres repérages complémentaires de la résidence en établissement pour personnes âgées, et d'autre part par le fait que le financement des aides à l'autonomie en France passe à la fois par des prestations telles que l'APA et l'ASH, mais aussi par des réductions ou crédit d'impôt.

Sélection des informations fiscales

Les données fiscales appariées à l'EIR correspondent aux informations des déclarations n°2042 (principale et complémentaires) pour l'impôt sur le revenu, pour l'ensemble des foyers fiscaux auxquels les individus de l'EIR 2016 résidents en France ont appartenu au cours de l'année. On a d'abord repéré, dans chaque foyer fiscal, à quel individu de la déclaration ("vous" ou "conjoint") correspondait l'individu EIR, en comparant les années de naissance et les montants de pensions dans les deux sources. On a ensuite agrégé, pour chaque individu, toutes les informations des foyers fiscaux auxquels a appartenu l'individu. La plupart des retraités de l'EIR ont une seule déclaration dans l'année, mais il peut y en avoir davantage, par exemple pour une personne qui devient veuve en cours d'année, et qui aura donc une déclaration pour le ménage du vivant du conjoint et une autre par la personne seule après le décès.

On construit ensuite les variables suivantes, à partir des informations déclarées au fisc :

- dépenses d'hébergement en établissement de la personne (variable *depindiv_etab_fisc*) ou du foyer dans son ensemble (*depfoyer_etab_fisc*), à partir des cases 7CD et 7CE de la déclaration n°2042-RICI,

- invalidité à 80 % donnant droit à une demi-part fiscale (variable *inval80_fisc*, définie à partir des cases “P” et “F” de la déclaration 2042, selon que la personne soit “vous” ou “conjoint”) et nombre de personnes invalides à 80 % dans le foyer (*nbfoy_inval80_fisc* : on retient la valeur la plus élevée si la personne a eu plusieurs foyers fiscaux en cours d’année),
- dépenses du foyer pour l’emploi d’un salarié à domicile (variable *depfoy_saldom_fisc*, correspondant à la case 7DF de la déclaration 2042-RICI),
- dépenses d’équipements spécialement conçus pour les personnes âgées ou handicapées (variable *depfoy equip_paph*, correspondant à la case 7WJ de la déclaration 2042-RICI).

On retient également trois variables sur les revenus globaux et les montants d’impôt du foyer (ou des foyers en cas de changement en cours d’année). Cette information est en effet importante, car les réductions d’impôt ne s’appliquent de fait qu’aux foyers payant des impôts : revenu brut global avant charges (*rbg*), revenu imposable (*revimp*), et montant d’impôt sur le revenu (*zimpof*).

On construit également une variable *absent_th*, qui repère les personnes vivant dans un foyer absent du fichier relatif à la taxe d’habitation tout en étant bien présent dans le fichier relatif à l’impôt sur le revenu. Cette absence peut en effet constituer un indice de la résidence en établissement pour personnes âgées, même si cela n’est pas systématique et si la variable doit être utilisée avec prudence.

On construit enfin deux variables : *persfip*, qui désigne la place de la personne dans la déclaration fiscale (“vous” ou “conj”, “vous+conj” si deux déclarations fiscales ont été réalisées dans l’année et que le rôle de la personne a changé – cas notamment lors des veuages survenant en cours d’année –, ou encore “pac” si la personne est considérée comme à charge dans son foyer fiscal), et *statut_matri_fisc*, qui correspond au statut matrimonial pour le fisc (M pour marié, C pour célibataire, etc. – les divers statuts étant accolés en cas de déclarations multiples en cours d’année.)

Rappelons que l’enrichissement de l’EIR de 2016 avec les données fiscales n’est pas exhaustif, car certains individus n’ont pas pu être retrouvés dans les fichiers fiscaux. Cela concerne en pratique 5,2 % des personnes observées dans le volet “retraite” de l’EIR 2016 (et résidentes en France). En comptabilisant en outre les personnes qui ne sont pas observées dans l’EIR 2016, les données fiscales sont manquantes pour 24 % des individus présents dans la base appariée EIR + RI-APA-ASH. Ces proportions de personnes non retrouvées dans l’enrichissement fiscal de l’EIR 2016 s’avèrent plus élevées pour les bénéficiaires de l’APA (7,9 % et 9,4 % respectivement pour l’APA à domicile et en établissement parmi les personnes observées dans l’EIR 2016, soit 17,7 % et 17,7 % parmi l’ensemble des personnes de la base) que pour les non-bénéficiaires (5,1 % parmi les personnes observées dans l’EIR 2016). Une hypothèse pour expliquer cet écart pourrait être que les bénéficiaires de l’APA sont plus souvent personnes à charge au sein de leur déclaration fiscale, le repérage dans les fichiers fiscaux étant plus difficiles pour ces personnes.

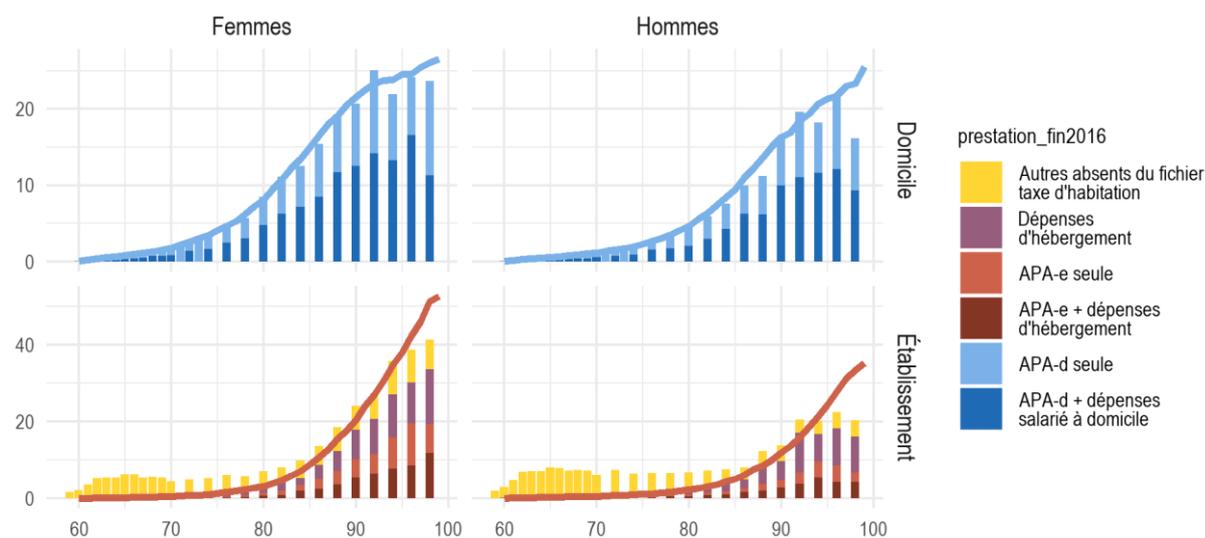
Proportion de résidents en établissement et de personnes déclarant des dépenses d’aide à domicile

Le repérage des personnes résidant en établissement par la combinaison des deux critères (bénéfice de l’APA en établissement d’après les RI-APA et/ou déclaration de dépenses d’hébergement dans les données fiscales) permet d’estimer des prévalences selon l’âge plus proches des prévalences de référence issues de l’enquête *Aide sociale*, même si elles restent malheureusement plus basses. Les personnes restant non repérées correspondent *a priori* aux résidents en établissement sous dotation globale (ce qui explique l’absence dans la base RI-APA-ASH) et n’ayant pas déclaré leurs dépenses d’hébergement au fisc (par exemple parce qu’ils sont déjà non-imposables du fait de faibles revenus, ou parce que leurs dépenses ont été réglées par un tiers). Remarquons qu’on pourrait ici, par extension, inclure également les personnes dont le conjoint a déclaré des dépenses d’hébergement en établissement, en supposant que celui-ci a en réalité payé les dépenses pour les deux membres du couple. Cette inclusion ne rehausserait toutefois que de façon marginale les prévalences estimées.

Si l’on considère également les personnes absentes des fichiers de la taxe d’habitation, les proportions d’individus de l’EIR repérés par l’un au moins des trois repérages de la résidence en établissement se rapprochent nettement, aux grands âges, des prévalences attendues (figure 3). Cependant, elles apparaissent maintenant significativement plus élevées que les prévalences attendues aux âges les plus jeunes, jusqu’à 75 ans voire 80 ans. Cela signifie sans doute que l’absence des fichiers de la taxe d’habitation capte également d’autres types de situations, par exemple le fait de résider dans un autre type d’établissement que ceux destinés aux personnes âgées (résidences sociales, foyers pour adultes handicapés, communautés religieuses, etc.).

Parmi les bénéficiaires de l’APA à domicile, une majorité semble déclarer également des dépenses d’aide à domicile au fisc, mais une minorité loin d’être négligeable n’en déclare pas. Là encore, il ne s’agit pas forcément d’une anomalie, puisque les personnes âgées non imposables n’avaient en réalité aucune utilité de déclarer leurs dépenses d’aide à domicile. Cette remarque vaut pour l’année 2016, observée ici : avec la transformation de la réduction d’impôt en crédit d’impôt en 2017, toutes les personnes âgées dépendantes, imposable ou non, ont intérêt à déclarer leurs dépenses d’emploi à domicile.

Figure 13 Proportion de bénéficiaires de l'APA à domicile ou en établissement fin 2016 (en %)



Note > Âge au 31/12/2016. Les données de prévalences d'après l'enquête Aide sociale (courbes en trait épais) ont ici été lissées pour obtenir des courbes par âge fin, en utilisant la structure par âge fin observée dans l'ensemble des RI-APA.

Champ > Personnes échantillonnées dans l'EIR et en vie fin 2016, résidentes en France, âgées de moins de 100 ans. Bénéficiaires payés pour l'enquête *Aide sociale* ; bénéficiaires en droit ouvert pour les RI-APA-ASH.

Source > Appariement EIR + données fiscales + RI-APA-ASH et enquête Aide sociale, DREES.

■ IMPUTATIONS ET REDRESSEMENTS COMPLÉMENTAIRES DES VARIABLES

Cette partie décrit les corrections appliquées à certaines informations issues des données-sources (EIR, RI-APA-ASH, EDP et données fiscales appariées à l'EIR), qui ont été réalisées spécifiquement dans le cadre de la construction de l'appariement EIR / RI-APA-ASH afin d'en améliorer la qualité. Ces corrections portent sur l'essentiel sur les informations relatives à la résidence en établissement pour personnes âgées et sur le bénéfice de l'APA dans ces établissements.

La présentation reste ici succincte, dans un souci de lisibilité. Le détail précis des redressements effectués est toutefois disponible dans le code source publié sur la [page internet d'ouverture des codes de la DREES](#).

Variables relatives à l'APA

Les données des RI-APA-ASH appariées à l'EIR proviennent soit des bases statistiques redressées par la DREES pour les bénéficiaires d'un droit ouvert fin 2017, soit des bases brutes initiales pour les autres. Pour ces dernières, certains des redressements réalisés par la DREES ont été reproduits :

- Les informations brutes fournies par les départements ont été remplacées par les informations redressées par le prestataire de collecte de la DREES (le cas échéant).
- Les "ressources au sens de l'APA", utilisées pour le calcul du ticket modérateur (variable *RESSOURC_APA*), ont été systématiquement mensualisées et individualisées. Certaines valeurs avaient en effet été fournies sur une base annuelle, ou n'avaient pas été divisées par le nombre d'unités de consommation au sein du ménage.
- Pour certaines valeurs pour lesquelles la modalité "999999999" avait été utilisée pour signaler une valeur inconnue, celle-ci a été remplacée par une valeur manquante (*NA_real*). L'utilisation de la modalité 999999999 se justifiait initialement par le souci de distinguer les valeurs inconnues des valeurs non renseignées car la personne n'était pas concernée. Le choix fait ici de redresser systématiquement à valeur manquante fait qu'il ne sera plus possible de distinguer les deux situations, mais il évite que les calculs sur les montants soient biaisés par des valeurs en réalité non significatives.

Par ailleurs, plusieurs variables de synthèse ont été calculées à partir de la chronique des évaluations du GIR (variables *DATEGIR_EVi*, *GIR_EVi* et *TYPAPA_EVi* pour $i = 1$ à 6 dans les données-sources) :

- Deux variables relatives à la situation au 31 décembre 2016 : type d'APA perçue (*TYPAPA_dec2016*, prenant les valeurs *D* ou *E*) et GIR (*GIR_dec2016*). Elles sont analogues aux variables *TYPAPA* et *GIR_FIN_APA*, relatives à la dernière situation observée en 2017. Les variables fin 2016 ont été calculées en utilisant les valeurs de la dernière évaluation antérieure à cette date. Certaines valeurs sont manquantes, car plusieurs départements n'ont pas fourni l'information sur la chronique complète des évaluations du GIR.
- Dates de première évaluation pour l'APA (ainsi que pour l'APA à domicile et en établissement) et date de dernière évaluation : *DATE_premevalAPAD*, *DATE_premevalAPAE*, *DATE_premevalAPA* et *DATE_derevalAPA*.
- Date de première évaluation en GIR 1 ou 2 (si l'information est disponible) : *DATE_premevalGIR12*.
- Séquences des types d'APA (variable *sequence_TYPAPA_eval*) et des GIR au fil des diverses évaluations (variable *sequence_GIR_eval*) : Par exemple, une valeur "DDD" pour la variable *sequence_TYPAPA_eval* signifie que trois évaluations ont été menées, chaque fois pour l'APA à domicile ; et une valeur "442" pour la variable *sequence_GIR_eval* signifie que les deux premières évaluations ont conduit à un classement en GIR 4 et la troisième en GIR 2.

Enfin, un nombre non négligeable d'incohérences ont été repérées pour les variables de dates (ouverture et fermeture des droits, décès), qui ont nécessité des redressements. Le tableau suivant résume, selon leur nature, le nombre et la fréquence de ces incohérences (pour l'ensemble des RI-APA-ASH, et non pour la seule partie appariée à l'EIR), à la fois avant et après redressement (figure 14). Ces redressements n'ont en effet pu être réalisés que lorsque les informations auxiliaires étaient suffisamment précises pour corriger ou pour imputer les dates repérées comme des anomalies. Dans le cas contraire, les anomalies ont été laissées en l'état, ce qui explique que des incohérences subsistent après redressement.

Figure 14 Nombres et proportions d'anomalies relatives aux dates d'ouverture ou de fermeture des droits à l'APA, avant et après redressements

pb	NbObs_ avant_ redressement	Pct_ avant_ redressement	NbObs_ après_ redressement	Pct_ après_ redressement
Au moins une anomalie de date	272 039	20,54	59 746	4,51
DATEOUVAPA < min(DATE_APAD, DATE_APAE)	192 511	14,53	40 078	3,03
DATE_APAD abs, mais APA-D dans séquence éval	100 296	7,57	1 545	0,12
DATE_APAD et DATEDECES renseignées, mais FERM_APA_D absente	27 908	2,11	314	0,02
DATE_APAD abs, mais TYPAPA==D	16 788	1,27	2 961	0,22
DATE_APAE et DATEDECES renseignées, mais FERM_APA_E absente	14 589	1,10	137	0,01
FERM_APA_D < DATE_APAD	8 561	0,65	600	0,05
FERM_APA_E < DATE_APAE	5 468	0,41	121	0,01
DATE_APAE abs, mais APA-E dans séquence éval	4 545	0,34	2 365	0,18
DATE_APAE abs, mais TYPAPA==E	4 509	0,34	1 976	0,15
DATEOUVASH < FERM_APA_D	3 791	0,29	9 131	0,69
DATE_APAE abs mais FERM_APA_E renseignée	3 414	0,26	1 274	0,10
DATE_APAD < DATE_APAE renseignées, mais FERM_APA_D absente	2 338	0,18	941	0,07
Bénéficiaire APA sans aucune date renseignée	2 263	0,17	2 974	0,22
DATE_APAD > DATEDECES	2 143	0,16	305	0,02
DATE_APAE < DATE_APAD renseignées, mais FERM_APA_E absente	1 695	0,13	427	0,03
DATE_FERM_ASH > DATEDECES	1 171	0,09		
DATE_APAD abs mais FERM_APA_D renseignée	938	0,07	716	0,05
FERM_APA_D = FERM_APA_E	265	0,02	305	0,02
DATE_APAD = DATE_APAE	92	0,01	126	0,01
DATEOUVASH > DATEDECES	54	0,00	54	0,00
DATE_APAE > DATEDECES	39	0,00	40	0,00

Champ > Bénéficiaires de l'APA en 2017.

Source > RI-APA-ASH, DREES.

Les principaux redressements réalisés sur les dates d'ouverture et de fermeture des droits sont résumés dans le tableau suivant (figure 15). Leur nombre et leur fréquence sont indiqués à la fois pour l'ensemble des RI-APA-ASH (colonnes 2 et 3 du tableau) et pour la sous-partie appariée à l'EIR (colonnes 4 et 5). Les principaux changements réalisés sont les suivants (un même individu pouvant être concernés par plusieurs redressements) :

- On a d'abord corrigé les dates manifestement erronées propres à certains départements. Par exemple, un département n'avait aucune valeur manquante pour la date d'ouverture des droits à l'APA à domicile (variable *DATE_APAD*), mais un grand nombre d'observations prenant la valeur "2018-03-08", sachant que toutes les autres variables relatives à l'APA à domicile était non-renseignées. On a considéré dans ce cas que la valeur avait en réalité été utilisée comme indicatrice de "non-concerné", et on a remplacé les valeurs correspondantes par des valeurs manquantes (ligne "*DATE_APAD mise à NA*" dans le tableau).
- Les anomalies les plus fréquentes sont celles où les dates d'ouverture des droits à l'APA à domicile (*DATE_APAD*) ou en établissement (*DATE_APAE*) ne sont pas renseignées, alors que la date d'ouverture des droits à l'APA tous lieux confondus (*DATEOUVAPA*) l'est. On a alors imputé les premières à partir de la dernière. Pour savoir, dans ces cas, si la personne avait bénéficié de l'APA à domicile et/ou en établissement, et dans quel ordre, on a mobilisé la séquence des évaluations du GIR.
- De même, pour un nombre non négligeable de cas, les dates de fermeture des droits ne sont pas renseignées, alors que la personne est décédée. On a alors supposé que les droits ont été fermés à la date de

décès, soit repérée dans les RI-APA-ASH (variable *DATEDECES*), soit dans l'EIR pour les décès survenus en 2017 ("*andeces-moisdeces EIR*").

- On a également imputé la date de fermeture des droits à l'APA à domicile avec la date d'ouverture des droits à l'APA en établissement (moins 1 jour) dans les cas où la chronique des évaluations du GIR faisait apparaître un parcours "domicile puis établissement".
- Lorsqu'aucune date d'ouverture des droits à l'APA n'était renseignée, mais qu'une évaluation pour l'APA à domicile ou pour l'APA en établissement apparaissait dans la chronique des évaluations du GIR, on a, par approximation, retenu cette date d'évaluation (*DATE_premevalAPAD* ou *DATE_premevalAPAE*) comme date d'ouverture des droits. Cela conduit généralement à anticiper de quelques semaines la "vraie" date (inconnue dans les données) d'ouverture des droits.
- On a également corrigé des cas où les dates d'ouverture et de fermeture des droits à l'APA à domicile (variables *DATE_APAD* et *FERM_APA_D*) avaient manifestement été inversées (lignes "*DATE_APAD <> FERM_APA_D*"). Notons qu'il reste malgré tout dans la base des observations où *FERM_APA_D < DATE_APA_D* : c'est notamment le cas en cas de réouverture des droits à l'APA à domicile, pour les personnes qui seraient entrées en établissement après une première période de perte d'autonomie à domicile, puis sorties d'établissement et retournées à domicile.

Figure 15 Nombre d'observations dans la base RI-APA-ASH pour lesquelles les dates d'ouverture ou de fermeture des droits à l'APA ont été redressés

type_redressement_date	NbObs_redressées_RIAPA	Pct_RIAPA	NbObs_redressées_EIR	Pct_EIR
DATE_APAD = DATEOUVAPA / FERM_APA_D = DATE_APAE - 1	82 142	6,2	909	4,3
DATE_APAD = DATEOUVAPA	72 157	5,4	1 104	5,2
FERM_APA_D = DATEDECES	23 700	1,8	365	1,7
FERM_APA_E = DATEDECES	13 337	1,0	162	0,8
DATE_APAD = DATE_premevalAPAD / FERM_APA_D = DATE_APAE - 1	11 692	0,9	139	0,7
FERM_APA_D = andeces-moisdeces EIR			113	0,5
DATE_APAD = DATEOUVAPA / FERM_APA_D = DATEDECES	6 069	0,5	97	0,5
DATE_APAE = DATEOUVAPA	5 957	0,4	74	0,3
FERM_APA_D = DATE_APAE - 1	2 488	0,2	30	0,1
DATE_FERM_ASH = andeces-moisdeces EIR			39	0,2
DATE_APAD mise à NA / DATE_APAE = DATEOUVAPA / FERM_APA_E = DATEDECES	1 929	0,1	34	0,2
FERM_APA_E = andeces-moisdeces EIR			22	0,1
DATE_APAD mise à NA	1 291	0,1	6	0,0
DATE_APAD = DATEOUVAPA / FERM_APA_D = DATE_premevalAPAE - 1 / DATE_APAE = DATE_premevalAPAE / FERM_APA_E = DATEDECES	1 205	0,1	13	0,1
DATE_APAD mise à NA / FERM_APA_E = DATEDECES	1 139	0,1	10	0,0
DATE_APAD = DATEOUVAPA / FERM_APA_D = DATE_APAE - 1 / FERM_APA_E = DATEDECES	998	0,1	9	0,0
DATE_APAD = DATEOUVAPA / FERM_APA_D = andeces-moisdeces EIR			15	0,1
DATE_FERM_ASH = DATEDECES	867	0,1	7	0,0
DATE_APAE = DATE_premevalAPAE	720	0,1	4	0,0

DATE_APAD mise à NA / DATE_APAE = DATEOUVAPA	623	0,0	10	0,0
DATE_APAD = DATEOUVAPA / FERM_APA_D = DATE_premevalAPAE - 1 / DATE_APAE = DATE_premevalAPAE	457	0,0	5	0,0
DATE_APAE = DATEOUVAPA / FERM_APA_E = DATEDECES	313	0,0	7	0,0
DATE_APAD = DATE_premevalAPAD / FERM_APA_D = DATE_APAE - 1 / FERM_APA_E = DATEDECES	321	0,0	4	0,0
DATE_APAE = FERM_APA_D + 1	293	0,0	5	0,0
DATE_APAD mise à NA / FERM_APA_E = DATEDECES / DATE_FERM_ASH = DATEDECES	126	0,0	4	0,0
FERM_APA_D = DATE_premevalAPAE - 1 / DATE_APAE = DATE_premevalAPAE	213	0,0	3	0,0
DATE_APAD = DATEOUVAPA / FERM_APA_E = DATEDECES	121	0,0	2	0,0
DATE_APAD = DATEOUVAPA / FERM_APA_D = DATE_APAE - 1 / DATE_FERM_ASH = DATEDECES	86	0,0	1	0,0
DATE_APAD <> FERM_APA_D	85	0,0	1	0,0
DATE_APAD = DATEOUVAPA / DATE_APAE = DATEOUVAPA	48	0,0	1	0,0
FERM_APA_D = DATEDECES / DATE_APAE = DATEOUVAPA	30	0,0	1	0,0
DATE_APAD = DATEOUVAPA / FERM_APA_D = DATE_APAE - 1 / FERM_APA_E = DATEDECES / DATE_FERM_ASH = DATEDECES	57	0,0		
FERM_APA_D = DATE_APAE - 1 / FERM_APA_E = DATEDECES	47	0,0		
DATE_APAD = DATEOUVAPA / FERM_APA_D = DATE_APAE - 1 / DATE_APAE mise à NA / FERM_APA_E = DATEDECES	24	0,0		
DATE_APAD mise à NA / DATE_APAE mise à NA / FERM_APA_E = DATEDECES	23	0,0		
DATE_APAD = DATE_premevalAPAD	19	0,0		
DATE_APAD = DATEOUVAPA / DATE_FERM_ASH = DATEDECES	15	0,0		

Note > Seuls les redressements concernant au moins 10 observations sont présentés dans le tableau. Les colonnes suffixées “_EIR” portent sur le champ de l'appariement EIR-RIAPA. Les proportions (colonnes “Pct_”) sont exprimées en % du total des bénéficiaires de l'APA observés.

Champ > Bénéficiaires de l'APA en 2017.

Source > RI-APA-ASH et appariement EIR - RI-APA-ASH, DREES.

Variables complémentaires sur les retraites

Afin de faciliter les analyses, on enrichit la table de quelques variables de synthèse sur la retraite, absentes de l'EIR :

- La variable *cc1_reg* correspond au régime de retraite de base principal en cours de carrière, regroupé en grands catégories avec un intitulé en clair (régimes de salariés du privé, de la fonction publique, etc.) Dans la plupart des usages, il sera plus adapté et plus simple de retenir cette variable regroupée, plutôt que le code caisse précis du régime principal *cc1*, exprimé selon la nomenclature spécifique à l'EIR.
- La variable *nb_enfants* est égale au nombre d'enfants de la personne telle qu'elle peut être inférée d'après les informations administratives dont disposent les caisses de retraite. Elle est calculée à partir des nombres de trimestres de majoration de durée d'assurance pour enfants (variable *TDD3A* de l'EIR) lorsque de tels trimestres sont renseignés (c'est-à-dire pour l'essentiel pour les mères ayant un droit direct de retraite) ou, à défaut, à partir des majorations de pension pour trois enfants ou plus (variable *MTOT5* dans l'EIR). Selon les individus, elle peut donc être égale au nombre exact d'enfants, ou bien à une catégorie “Moins de 3 enfants” (catégorie incluant les personnes sans enfant) ou “3 enfants ou plus”.
- Les variables *agedebcotis* et *agefincotis* correspondent aux âges atteints par les assurés aux 31 décembre respectivement de l'année de première cotisation d'un trimestre pour la retraite et de l'année de dernière cotisation d'un trimestre. Ils renseignent donc sur les âges de début et de fin de carrière des personnes ayant un droit direct de retraite. Ils sont construits à partir des variables *PACDD* et *DACDD* de l'EIR, en

corrigeant du fait que certains régimes (de la fonction publique et de la mutualité sociale agricole notamment) n'ont pas fourni cette information : dans ces derniers cas, on a fait l'hypothèse que l'année de dernière cotisation correspond à l'année de départ à la retraite, et que l'année de première cotisation correspond à cette année de départ à la retraite diminuée de la durée validée au titre de l'emploi dans le régime (ce qui revient donc à faire l'hypothèse que l'assuré a validé une période d'emploi continue dans le régime). Du fait de cette dernière approximation, les variables *agedebcotis* et *agefincotis* sont *a priori* plutôt des majorants des âges effectifs de début et de fin de carrière. Les variables ont en outre été tronquées : entre 15 et 26 ans pour l'âge de début de carrière, et entre 55 et 66 ans pour l'âge de fin de carrière. Ces variables sont, par construction, à valeurs manquantes pour les personnes qui ne disposent d'aucun droit direct de retraite.

Résidence en établissement pour personnes âgées

L'appariement des données de l'EIR (enrichi de certaines informations issues de déclarations fiscales) et des RI-APA-ASH fournit plusieurs repérages des personnes âgées résidentes en établissement, mais aucune n'est malheureusement exhaustive : le bénéfice de l'APA en établissement dans les RI-APA-ASH n'est connu que pour une partie du champ, puisque les systèmes d'information des conseils départementaux (dans lesquels sont extraites les RI) n'ont généralement pas d'information individuelle sur les bénéficiaires résidents en établissement sous dotation globale ; les dépenses d'hébergement sont vraisemblablement sous-déclarées au fisc puisque, ouvrant droit à une réduction d'impôt et non à un crédit d'impôt, les personnes non-imposables n'ont aucune incitation à les déclarer ; l'absence des fichiers de la taxe d'habitation ou du champ de FILOSOFI (dans les données EDP) peut être pris comme un indice de résidence en communauté, mais cette dernière n'est pas forcément en établissement pour personnes âgées, etc.

Le repérage de l'hébergement en établissement pour personne âgée étant de première importance dans les analyses de la perte d'autonomie, on cherche ici à imputer au mieux ce statut, en synthétisant les divers *proxies* disponibles. On impute ensuite le bénéfice de l'APA en établissement à certains individus non repérés dans les RI-APA-ASH, afin que la base de données statistique soit la plus représentative possible de l'ensemble des bénéficiaires de l'APA. Cette sous-section du document présente à la fois la méthodologie d'imputation retenue et des éléments de diagnostic sur les biais potentiels.

La résidence dans d'autres types d'établissements, notamment les foyers pour adultes handicapés, est également d'intérêt pour l'analyse des politiques d'autonomie, et l'information relative à ces autres établissements est bien sûr conservée dans la base de données. Néanmoins, du fait du poids nettement plus important des établissements pour personnes âgées parmi les personnes de 60 ans et plus, c'est sur cette seule catégorie d'établissement que cette sous-section et la suivante se centreront.

Redressement du repérage des "communautés" dans l'EDP

Les données issues de FIDELI disponibles dans l'EDP permettent de disposer d'une variable de "communauté la plus importante à l'adresse" déclarée par la personne au fisc (variable *COM_SCAT1*). Comme l'indique le guide utilisateur de FIDELI, cette variable est construite par l'Insee en sélectionnant d'abord les adresses fiscales où les foyers fiscaux inconnus de la taxe d'habitation sont suffisamment nombreux (5 ou plus) et où la proportion de foyers inconnus de la taxe d'habitation est suffisante (30% ou plus). Pour chaque adresse sont conservées les dénominations les plus fréquentes utilisées par les personnes pour décrire textuellement un complément d'adresse à destination des services fiscaux (EHPAD, PRISON, MONASTERE, CCAS, ADOMA,...). Dans ces adresses fiscales les communautés sont repérées grâce à un rapprochement avec le répertoire des communautés au moyen d'algorithmes de reconnaissance des noms et des adresses. Afin de compléter les caractérisations, un certain nombre d'imputations sont effectuées par l'Insee sur la base des mots contenus dans le complément d'adresse indiqué par le foyer aux services fiscaux (notamment et par exemple le terme "Ehpad").

Le repérage des personnes âgées résidant en EHPAD ou en USLD au moyen de cette variable (modalités "11" ou "15" de la variable *COM_SCAT1*⁶) reste malheureusement imparfait. Il ne repère qu'environ la moitié des effectifs attendus d'après les données de cadrage des enquêtes Aide sociale ou EHPA de la DREES. On la complète donc par des imputations complémentaires, en créant une variable redressée *COM_SCAT_red*.

Le redressement est réalisé de la façon suivante :

- Pour chaque individu de l'EDP, on crée d'abord une variable intermédiaire de "suspicion de résidence hors ménage ordinaire" si au moins une des raisons suivantes est observée : soit la personne est repérée comme vivant en communauté d'après la variable *COM_SCAT1*, soit elle a répondu à l'enquête annuelle de recensement (EAR) et a été qualifiée comme vivant hors ménage ordinaire dans cette enquête

⁶ Les intitulés de ces deux modalités dans la nomenclature des communautés de l'Insee sont plus précisément "Maison de retraite, EHPAD" et "Structure pour personnes nécessitant des soins médicaux".

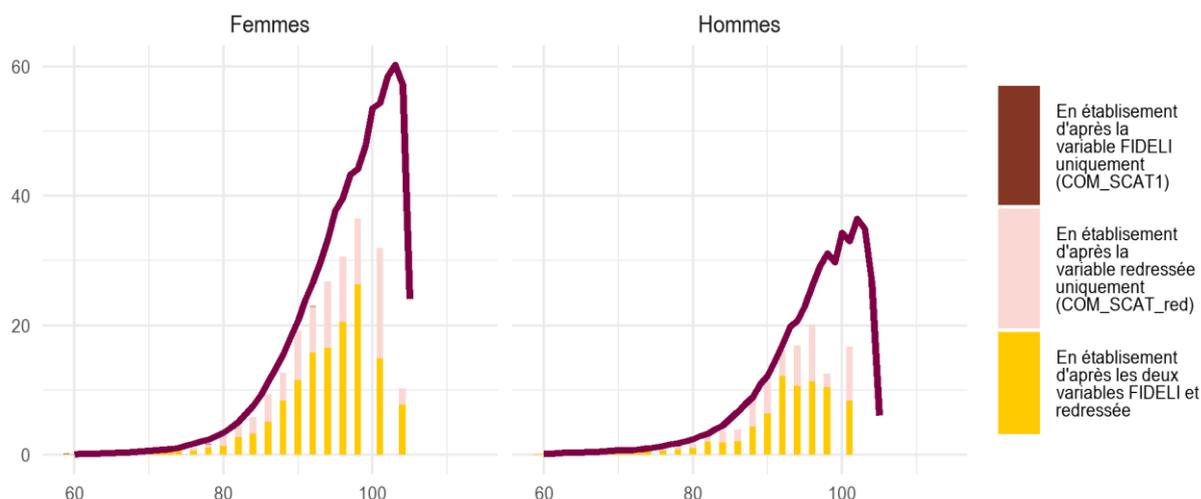
(modalité de la variable *CPOP* différente de "01"⁷), soit les informations sur le logement ou le ménage (au sens des données FILOSOFI) ne sont pas disponibles dans l'EDP (variables *L_FISC_LOGEMENT* ou *L_FISC_REVENU* de l'EDP prenant la modalité "0"), soit la personne est décédée et son lieu de décès indiqué à l'état-civil correspond à l'intitulé "Hospice, maison de retraite" (modalité "HOSMAR" de la variable *CTX_TYPE_LIEU* dans la table "décès" de l'EDP).

- On retient ensuite, pour chaque personne, la dernière période continue pendant laquelle la personne est suspectée de vivre hors ménage ordinaire. Pour les personnes très âgées, cette période est généralement celle qui s'achève par le décès. On exclut de cette étape les individus dont les revenus individuels ne sont jamais observés dans les données fiscales (issues de FIDELI) de l'EDP.
- Si la personne est repérée comme vivant en établissement (par la variable *COM_SCAT1* de FIDELI, par la variable *CTX_TYPE_LIEU* de la table "décès" de l'état-civil, ou par la variable *CPOP* de l'EAR) pour au moins *une* année de cette période, alors on considère qu'elle réside dans le même établissement pour *toutes* les années de la période. Par exemple, si une personne âgée est "suspectée" d'être hors ménage ordinaire de 2016 à 2019 et que la variable *COM_SCAT1* prend la modalité correspondant aux EHPAD (modalité "11") en 2019 seulement, la variable redressée *COM_SCAT_red* prendra la même modalité "11" pour toutes les années de 2016 à 2019. En outre, on fait l'hypothèse qu'une personne observée en EHPAD le reste jusqu'à son décès : toutes les années de vie postérieures à une année d'observation en EHPAD (modalité "11" de la variable *COM_SCAT1* ou résidence en communautés après 75 ans d'après l'enquête annuelle de recensement) sont donc considérées comme des années de résidence en EHPAD.

L'imputation de la variable *COM_SCAT_red* relève de 53 % le nombre de personnes âgées repérées comme en EHPAD ou en USLD (modalités "11" et "15" de la variable), par rapport à la variable *COM_SCAT1* de FIDELI disponible dans l'EDP (figure 16). Les proportions de résidents à chaque âge restent malgré tout inférieures à celles attendues d'après l'enquête EHPA de la DREES.

Notons qu'il est possible en théorie que la variable redressée *COM_SCAT_red* ne corresponde pas à un EHPAD, alors que c'est le cas pour la variable issue de FIDELI *COM_SCAT1*. Cette situation se produit lorsque l'observation dans FIDELI ne fait pas partie de la dernière période continue de "suspicion de résidence hors ménage ordinaire", c'est-à-dire s'il existe au moins une année postérieure pour laquelle les informations sur le logement et le ménage sont bien observés dans FILOSOFI et où la résidence en établissement n'est repérée ni par FIDELI ni par l'EAR. Ces situations s'avèrent néanmoins extrêmement rares.

Figure 16 Proportion de personnes résident en établissement pour personnes âgées fin 2016, avec et sans redressement de la variable *COM_SCAT* de l'EDP (en %)



Note > Les prévalences de référence (courbe sur le graphique, issue des données de l'enquête EHPA) sont calculées fin 2015. On suppose l'écart entre fin 2015 et fin 2016 négligeable.

Champ > Personnes dans le champ de l'EDP et de l'EIR 2016.

Source > Appariement EIR + EDP, DREES et Insee. Enquête EHPA 2015, DREES.

À l'issue de ce redressement, 3 523 observations, représentant environ 426 000 personnes, sont repérées comme résidant en EHPAD fin 2016 (modalité "11" de la nomenclature de communautés de l'Insee) sur le champ des personnes observées dans l'EDP. Par ailleurs, 293 observations, représentant environ 12 000 personnes sont repérées comme résidant en USLD (modalité "15" de la nomenclature de communautés), et 457 observations,

⁷ Pour les personnes qui disent ne pas avoir changé de logement depuis un an dans l'EAR (variable *IRAND* égale à "1"), la variable *CPOP* a été également imputée à l'année précédant l'interrogation dans l'EAR.

représentant environ 11 000 personnes de 60 ans et plus sont repérées comme résidant en établissement pour adultes handicapés (modalité "17" de la nomenclature de communautés). À titre de comparaison, 595 000 personnes résident en EHPAD et 30 000 personnes en USLD fin 2019 [d'après l'enquête EHPA de la DREES](#), et environ 20 000 personnes de 60 ans ou plus résident en maisons d'accueil spécialisées ou en foyers pour adultes handicapés fin 2018 [d'après l'enquête ES-Handicap](#).

Couverture de la résidence en établissement par les divers repérages disponibles

Le tableau suivant synthétise les taux de couverture de la résidence en établissement associés à chacun des repérages envisageables, par rapport aux effectifs attendus d'après l'enquête EHPA de 2015 de la DREES⁸. Les taux de couverture sont calculés pour trois grandes tranches d'âge : moins de 75 ans, 75 à 84 ans, et 85 ans et plus.

Parmi les plus âgés, aucune variable de repérage ne couvre en totalité les nombres de résidents attendus (figure 17). Cette classe d'âge des 85 ans et plus est ici mise en avant comme la plus pertinente pour établir le diagnostic, car c'est celle dans laquelle se situent la plus grande partie des résidents, et celle pour laquelle le risque de confusion avec d'autres types d'établissements médico-sociaux (notamment ceux pour adultes handicapés) est *a priori* le plus faible. La déclaration de dépenses d'hébergement au fisc concerne environ 60 % des résidents attendus, le repérage d'après FIDELI environ 50 % (mais 70 % lorsqu'on utilise la variable redressée décrite dans la sous-section précédente) et le bénéfice de l'APA en établissement ou de l'ASH d'après les RI-APA-ASH environ 50 % également. Notons que cette dernière proportion est plus élevée que la part de bénéficiaires de l'APA en établissement sous dotation globale parmi l'ensemble des bénéficiaires de l'APA en établissement, qui est de 34 % fin 2016 d'après l'enquête *Aide sociale* : une partie des résidents en établissement sous dotation globale est donc bien repérée et couverte dans les RI-APA.

Parmi les personnes de moins de 75 ans, les taux de couverture sont plus élevés, mais ils dépassent alors parfois 100 %. Dans ce cas, cela peut signifier que sont également repérée des personnes qui, en réalité, résident dans d'autres types d'établissements que ceux destinés aux personnes âgées, en particulier – en toute vraisemblance – des foyers pour adultes handicapés. Signalons que, concernant le bénéfice de l'ASH, on a écarté ici toutes les personnes explicitement repérées comme bénéficiaires de l'aide sociale à l'hébergement au titre du handicap (modalité "1" de la variable *HANDI_ASH*) ; on a également écarté les cas où cette situation n'est pas connue dans les données (modalité "9" de *HANDI_ASH*).

Figure 17 Taux de couverture des divers repérages de la résidence en établissement par rapport aux effectifs attendus d'après l'enquête EHPA 2015, selon la tranche d'âge (en %)

Champ	[60,75]	[75,85]	[85,Inf]
Dépenses d'hébergement déclarées au fisc (individu)	74,1	60,2	59,3
Dépenses d'hébergement déclarées au fisc (foyer fiscal)	102,9	76,8	62,7
Bénéfice de l'APA-é ou de l'ASH	68,6	55,5	51,2
Bénéfice de l'APA-é	47,9	42,8	45,8
Bénéfice de l'ASH	69,6	94,0	88,4
Bénéfice de l'APA-é ou l'ASH ou dépenses d'hébergement (fisc)	121,7	92,5	86,1
Résidence en EHPAD ou USLD (FIDELI-EDP)	64,9	43,5	50,4
Résidence en EHPAD ou USLD (EDP redressé)	72,8	76,4	80,5
Résidence EHPAD ou USLD (FIDELI) ou dépenses d'hébergement (fisc)	116,8	79,3	79,8
Résidence EHPAD ou USLD (EDP redressé) ou dépenses d'hébergement (fisc)	121,3	98,0	94,5
Absent dans FIDELI (EDP)	781,4	176,5	48,2
Absent dans FILOSOFI (EDP) (présent FIDELI)	437,1	96,7	71,5
Absence fichier Taxe d'habitation (EIR-données fiscales)	1 110,2	174,3	83,3

⁸ L'enquête EHPA permet de calculer des prévalences de la résidence en établissement par sexe et âge fin 2015, et ces prévalences sont appliquées à la population fin 2016 pour estimer les effectifs attendus. La résidence en établissement a été ici définie par le fait de percevoir l'APA en établissement ou l'ASH, sauf lorsqu'on considère chacune de ces deux prestations séparément dans la table appariée, auquel cas on calcule le taux de couverture par rapport aux seuls bénéficiaires de l'APA ou de l'ASH respectivement.

Lecture > Les personnes de 85 ans ou plus déclarant des dépenses d'hébergement au fisc représente 59,2 % des personnes résidentes en établissement de cette tranche d'âge.

Champ > Personnes dans le champ de l'EDP et de l'EIR 2016.

Source > Appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee. Enquête EHPA 2015, DREES.

Comparaison des divers repérages de la résidence en établissement

On cherche maintenant à illustrer la cohérence (ou non) entre les divers repérages disponibles de l'hébergement en établissement pour personnes âgées. Concernant le fait d'être absent des fichiers de la taxe d'habitation (dans l'enrichissement de l'EIR avec les données fiscales) ou de FILOSOFI (dans l'EDP), on ne définit la variable que sur les sous-champs des personnes qui, respectivement, sont bien repérées dans les fichiers de l'impôt sur le revenu ou ont des valeurs renseignées pour les variables issues de FIDELI, afin de ne pas tenir compte des cas "d'absences" qui s'expliqueraient simplement par un défaut d'appariement. À noter par ailleurs que les fichiers de la taxe d'habitation disponibles avec l'enrichissement fiscal de l'EIR 2016 semblent relatifs à la taxe payée en 2016 : certaines personnes entrées en établissement au cours de l'année 2016 pourraient donc être encore présentes dans ces fichiers.

La cohérence entre les divers repérages apparaît aussi assez limitée : parmi les bénéficiaires de l'APA en établissement ou de l'ASH, par exemple, seuls un tiers environ déclarent des dépenses d'hébergement au fisc, et de l'ordre de 60 % sont absents des fichiers de la taxe d'habitation ou de FILOSOFI (figure 18). Réciproquement, parmi les personnes qui déclarent des dépenses d'hébergement au fisc, un peu plus d'un tiers seulement bénéficie d'une prestation en établissement d'après les RI-APA-ASH, et la moitié sont absents des fichiers de la taxe d'habitation ou de FILOSOFI. Soulignons également que l'absence dans ces deux derniers fichiers apparaît assez faiblement corrélée, alors qu'on s'attendrait à une similarité plus forte : seulement la moitié environ des personnes absentes de FILOSOFI (mais présentes dans FIDELI) sont également absentes des fichiers de la taxe d'habitation, et seulement un quart des personnes absentes de la taxe d'habitation sont également absentes de FILOSOFI (hors celles qui sont également absentes de FIDELI, cas qui a été écarté car il pourrait correspondre à un problème d'appariement dans l'EDP).

Figure 18 Cohérence entre les divers repérages de la résidence en établissement (en %)

champ	Nombre d'observations	Dépenses d'hébergement déclarées au fisc (individu)	Bénéfice de l'APA-é	Bénéfice de l'ASH	Résidence en EHPAD ou USLD (FIDELI-EDP)	Résidence en EHPAD ou USLD (EDP redressé)	Absent dans FILOSOFI (EDP) (présent FIDELI)	Absence fichier Taxe d'habitation (EIR-données fiscales)
Dépenses d'hébergement déclarées au fisc (individu)	3 212	100,0	27,7	10,4	37,9	52,3	50,7	49,8
Dépenses d'hébergement déclarées au fisc (foyer fiscal)	4 052	79,3	23,1	8,2	30,2	42,2	40,8	40,1
Bénéfice de l'APA-é ou de l'ASH	2 916	37,3	69,4	48,8	40,8	57,2	62,1	57,9
Bénéfice de l'APA-é	2 025	44,0	100,0	25,3	41,7	60,7	61,4	57,0
Bénéfice de l'ASH	1 385	23,8	35,7	100,0	41,4	53,6	67,3	62,7
Bénéfice de l'APA-é ou l'ASH ou dépenses d'hébergement (fisc)	5 041	63,7	40,2	27,9	36,5	51,0	52,9	50,5
Résidence en EHPAD ou USLD (FIDELI-EDP)	2 739	44,4	30,9	21,2	100,0	96,5	85,3	69,4
Résidence en EHPAD ou USLD (EDP redressé)	3 724	45,1	33,0	20,2	71,0	100,0	79,6	66,5
Résidence EHPAD ou USLD (FIDELI) ou dépenses d'hébergement (fisc)	4 734	67,8	27,2	15,0	57,9	65,7	60,1	53,1
Résidence EHPAD ou USLD (EDP redressé) ou dépenses d'hébergement (fisc)	5 256	61,1	28,2	16,1	50,4	70,9	60,9	53,3

Absent dans FIDELI (EDP)	106 265	0,1	0,2	0,1	0,0	0,0	0,9	0,9
Absent dans FILO-SOFI (EDP) (présent FIDELI)	14 912	10,9	8,3	6,3	15,7	19,9	100,0	43,4
Absence fichier Taxe d'habitation (EIR-données fiscales)	28 434	5,6	4,1	3,1	6,7	8,7	22,7	100,0

Lecture > Parmi les personnes déclarant au fisc des dépenses d'hébergement en établissement (soit 3 209 individus dans la base), 27,7 % sont bénéficiaires de l'APA en établissement, 37,9 % résident dans un établissement social ou médico-social d'après les données de FIDELI disponibles dans l'EDP, et 49,8 % sont absents des fichiers de la taxe d'habitation dans les données de l'appariement de l'EIR avec les fichiers fiscaux.

Champ > Personnes dans le champ de l'EDP et de l'EIR 2016. Données non pondérées.

Source > Appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee.

La résidence en établissement d'après les données FIDELI présentes dans l'EDP (variable *COM_SCAT1*) apparaît comme l'une des variables les plus corrélées aux autres : environ 30 % sont repérées comme bénéficiaires de l'APA ou de l'ASH, environ 40 % déclarent des dépenses d'hébergement au fisc, et entre 69 % et 83 % sont absents des fichiers de la taxe d'habitation ou de FILO-SOFI. Le tableau ci-après détaille les résultats selon le type de communautés de résidence, d'après la nomenclature de la variable *COM_SCAT1* dans l'EDP. L'illustration est ici réalisée à partir de la variable redressée (*COM_SCAT_red*). Parmi les personnes repérées comme résident en EHPAD (modalité 11 de la nomenclature des communautés de l'Insee), un tiers environ sont bénéficiaires de l'APA en établissement dans les RI-APA-ASH, et 40 % environ déclarent des dépenses d'hébergement au fisc (figure 19). Ces proportions sont respectivement de une sur six et une sur cinq dans les structures "pour personnes nécessitant des soins médicaux" (modalité 15). Pour celles résident en établissement pour adulte handicapé, une petite minorité seulement (5 %) déclare des dépenses d'hébergement ou est bénéficiaire de l'APA, mais un tiers est bénéficiaire de l'ASH. Enfin, parmi les résidents dans les autres types de structures (foyers sociaux, communautés religieuses, etc.), très peu perçoivent une prestation d'autonomie ou déclarent des dépenses d'hébergement. Les proportions d'individus absents des fichiers de la taxe d'habitation ou de FILO-SOFI parmi ces résidents sont plus faibles que pour les résidents en structures médicales ou médico-sociales, tout en restant non négligeables.

Figure 19 Situation au regard des divers repérages de la résidence en établissement, selon le lieu de résidence d'après FIDELI (en %)

type_com	Nombre d'observations	Dépenses d'hébergement déclarées au fisc (individu)	Bénéfice de l'APA-é	Bénéfice de l'ASH	Bénéfice de l'APA-é ou l'ASH ou dépenses d'hébergement (fisc)	Absent dans FILO-SOFI (EDP) (présent FIDELI)	Absence fichier Taxe d'habitation (EIR-données fiscales)
Maison de retraite, EHPAD (COM_SCAT_red == 11)	3 435	46,9	34,4	20,5	71,6	79,3	67,1
Structure pour personnes nécessitant des soins médicaux (COM_SCAT_red == 15)	289	23,5	16,3	16,5	39,7	83,4	59,2
Structure pour les adultes handicapés (COM_SCAT_red == 17)	453	5,5	3,8	9,3	17,0	79,2	55,0
Autres structures	984	1,3	1,3	1,8	3,0	58,3	35,8
Non renseigné	582 505	0,3	0,1	0,1	0,4	1,9	4,4

Lecture > Parmi les personnes qui résident en maison de retraite ou EHPAD d'après les données FIDELI disponibles dans l'EDP et/ou redressées (soit 3 850 individus dans la base), 41,3 % déclarent au fisc des dépenses d'hébergement en établissement, 32 % sont bénéficiaires de l'APA en établissement, et 69,7 % sont absents des fichiers de la taxe d'habitation dans les données de l'appariement de l'EIR avec les fichiers fiscaux.

Champ > Personnes dans le champ de l'EDP et de l'EIR 2016, présentes parmi les informations de FIDELI disponibles dans l'EDP. Données non pondérées.

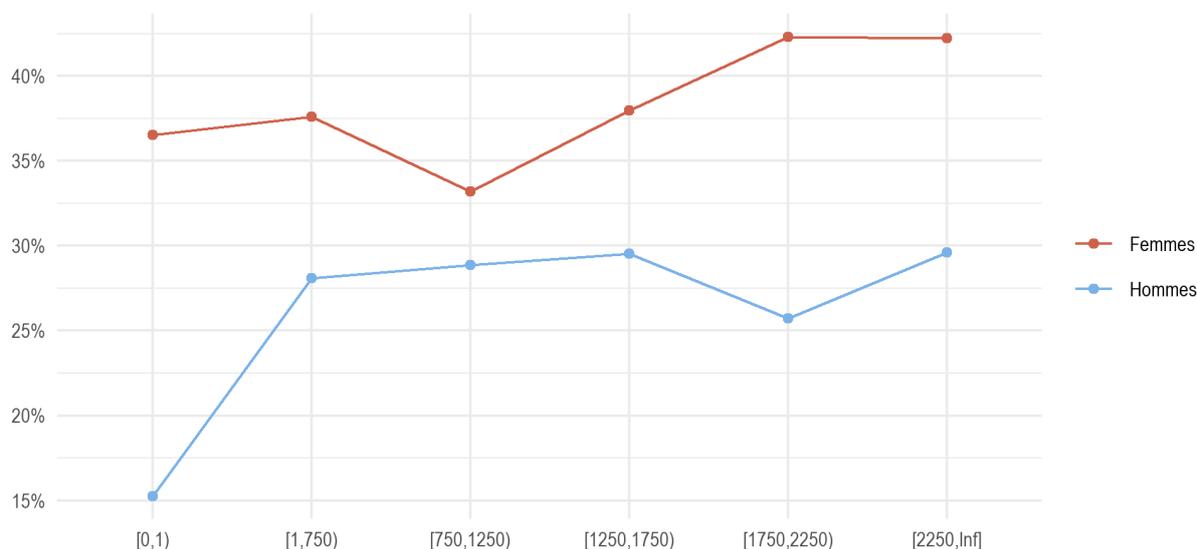
Source > Appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee.

Le graphique suivant cherche enfin à illustrer les éventuelles disparités de taux de couverture de l'APA en établissement dans les RI-APA selon les caractéristiques des individus. On retient ici un critère de ventilation de la population selon le revenu. On sait en effet que la couverture des RI-APA est moins élevée dans les établissements sous dotation globale ; or le fait d'être sous dotation globale est nettement moins fréquent parmi les établissements privés, et la probabilité de résider dans un établissement public ou privé varie fortement selon le revenu des personnes, du fait des tarifs généralement plus élevés des premiers. L'analyse consiste ici à représenter le taux de couverture de l'APA en établissement dans les RI-APA-ASH, à sexe et tranche de revenus donnés, parmi les personnes repérées comme résidentes en établissement. Pour se placer sur le champ le plus large possible, on retient toutes les personnes déclarant des dépenses d'hébergement en 2016 dans les données fiscales ou repérées comme en maison de retraite ou USLD d'après l'information redressée de l'EDP (modalités "11" et "15" de la

variable *COM_SCAT_red*). Par simplification, le revenu considéré ici correspond à la pension moyenne des deux membres du couples (ou la pension de l'individu s'il vit seul) mensualisée en 2015.

Le taux de couverture en établissement par les RI-APA-ASH apparaît effectivement croissant avec le montant moyen de retraite dans le foyer (figure 20). Pour les femmes, l'écart est de près de 10 points (un peu plus de 40 % vs. un peu plus de 30 %). À tranche de retraite donnée, le taux de couverture apparaît aussi significativement plus bas pour les hommes que pour les femmes. L'explication semble ici moins évidente.

Figure 20 Proportion de bénéficiaires de l'APA en établissement repérés dans les RI-APA-ASH parmi les résidents en établissements, selon le montant mensuel moyen de retraite du foyer



Champ > Personnes dans le champ de l'EDP et de l'EIR 2016, présentes parmi les informations de FIDELI disponibles dans l'EDP, déclarant des dépenses d'hébergement au fisc ou repérées comme vivant en établissement d'après FIDELI (variable redressée *COM_SCAT_red*).

Source > Appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee.

Imputation de l'APA en établissement

Le défaut de couverture de l'APA en établissement significatif dans les RI-APA-ASH pose problème en termes de représentativité de la population des bénéficiaires de cette prestation. On complète donc les données en imputant d'autres bénéficiaires dans la base de données, de façon à ce que les prévalences de l'APA en établissement dans cette base soient, après imputation, plus proches de celles attendues dans la population générale d'après les données de référence. Cette imputation est réalisée avec le souci de conserver, au maximum, la cohérence avec les autres repérages de la résidence en établissement disponibles dans les données.

L'imputation est réalisée par étapes, de la façon suivante :

Étape 1 : On crée d'abord une indicatrice de repérage de la résidence en établissement pour personnes âgées en 2017 (y compris pour celles qui décèdent en cours d'année). Cette indicatrice vaut 1 dès lors qu'au moins un type de repérage est disponible : bénéfice de l'APA en établissement ou de l'ASH dans les RI-APA-ASH (à l'exception de l'ASH versée au titre de l'aide aux personnes handicapées et non de celle aux personnes âgées⁹), déclaration de dépenses d'hébergement au fisc (sauf si la personne a par ailleurs des droits encore ouverts à l'APA à domicile – ce qui pourrait signaler une résidence dans un établissement de moins de 25 places), résidence en établissement d'après l'EDP (d'après la variable de FIDELI ou bien d'après la variable redressée selon la méthode présentée dans la sous-section précédente), sortie de l'APA à domicile en cours d'année 2017 au moins 1 mois avant le décès.

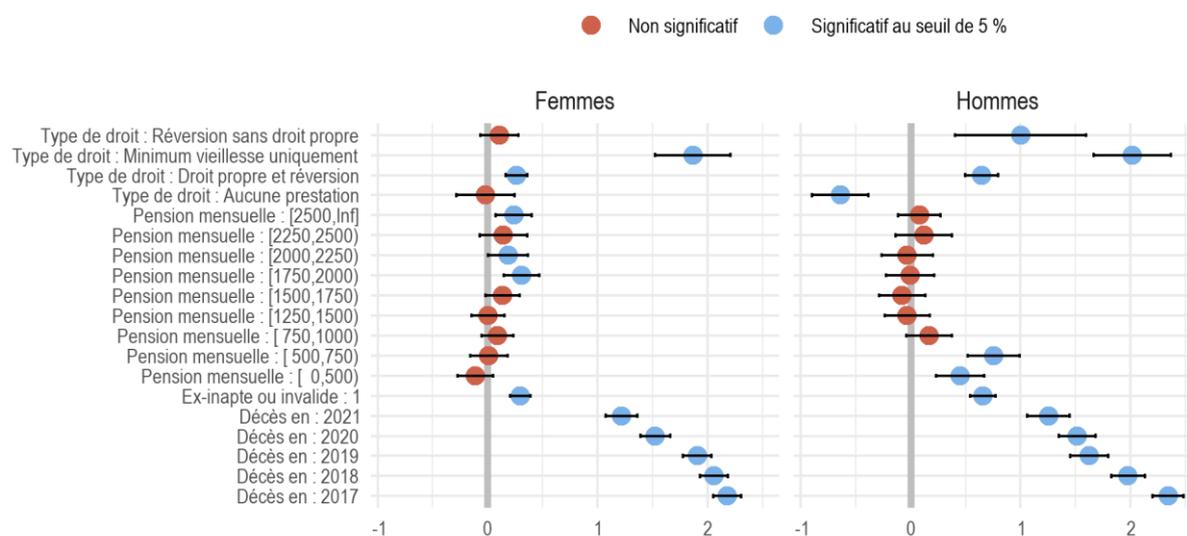
Étape 2 : Pour chaque individu de la table, vivant au 31 décembre 2016, on impute une probabilité de résider en établissement en 2017 à partir de la régression par un modèle *logit* de la variable créée à l'étape 1 sur les variables explicatives suivantes : polynôme de degré 4 de l'âge, montant de pension (droit propre et réversion) par tranche, type de droit dans l'EIR (droit propre seul, réversion, etc.), indicatrice de départ à la retraite au titre de l'inaptitude

⁹ On a, en pratique, retenu uniquement les cas où la variable *HANDI_ASH* prend la modalité "0". Les observations pour lesquelles *HANDI_ASH* est non renseignée (modalité "9") ont ainsi, implicitement, été supposée relever de l'aide aux personnes handicapées ; ce choix s'appuie sur le fait que la majorité des bénéficiaires pour lesquels *HANDI_ASH* == "9" ont une date d'ouverture des droits antérieure à leurs 60 ans, dans des proportions très proches des bénéficiaires pour lesquels *HANDI_ASH* == "1", alors que ce n'est le cas que d'une minorité pour les bénéficiaires pour lesquels *HANDI_ASH* == "0".

ou de l'ex-invalidité, année de décès entre 2017 et 2021 (pour les personnes décédées). Le modèle est estimé séparément sur chaque sexe. Les coefficients associés à chacune de ces variables sont représentés dans le graphique suivant.

D'après les probabilités imputées compte tenu des caractéristiques individuelles des personnes observées dans la table, 556 000 personnes environ encore en vie fin 2017 résideraient en établissement, soit un effectif proche des 541 000 bénéficiaires de l'APA en établissement à cette date dans l'enquête *Aide sociale* de la DREES.

Figure 21 Coefficients associés aux variables explicatives du modèle *logit* de probabilité de résider en établissement en 2017



Note > Un coefficient supérieur à 0 signifie que la modalité est associée à une probabilité de résider en établissement plus élevée que pour la modalité de référence. Cette dernière correspond aux modalités non mentionnées sur le graphique : bénéficie d'un droit propre uniquement pour le type de droit, montant compris entre 1 000 et 1 250 euros par mois pour la pension, retraités non ancien inapte ou invalide, et encore en vie début 2022. Pour plus de lisibilité, la constante du modèle et les coefficients associés à l'âge (polynôme de degré 4) ne sont pas représentés.

Lecture > Le fait de décéder en 2017 est associé à un coefficient d'un peu plus de 2 dans la régression logistique pour les femmes, ce qui signifie qu'une femme décédant en 2017 a, toutes choses égales par ailleurs, une probabilité significativement plus élevée d'être résidente en établissement cette année-là qu'une femme encore en vie début 2022.

Champ > Personnes dans le champ de l'EDP et de l'EIR, en vie fin 2016.

Source > Appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee.

Étape 3 : On impute ensuite aléatoirement le fait d'être ou non en établissement en 2017 (éventuellement sur une partie de l'année seulement pour les personnes qui décèdent en cours d'année), compte tenu de la probabilité calculée à l'étape précédente.

Pour cela, on met en œuvre une méthode d'alignement, telle que décrite dans le [Document de travail M 2014/01 de l'Insee](#) (pages 40 à 42). Le principe de la méthode peut être brièvement décrit de la façon suivante : on tire d'abord pour chaque individu un aléa u uniforme entre 0 et 1 ; on calcule ensuite la différence $v = \text{logit}(u) - \text{logit}(p)$ où p représente la probabilité d'être en établissement en 2017, modélisée lors de l'étape précédente ; on ordonne enfin les individus selon les valeurs croissantes de v et on retient le nombre de personnes qui permet d'obtenir les effectifs "attendus" compte tenu des données de cadrage sur la prévalence de l'APA en établissement (données de l'enquête *Aide sociale* de la DREES avec la ventilation par âge issue des RI-APA pour l'APA en établissement, corrigées pour tenir compte de la différence entre la probabilité d'être bénéficiaire de l'APA en établissement en cours d'année vs. au 31 décembre de l'année).

La procédure a été adaptée de façon à retenir en premier les personnes dont on a déjà repéré qu'elles étaient en établissement en 2017 d'après l'étape 1), et pour placer à l'inverse en dernier les personnes dont on sait qu'elles bénéficient (encore) de l'APA à domicile fin 2017, ou qu'elles en ont bénéficié jusqu'à leur décès courant 2017 (afin de ne pas les retenir comme personnes âgées en établissement). Cette adaptation se justifie par le souci de conserver la cohérence interne entre l'imputation de l'APA en établissement et les autres variables relatives à la résidence en établissement. En outre, le tri et la sélection sont réalisés séparément sur chaque sous-groupe définis par le croisement du sexe, de l'âge (en regroupant tous les plus 100 ans ou plus), mais aussi d'indicatrices de présence dans le champ de l'EDP et dans celui de l'appariement de l'EIR avec les données fiscales, et d'une indicatrice de décès en 2017¹⁰. En effet, dans la mesure où les données de l'EDP et de l'appariement fiscal de

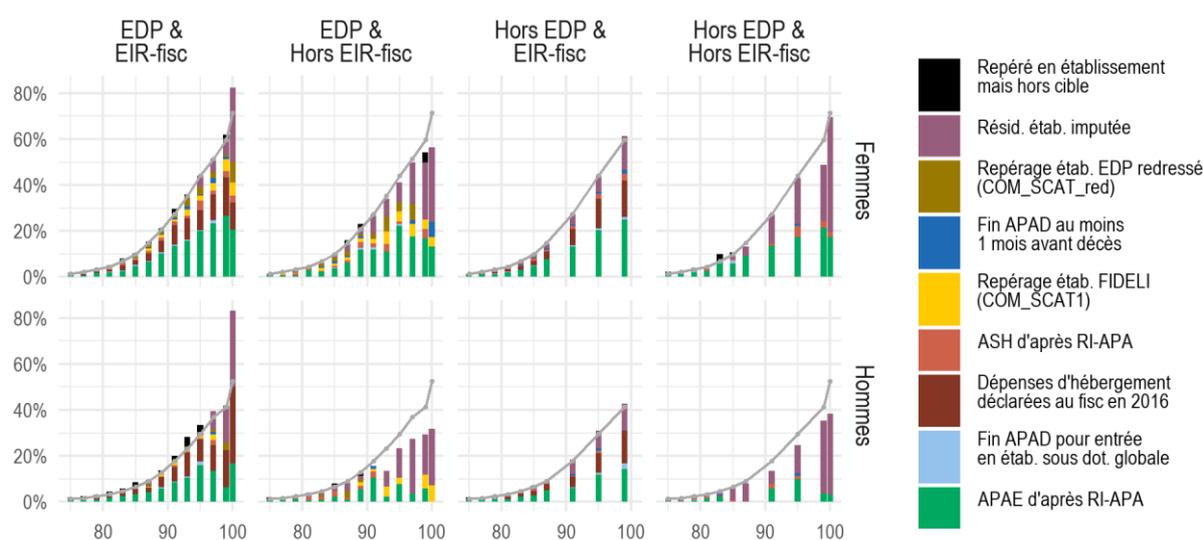
¹⁰ Pour les âges les plus élevés, les divers sous-champ EDP et données fiscales ont toutefois été regroupés, pour éviter de créer de cellules de taille trop petite, dans lesquelles l'imputation se heurterait au nombre trop faible d'observations.

l'EIR sont utilisées pour repérer la résidence en établissement, la proportion de résidents "manquant" varie significativement selon qu'un individu est ou non dans le champ de ces données.

Les effectifs-cibles attendus sont calculés en multipliant les effectifs pondérés totaux pour chaque sous-champ, sexe et âge par les prévalences de l'APA en établissement par sexe et âge. On a ainsi considéré le fait d'être dans le champ des sources EDP et/ou EIR-données fiscales comme aléatoire, et on a donc supposé que les prévalences par sexe et âge étaient identiques dans chaque sous-champ. Les prévalences de l'APA sont elles-mêmes calculées à partir des prévalences au 31 décembre 2017 issues de l'enquête *Aide sociale* et de ratios calculés à partir des RI-APA-ASH exhaustives : par exemple, la prévalence des bénéficiaires de l'APA au moins une fois en 2017 parmi les vivants en début d'année est égale à la prévalence de l'APA au 31 décembre 2017 parmi les vivants à cette date, divisée par la probabilité de survie en cours d'année (d'après les données de mortalité de l'Insee), et multipliée par le rapport entre les nombres de bénéficiaires de l'APA en établissement au moins une fois en 2017 et de bénéficiaires au 31 décembre 2017 observés dans les RI-APA-ASH.

Le graphique suivant résume la procédure d'imputation en présentant les prévalences de l'APA en établissement ventilées selon la "source" du repérage (observation dans les RI-APA, repérage grâce à une autre observation, ou imputation aléatoire).

Figure 22 Prévalences de l'APA en établissement en 2017 : valeurs attendues vs. personnes repérées comme résidentes en établissement



Note > Les prévalences en 2017 ne sont pas comparables aux graphiques précédents, car elles correspondent ici aux bénéficiaires de l'APA en établissement au moins une fois en cours d'année, y compris ceux qui sont décédés au 31 décembre 2017. Chaque personne est associée à une catégorie unique, attribuée de façon séquentielle en suivant l'ordre inverse à celui d'apparition dans la légende.

Lecture > Parmi les femmes de 100 ans fin 2016 qui sont observées dans les données de l'EDP et dans les données fiscales appariées à l'EIR (champ "EDP & EIR-fisc"), 20 % sont observées comme bénéficiaire de l'APA en établissement (APAE) d'après les données RI-APA, 12 % supplémentaires ont des dépenses d'hébergement déclarées au fisc en 2016, etc.

Champ > Personnes en vie fin 2016, de 75 ans ou plus fin 2017. Les âges sont calculés au 31 décembre 2017. Pour des raisons de lisibilité du graphique, seuls les âges à partir de 75 ans sont représentés.

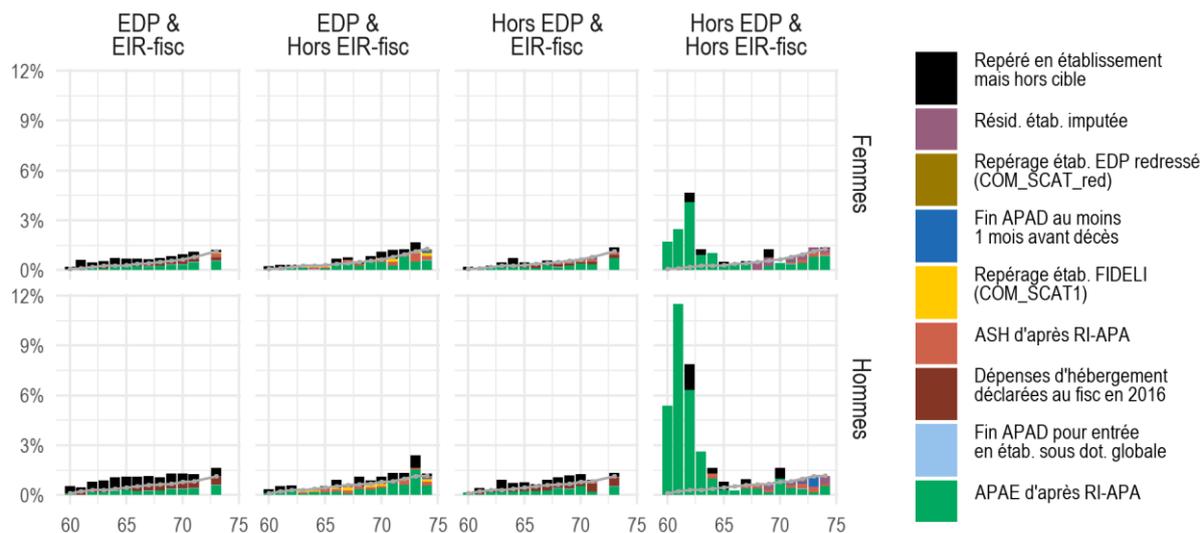
Source > Appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee.

À l'issue de cette étape, 5 175 individus dans la base sont imputés comme bénéficiaires de l'APA en établissement, dont 4 147 pour lesquels la résidence en établissement est repérée grâce aux informations auxiliaires (dépenses d'hébergement déclarées au fisc, bénéfice de l'ASH, repérage dans FIDELI, sortie de l'APA à domicile hors décès) et 1 028 pour lesquels le bénéfice de l'APA en établissement est entièrement imputé. Par ailleurs (pour rappel), 4 788 individus étaient déjà observés comme bénéficiaire de l'APA en établissement dans les RI-APA-ASH appariées à l'EIR.

À l'inverse, 2 286 individus sont repérés en établissement d'après les informations auxiliaires mais ne sont pas considérés comme bénéficiaires de l'APA en établissement, car une telle imputation aboutirait à une prévalence de l'APA en établissement trop élevée, par rapport aux données de cadrage, aux âges de ces individus. Il s'agit souvent de personnes relativement jeunes. On fait l'hypothèse qu'elles correspondent soit aux personnes résidant en établissement pour personnes âgées mais non bénéficiaires de l'APA¹¹, soit à des personnes résidant en réalité dans des foyers pour personnes handicapées.

¹¹ Les résidents en EHPAD catégorisés en GIR 5 ou 6, et à ce titre non éligibles à l'APA, représentent par exemple 6,6 % de l'ensemble des résidents dans ces établissements fin 2019 d'après l'enquête EHPA de la DREES.

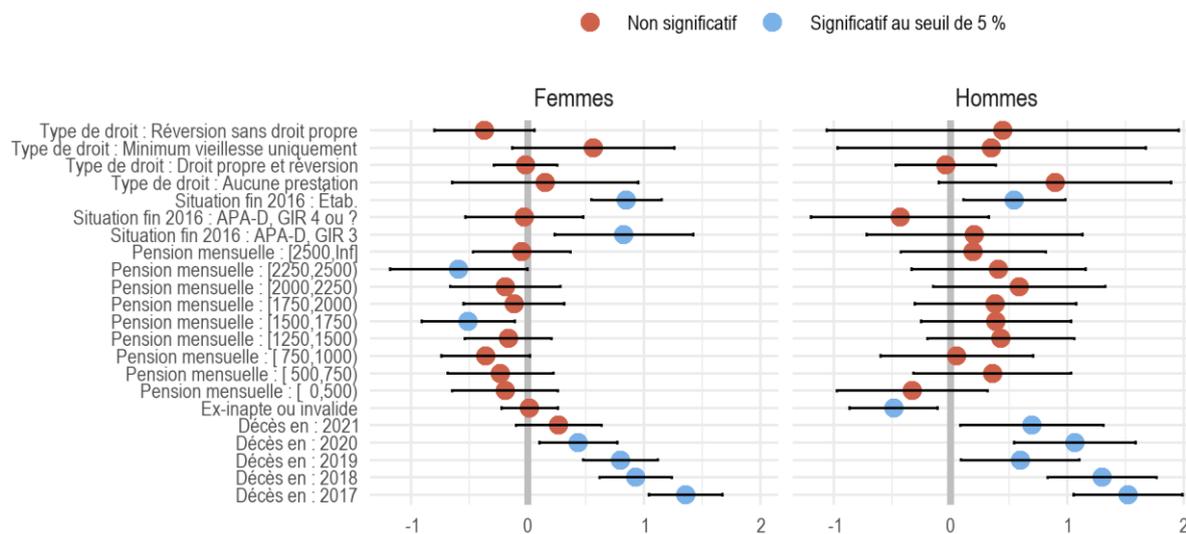
Figure 23 Prévalences de l'APA en établissement en 2017 : valeurs attendues vs. personnes repérées comme résidentes en établissement (personnes de moins de 75 ans)



Champ > Personnes en vie fin 2016, de moins de 75 ans fin 2017. Les âges sont calculés au 31 décembre 2017.
Source > Appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee.

Étape 4 : Pour les personnes considérées comme bénéficiaires de l'APA en établissement en 2017, on impute également le fait d'être en GIR 1 ou 2 plutôt qu'en GIR 3 ou 4. Le modèle est estimé par un *logit* sur le sous-échantillon des personnes en établissement repérées dans les RI-APA-ASH, en utilisant les mêmes variables explicatives que pour l'étape 2, auxquelles on ajoute des indicatrices de situation fin 2016 (bénéfice de l'APA à domicile, en établissement ou non). Les coefficients associés à ces diverses variables explicatives sont représentés dans le graphique ci-après. Le calcul des cibles et l'imputation sont ensuite réalisés sur les mêmes cellules que l'étape précédente (c'est-à-dire pour chaque croisement du sexe, de l'âge, de fait d'être décédé en 2017 ou non, et de l'appartenance au champ de l'EDP et/ou de l'EIR-données fiscales).

Figure 24 Coefficients de la régression logit du fait d'être en GIR 1 ou 2 plutôt que 3 ou 4



Note > Pour plus de lisibilité, la constante du modèle, les coefficients associés à l'âge (polynôme de degré 4), et le coefficient associé au fait d'être déjà en GIR 1 ou 2 dans le cadre de l'APA à domicile fin 2016 (valeur positive très élevée) ne sont pas représentés.

Champ > Bénéficiaires de l'APA en établissement repérés dans les RI-APA-ASH en 2017, dont le GIR est connu.

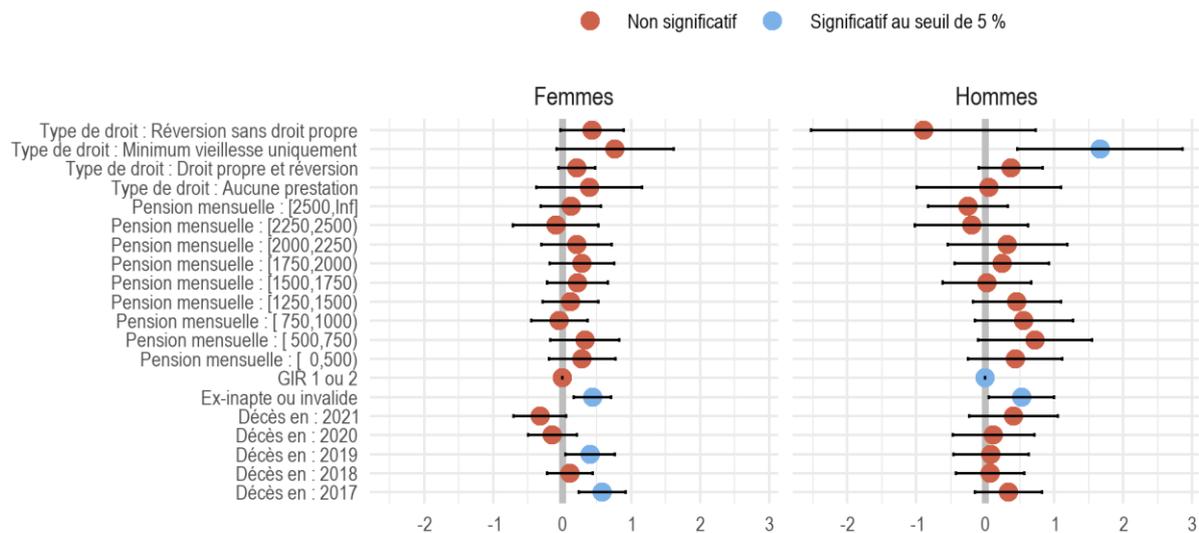
Source > Appariement EIR + RI-APA-ASH, DREES et Insee.

Étape 5 : On impute de même une probabilité d'être déjà dans l'APA en établissement fin 2016 pour les bénéficiaires en 2017. Le modèle est, comme pour l'étape précédente, estimé par un *logit* sur le sous-échantillon des personnes en établissement repérées dans les RI-APA-ASH, en utilisant les mêmes variables explicatives que pour l'étape 2, auxquelles on ajoute une indicatrice de GIR 1 ou 2. Les coefficients associés à ces diverses variables explicatives sont représentés dans le graphique ci-après. À partir des probabilités modélisées, un effectif-cible de bénéficiaires de l'APA en établissement qui l'étaient déjà fin 2016 est ensuite calculé, et la résidence en

établissement fin 2016 est imputée au niveau individuel de façon à atteindre cet effectif cible (en retenant en priorité les personnes dont on sait déjà qu'elles étaient en établissement à cette date, compte tenu des autres informations disponibles : par exemple, une personne qui déclarait des dépenses d'hébergement au fisc au titre de 2016 est considérée comme déjà en établissement en 2016).

Si les cibles sont déjà atteintes au sein d'une cellule d'imputation, les personnes sont considérées comme non-bénéficiaires de l'APA en établissement fin 2016, *même si* elles sont repérées par ailleurs comme résidentes dès 2016 d'après les autres informations. On pourra donc trouver dans la base des observations pour lesquelles le droit à l'APA en établissement n'est ouvert qu'en 2017, alors même que la personne est déjà en établissement en 2016 d'après les données fiscales ou le repérage des communautés de l'EDP. Remarquons que cela n'est pas un *artefact* lié à la simulation : de telles situations sont en réalité également observées pour les bénéficiaires de l'APA en établissement qui sont bien repérés dans les RI-APA-ASH.

Figure 25 Coefficients de la régression logit du fait qu'un bénéficiaire de l'APA en établissement en 2017 l'était déjà fin 2016



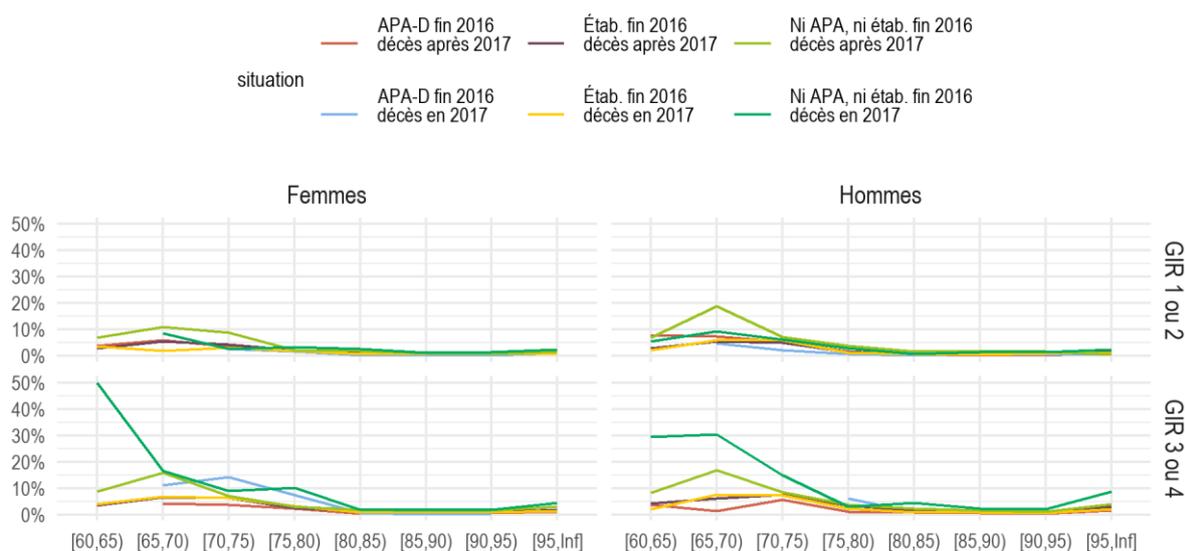
Note > Pour plus de lisibilité, la constante du modèle, le coefficient associé à l'âge n'est pas représenté.

Champ > Bénéficiaires de l'APA en établissement repérés dans les RI-APA-ASH en 2017.

Source > Appariement EIR + RI-APA-ASH, DREES et Insee.

Étape 6 : On crée une table de "donneurs" potentiels parmi les personnes bénéficiaires de l'APA en établissement dans les RI-APA-ASH qui n'ont pas été retrouvés dans l'EIR (afin que les caractéristiques imputées soient distinctes de celles déjà observées dans la table appariée EIR + RI-APA-ASH). Le graphique suivant illustre les rapports entre les nombres de receveurs (c'est-à-dire les individus qu'on a imputés comme bénéficiaires de l'APA en établissement en 2017 mais qui sont absents des RI-APA-ASH) et le nombre de donneurs potentiels, pour chaque sous-groupe sur lequel l'imputation sera réalisée. Les sous-groupes sont définis par le croisement des caractéristiques suivantes : sexe, âge (en regroupant les plus de 100 ans), situation fin 2016 (APA à domicile, APA en établissement, non-bénéficiaire de l'APA), GIR 1 ou 2 (vs. 3 ou 4), fait d'être décédé en 2017 ou après.

Figure 26 Rapport entre le nombre de “receveurs” et le nombre de “donneurs” potentiels pour l’imputation des caractéristiques d’APA en établissement, par âge et champ



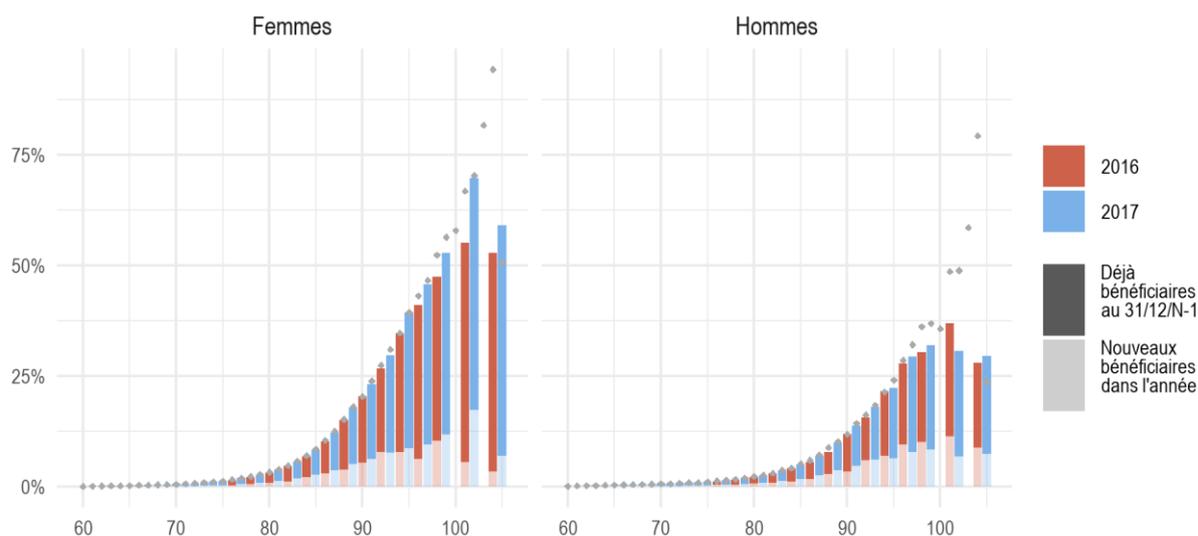
Champ > Personnes repérées comme en établissement en 2017. Données non pondérées.

Source > Appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee.

Étape 7 : Pour chaque individu de la table appariée, reconnu comme bénéficiaire de l’APA en établissement en 2017 (par repérage d’après d’autres variable ou par imputation) mais absent des RI-APA-ASH, on impute des caractéristiques de l’APA en établissement en sélectionnant aléatoirement une personne dans les RI-APA-ASH hors champ de l’EIR, et en lui attribuant les mêmes caractéristiques. L’individu “donneur” est sélectionné parmi les personnes de même sexe, même âge (en regroupant toutes les personnes de 100 ans ou plus), même GIR (1 et 2 vs. 3 et 4), même situation fin 2016 (bénéficiaire de l’APA à domicile, déjà bénéficiaire de l’APA en établissement, non-bénéficiaire de l’APA), et même date de décès (2017 ou plus tard). Au sein de ces individus de même caractéristiques, la sélection est réalisée en prenant les individus par ordre d’apparition dans la base des “donneurs”, après un tri aléatoire de celle-ci. Une fois cela-fait, on corrige certaines variables imputées pour les rendre cohérentes avec les variables observées. Plus précisément, la date d’ouverture des droits à l’APA en établissement est redressée pour être égale au jour suivant la fermeture des droits à l’APA à domicile, si la personne a été observée à domicile dans les RI-APA-ASH, avec une date de fermeture des droits ayant lieu en 2017.

Le graphique suivant représente les prévalences de l’APA en établissement fin 2016 et fin 2017 en incluant les informations imputées, comparées aux prévalences de référence fin 2017. Notons qu’il n’y a pas égalité stricte aux âges les plus élevés : c’est lié notamment au faible nombre de personnes disponibles à ces âges dans la base, qui ne permet pas de se caler *exactement* sur les prévalences attendues. En outre, le calage est réalisé se le fait de percevoir l’APA *courant* 2017, ce qui ne garantit pas de retrouver strictement la prévalence attendue sur le champ des bénéficiaires *fin* 2016 ou fin 2017.

Figure 27 Prévalences de l'APA en établissement dans la population fin 2016 et fin 2017 après imputation



Note > Les points sur le graphique indiquent les prévalences de l'APA en établissement attendues fin 2017, d'après les données des RI-APA-ASH recalées sur le nombre total de bénéficiaires de l'APA en établissement dans l'enquête *Aide sociale*. Les nouveaux bénéficiaires dans l'année correspondent aux personnes dont le droit à l'APA en établissement a été ouvert entre le 1^{er} janvier et le 31 décembre de l'année.

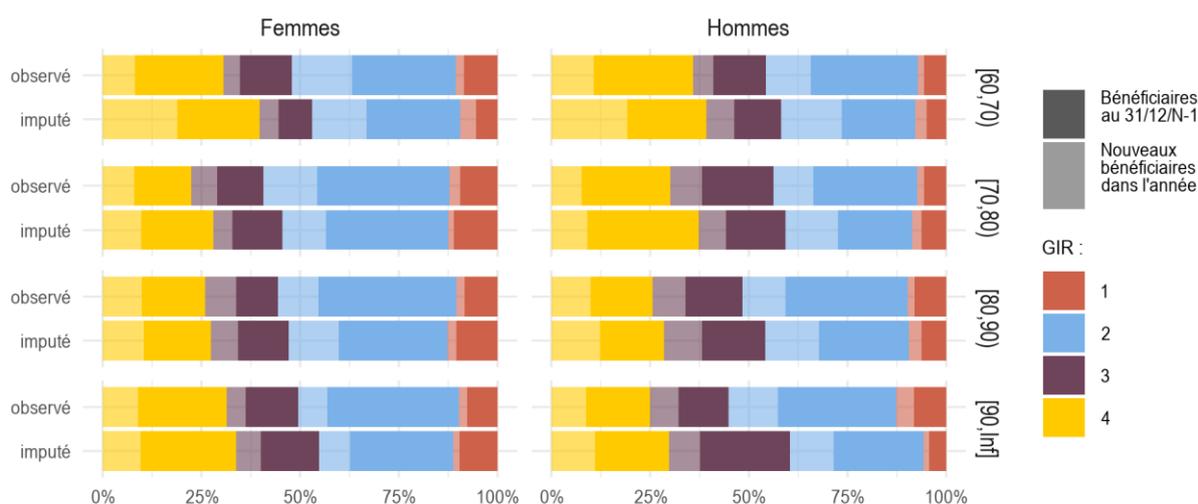
Champ > Personnes vivantes fin 2016 resp. fin 2017.

Source > Appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee.

Le graphique suivant compare quant à lui les répartitions par GIR des bénéficiaires de l'APA en établissement fin 2017 selon qu'ils sont observés dans les RI-APA-ASH ou bien qu'ils aient été imputés. Le but est de vérifier que l'imputation ne déforme pas trop cette répartition par GIR. Les différentes catégories de GIR sont en outre distinguées selon qu'elles concernent des nouveaux bénéficiaires de l'APA en établissement au cours de l'année 2017 (fond plus clair) ou des personnes qui étaient déjà bénéficiaires fin 2016 (fond plus foncé).

Les proportions de GIR 1 semblent similaires entre les bénéficiaires observés et imputés, mais les proportions de GIR 2 semblent, dans certaines classes d'âge (notamment les moins de 80 ans pour les deux sexes, et les plus de 90 ans parmi les hommes), légèrement moins nombreux parmi les bénéficiaires imputés, au profit des GIR 4. La proportion de nouveaux bénéficiaires de l'APA en établissement en 2017 semble par ailleurs plus élevée parmi les bénéficiaires imputés que parmi les bénéficiaires observés, notamment parmi les hommes.

Figure 28 Répartition par GIR des bénéficiaires de l'APA en établissement fin 2017, par sexe, tranche d'âge et statut d'imputation



Lecture > Parmi les femmes bénéficiaires de l'APA en établissement ayant entre 60 ans (inclus) et 70 ans (exclus), sur le champ des bénéficiaires observés dans les RI-APA (ligne "observé"), 8 % sont en GIR 4 et sont de nouvelles bénéficiaires au cours de l'année et 22 % sont en GIR 4 et étaient déjà bénéficiaires au 31 décembre de l'année précédente. Pour le même groupe sur le champ des bénéficiaires qui ont été imputés (ligne "imputé"), 19 % sont en GIR 4 et sont de nouvelles bénéficiaires au cours de l'année, et 21 % sont en GIR 4 et étaient déjà bénéficiaires au 31 décembre de l'année précédente.

Champ > Bénéficiaires de l'APA en établissement au 31 décembre 2017 (observés ou imputés), hors ceux pour lesquels le GIR est inconnu.

Source > appariement EIR + données fiscales + EDP + RI-APA-ASH, DREES et Insee.

■ PONDÉRATION

De façon générale, la pondération d'une base statistique vise à rendre celle-ci davantage représentative de son champ, qui correspond ici à l'ensemble des personnes vivant en France en ayant au moins 60 ans au 31 décembre 2016, que ces dernières soient bénéficiaires de prestations (retraite, minimum vieillesse, APA, ASH) ou non. La pondération peut chercher à corriger deux types de défauts de représentativité : un défaut "initial" lié à l'échantillonnage des données, si les probabilités d'inclusion dans l'échantillon dépendent de certaines caractéristiques des individus, et un défaut lié à la collecte, si les "trous de collecte" sont plus fréquents pour certains types d'individus (ils proviennent, ici, des données manquantes dans l'une ou l'autre des bases appariées et du fait que tous les individus n'ont pas pu être appariés).

Le défaut initial peut être corrigé de façon *a priori*, à partir de la connaissance du plan de sondage, par des poids égaux à l'inverse des probabilités de tirage. Ce défaut est important ici du fait du schéma spécifique d'échantillonnage de l'EIR : la correction est donc nécessaire. Le tirage de l'EIR fait en effet l'hypothèse que le jour de naissance au sein de l'année constitue un tirage aléatoire (ce qui n'est pas tout à fait le cas) et, surtout, retient un nombre de jours de naissance qui varie selon l'année et le trimestre de naissance. La correction est décrite dans la première sous-partie ci-après, qui aboutit à la création de pondérations qui seront dites "initiales".

Le second défaut de représentativité ne peut généralement se constater qu'*a posteriori*, par comparaison des agrégats calculés à partir de la base avec des données de cadrage. Il est généralement corrigé par un calage sur marges. Pour l'appariement EIR-RI APA, une double comparaison doit être faite : avec des agrégats relatifs aux retraites, tirés de l'EIR, et avec des agrégats relatifs aux prestations d'autonomie, tirés des RI-APA-ASH. Remarquons toutefois que la base qui est documentée ici vise à permettre des travaux d'analyse sur les liens entre caractéristiques de retraite ou de carrière et caractéristiques d'autonomie, et pas à produire des agrégats statistiques de référence ; il n'est donc pas forcément indispensable que les principaux agrégats soient *exactement* égaux aux données de cadrage. Dans la seconde sous-partie ci-après, on cherchera donc surtout à comparer les données de cadrage aux données calculées à partir de la base appariée en utilisant les pondérations initiales, afin de déterminer si celles-ci permettent déjà de rendre la base suffisamment représentative de son champ, ou si un calage sur marge est en outre nécessaire.

Pondérations initiales, pour tenir compte de l'échantillonnage de l'EIR

L'EIR ne sélectionne qu'une partie des générations, et ne retient pas les mêmes jours de naissance pour chaque génération (selon l'année de naissance, entre 3 et 36 jours parmi les 365 de l'année sont retenus). En outre, sur le champ des personnes échantillonnées dans l'EIR mais non bénéficiaires d'une retraite ou du minimum vieillesse, on a mobilisé les données de l'EDP, et la sélection dans le champ de l'EDP ne retient pour certaines générations qu'une partie des jours de naissance échantillonnés de l'EIR.

Il est donc nécessaire de pondérer les observations pour tenir compte du plan de sondage asymétrique, afin d'éviter que certaines générations se voient artificiellement donner (du fait d'un nombre de jours de tirages plus élevés) un poids supérieur à leur poids réel dans la population des personnes de 60 ans et plus. Ces pondérations sont qualifiées "d'initiales", car elles tiennent compte uniquement du plan de sondage et de la population totale résident en France, et non d'un calage sur marge sur des variables d'intérêt (relatives aux retraites ou aux prestations d'autonomie).

Deux jeux de pondérations sont créés, l'un pour rendre l'échantillon représentatif de chaque année de naissance présente dans la base, l'autre pour le rendre représentatif de la population ventilée par âge quinquennal. Dans le second cas, les générations absentes de l'EIR sont "représentées" en extrapolant les générations adjacentes ; par exemple, les observations relatives à la génération 1934 voient leur poids multiplié par 2 afin que ces observations représentent à la fois la génération 1934, la génération 1933 pour moitié (l'autre moitié étant représentée par extrapolation des observations de personnes nées en 1932) et la génération 1935 également pour moitié (l'autre moitié étant représentée par extrapolation des nés en 1936). Les variables de pondérations correspondantes s'intitulent respectivement *pond_recalpop_gen* (générations représentées dans l'EIR) et *pond_recalpop_agequin* (population totale ventilée par âge quinquennal). Cette dernière a été calculée à la fois pour être représentatif de la population par tranche d'âge quinquennal au 31 décembre 2016, mais aussi au 31 décembre 2017 ; la pondération correspond donc en réalité à deux variables, selon la date de référence, *pond_recalpop_agequin2016* et *pond_recalpop_agequin2017*.

Le calcul de ces pondérations est réalisé en deux temps. Dans un premier temps, on considère que chaque observation est représentative de l'ensemble des jours de naissance du trimestre (en considérant des trimestres de 365/4 jours), en fonction inverse du nombre de jours de naissance effectivement sélectionnés dans l'EIR. Puisque l'appariement avec la base RI-APA-ASH est disponible pour toutes les personnes échantillonnées dans la vague 2016 ou la vague 2012 de l'EIR, on retient le nombre maximal de jours de naissance sélectionnés entre ces deux vagues. Pour les personnes dont le mois de naissance est inconnu (variable *MNI* égale à 0 dans l'EIR), qui ont été sélectionnées d'après leur clé de NIR, on retient l'inverse du nombre de clés de NIR sélectionnées parmi les 97 disponibles. Par exemple :

- pour les personnes nées en octobre 1946, les nés du 1^{er} au 24 octobre sont retenus dans l'EIR 2012 et ceux du 1^{er} au 10 dans l'EIR 2016 : si elles sont observées comme retraitées ou bénéficiaires du minimum vieillesse dans l'EIR (2012 ou 2016), on considère donc que les observations correspondantes représentées (365/4) / 24 personnes ;
- si ces personnes n'ont aucune retraite ou minimum vieillesse dans l'EIR, et sont observées uniquement dans les données de l'EDP, elles ne peuvent être nées qu'entre le 1^{er} et le 4 octobre (critère de sélection de l'EDP pour ce mois de naissance) : on considère alors que les observations représentent (365/4) / 4 personnes ;
- si la personne est née en 1946 mais que son mois de naissance est inconnu dans les données d'état-civil de l'Insee, la sélection dans l'échantillon a été réalisée en retenant 6 clés de NIR parmi les 97 disponibles : on considère dans ce cas que les observations représentent 97 / 6 personnes.

Les poids calculés à l'issue de ce premier temps s'intitulent *pond_corrirage_gen* et *pond_corrirage_agequinAAAA* (ou AAAA = 2016 ou 2017).

Dans un second temps, on récupère les populations officielles de l'Insee aux 1^{er} janvier 2017 et 2018 (supposées égales à celles au 31 décembre 2016 et 2017 respectivement), et on corrige chaque poids par le rapport entre la population officielle (par sexe et âge fin ou quinquennal) et la somme des poids intermédiaires (pour la même catégorie) calculés à l'étape précédente. Ce recalage sur les populations officielles de l'Insee donne les pondérations *pond_recalpop_gen* et *pond_recalpop_agequinAAAA*.

On a également produit un jeu de pondérations sur le champ restreint aux individus qui font en outre partie du champ de l'échantillon démographique permanent (EDP). Les poids correspondants sont suffixés *edpgen* et *edpgequin* respectivement.

Examen de la pertinence d'un calage sur marge sur des variables d'intérêt

Afin d'apprécier s'il est besoin de réaliser un calage sur marge sur divers agrégats, on confronte quelques grandeurs calculées à partir de la base appariée EIR + RI-APA-ASH en utilisant les pondérations décrites dans la sous-section précédente avec les mêmes grandeurs calculées à partir des bases complètes. On s'intéresse en premier lieu à la cohérence avec les bases sources utilisées pour l'appariement : en ce qui concerne les prestations d'autonomie, on se compare donc aux RI-APA-ASH complètes, et non aux données de l'enquête *Aide sociale*.

Les comparaisons pour l'APA et l'ASH portent sur le champ de la base diffusée au CASD, c'est-à-dire les bénéficiaires d'un droit encore ouvert au 31 décembre 2017. Le champ diffère donc de celui qui a été analysé jusqu'ici, qui portait dans la plupart des illustrations sur la situation fin 2016. Par ailleurs, le bénéfice des prestations est apprécié différemment selon les bases : pour la base appariée, EIR + RI-APA-ASH, ce bénéfice (en droits ouverts) est calculé d'après les dates d'ouverture et de fermeture des droits, conformément à ce qui était fait dans les graphiques des parties précédentes¹², mais il est apprécié à partir des variables *BENAPA* (pour l'APA à domicile) ou *pres_APA_imput* et *pres_ASH* (construites par la DREES pour l'APA en établissement et l'ASH) pour les bases RI-APA-ASH exhaustives. Pour la base appariée, on utilise les pondérations recalées sur la population par âge quinquennal fin 2017 (variable *pond_recalpop_agequin2017*). Pour les RI-APA-ASH, on utilise les pondérations *ponds_p* construites par la DREES dans les bases statistiques diffusées (calculées pour rendre les bases représentatives du nombre de bénéficiaires payés de l'APA en France entière fin 2017).

Le nombre de bénéficiaire de l'APA à domicile fin 2017 dans la base appariée est inférieur d'environ 3 % à celui observé dans les RI-APA-ASH (figure 29), soit un taux très proche de la proportion de non-appariés (cf. parties précédentes). L'écart est un peu plus fort pour l'ASH (-7 %), mais il est quasi-nul pour l'APA en établissement, celle-ci ayant fait l'objet d'imputations pour assurer la cohérence avec les données de cadrage. Dans le premier cas, on a cependant classé tous les bénéficiaires de l'ASH en établissement dans les RI-APA-ASH, alors qu'une petite partie apparaît comme concernant des bénéficiaires de l'APA à domicile¹³ dans la base appariée EIR + RI-APA-ASH ; lorsqu'on corrige cela, l'écart se réduit.

Par ailleurs, la sous-représentation des bénéficiaires de l'APA dans la base appariée apparaît plus marquée pour les GIR les plus bas (GIR 1 et 2), ce qui pourrait s'expliquer par le taux d'appariement moins élevé pour les plus âgés, c'est-à-dire dans les tranches d'âges où la dépendance est la plus forte. L'imputation de l'APA en établissement pour une partie des personnes dans la base ne change pas cela, car la probabilité d'être en GIR 1 ou 2 pour les bénéficiaires imputés est estimée à partir des bénéficiaires déjà observés (grâce à l'appariement) : on reproduit donc la même sous-estimation des catégories de perte d'autonomie la plus lourde. Malgré tout, les répartitions par GIR restent assez proches entre la base appariée et la base exhaustive. Un diagnostic similaire peut être fait pour

¹² Ce choix est notamment réalisé car il permet d'avoir une approche homogène pour définir l'APA fin 2016 ou fin 2017, ces deux dates de référence étant pertinentes pour la base appariée.

¹³ Ce cas de figure est possible d'après la réglementation : dans les établissements de petite taille, par exemple, c'est l'APA à domicile qui est versée, et non l'APA en établissement.

les autres répartitions qui ont été regardées (selon les ressources au sens de l'APA, selon l'ancienneté dans l'APA, selon la part de la dépense à la charge du bénéficiaire...).

Figure 29 Distribution des principales variables d'intérêt relatives aux prestations d'autonomie : comparaison entre l'appariement EIR-RI APA et les RI-APA-ASH exhaustives

Catégorie	Appariement EIR-RI APA	RI-APA-ASH (Domicile)	Écart (Domicile)	Appariement EIR-RI APA (Établissement)	RI-APA-ASH (Établissement)	Écart (Établissement)
Nb APA (ensemble)	745890	766905	-2.7 %	540674	541685	-0.2 %
Nb GIR 1	16231	17584	-7.7 %	56422	86350	-34.7 %
Nb GIR 2	119336	128489	-7.1 %	209639	233282	-10.1 %
Nb GIR 3	154775	172398	-10.2 %	110174	94873	16.1 %
Nb GIR 4	434328	448434	-3.1 %	159051	127180	25.1 %
Part GIR 1 (%)	2.2	2.3	-0.1 pt(s)	10.5	15.9	-5.4 pt(s)
Part GIR 2 (%)	16.5	16.8	-0.3 pt(s)	39.2	43.1	-3.9 pt(s)
Part GIR 3 (%)	21.4	22.5	-1.1 pt(s)	20.6	17.5	3.1 pt(s)
Part GIR 4 (%)	59.9	58.5	1.4 pt(s)	29.7	23.5	6.2 pt(s)
Nb ASH (ensemble)	4251			113262	121349	-6.7 %
Part bénéf. aide humaine (%)	88	93	-5 pt(s)			
Part à la charge bénéf. (%) : [0,1)	20	18.6	1.4 pt(s)			
Part à la charge bénéf. (%) : [1,25)	47.5	48.4	-0.9 pt(s)			
Part à la charge bénéf. (%) : [25,50)	22.9	22.9	0 pt(s)			
Part à la charge bénéf. (%) : [50,75)	6.9	7.1	-0.2 pt(s)			
Part à la charge bénéf. (%) : [75,99)	2.8	2.9	-0.1 pt(s)			
Part ressources (%) : [0,500)	7.3	6.3	1 pt(s)	7.5	6.4	1.1 pt(s)
Part ressources (%) : [500,1000)	23.9	23.6	0.3 pt(s)	21.8	24.3	-2.5 pt(s)
Part ressources (%) : [1000,1250)	19.6	20.2	-0.6 pt(s)	13.7	14.5	-0.8 pt(s)
Part ressources (%) : [1250,1500)	17	17.4	-0.4 pt(s)	13.2	13	0.2 pt(s)
Part ressources (%) : [1500,2000)	19.3	19.4	-0.1 pt(s)	19.6	18.2	1.4 pt(s)
Part ressources (%) : [2000,Inf]	12.8	13.2	-0.4 pt(s)	24.3	23.6	0.7 pt(s)
Part ancienneté (mois) dans l'APA (%) : [0,6)	11.7	10.3	1.4 pt(s)	5.7	5.9	-0.2 pt(s)
Part ancienneté (mois) dans l'APA (%) : [6,12)	11.8	12.2	-0.4 pt(s)	8.6	7.8	0.8 pt(s)
Part ancienneté (mois) dans l'APA (%) : [12,24)	17.3	17	0.3 pt(s)	14.1	13.4	0.7 pt(s)
Part ancienneté (mois) dans l'APA (%) : [24,36)	13.3	13	0.3 pt(s)	12.5	13	-0.5 pt(s)
Part ancienneté (mois) dans l'APA (%) : [36,60)	17.5	18.3	-0.8 pt(s)	21.3	20.6	0.7 pt(s)
Part ancienneté (mois) dans l'APA (%) : [60,Inf]	28.4	29.3	-0.9 pt(s)	37.8	39.2	-1.4 pt(s)

Note > L'ancienneté dans l'APA est calculée comme la différence entre la date d'ouverture des droits DATEOUVAPA (tous types d'APA confondus) et le 31 décembre 2017.

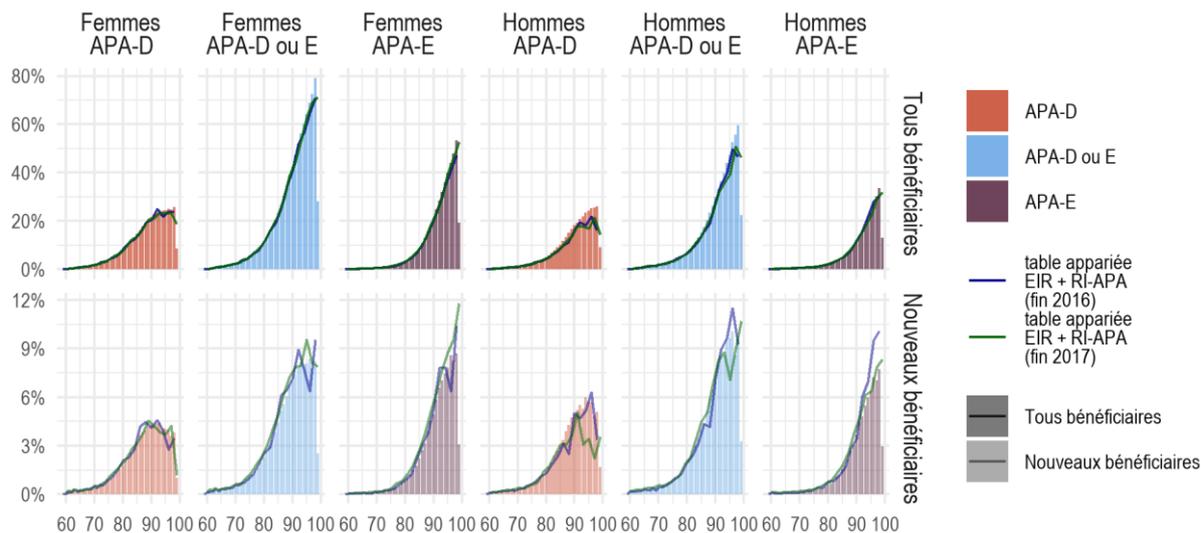
Champ > Personnes ayant un droit ouvert à l'APA ou à l'ASH fin 2017

Source > Appariement EIR + RI-APA-ASH et RI-APA-ASH complètes, DREES.

Le graphique suivant compare, plus finement, les proportions de bénéficiaires et de nouveaux bénéficiaires de l'APA en fin d'année calculées pour chaque âge fin. Pour les RI-APA-ASH, les prévalences sont estimées en rapportant les nombres totaux de bénéficiaires aux populations résidentes d'après le bilan démographique de l'Insee.

Les prévalences sont très proches, même si elles apparaissent un peu plus basses pour les hommes bénéficiaires de l'APA à domicile après 90 ans, et un peu plus hautes pour les nouveaux bénéficiaires de l'APA, notamment en établissement (figure 30).

Figure 30 Prévalences des bénéficiaires et des nouveaux bénéficiaires de l'APA fin 2017, par sexe et âge



Note > Âges au 31 décembre 2017 (resp. 2016 pour la courbe "situation fin 2016"). Pour des raisons de lisibilité du graphique, seuls les âges avant 100 ans sont représentés. Les nouveaux bénéficiaires de l'APA correspondent ici aux personnes dont le droit à l'APA (respectivement à l'APA à domicile ou à l'APA en établissement) a été ouvert en cours d'année, et qui sont encore en vie au 31 décembre de l'année.

Champ > Personnes en vie et résident en France fin 2017 (sauf courbe "situation fin 2016").

Source > Appariement EIR + RI-APA-ASH et RI-APA-ASH, DREES ; bilan démographique, Insee.

Pour ce qui concerne les pensions de retraite, on se compare à l'EIR de 2016 (table INDIV). Pour certaines variables (relatives aux conditions de liquidation et à la carrière complète), on a utilisé les valeurs de l'EIR "générationnel", c'est-à-dire la base construite par empilement de toutes les vagues de l'EIR, qui peuvent différer des valeurs renseignées dans la table de diffusion de l'EIR 2016 du fait de redressements et imputations supplémentaires.

On s'attend ici à des résultats beaucoup plus proches entre les bases, car la table appariée EIR + RI-APA contient *a priori* toutes les observations de l'EIR pour les personnes de 60 ans et plus résidant en France. Les seules différences tiennent aux pondérations, ainsi qu'à l'inclusion dans la base appariée de quelques générations observées dans la vague 2012 de l'EIR (générations nées en 1912, 1943 et 1945). La pondération ne tient compte que du plan de sondage et des populations Insee pour la base appariée, tandis qu'elle a été recalée sur les données de cadrage de l'enquête annuelle auprès des caisses de retraite (EACR) pour l'EIR 2016 (variable *pond16_cal* de cette base).

En pratique, pour les quelques variables d'intérêt considérées dans le tableau (ventilation par type de droit, par circonstance de liquidation, par montant de pension) les proportions sont quasi-identiques dans les deux bases (figure 31). Les écarts sont au plus de 0,6 point de pourcentage, et ne dépassent généralement pas 0,1 ou 0,2 point.

Figure 31 Distribution des principales variables d'intérêt relatives aux retraites : comparaison entre l'appariement EIR-RI APA et l'EIR 2016

Catégorie	Appariement EIR-RI APA	EIR 2016	Écart
Droit propre de retraite	97.1	97.3	-0.2 pt(s)
Minimum vieillesse uniquement	0.5	0.4	0.1 pt(s)
Réversion sans droit propre	2.5	2.2	0.3 pt(s)
Carrière longue	9	9.3	-0.3 pt(s)
Décote	6	5.9	0.1 pt(s)
Décote non applicable	13.4	13.4	0 pt(s)
Surcote	6.3	6.4	-0.1 pt(s)
Tx plein âge	14.2	14.1	0.1 pt(s)
Tx plein autre	0.6	0.7	-0.1 pt(s)
Tx plein durée	35.4	35.3	0.1 pt(s)
Tx plein inaptitude	8.7	8.6	0.1 pt(s)
Tx plein invalidité	5.7	5.6	0.1 pt(s)
Inconnu	0.6	0.6	0 pt(s)
Part reconnus inaptes (%)	15.7	15.7	0 pt(s)
Part pension DD (%) : [0,500)	16.2	15.6	0.6 pt(s)
Part pension DD (%) : [500,1000)	22.6	22.8	-0.2 pt(s)
Part pension DD (%) : [1000,1250)	10.6	10.8	-0.2 pt(s)
Part pension DD (%) : [1250,1500)	11.3	11.2	0.1 pt(s)
Part pension DD (%) : [1500,2000)	17.4	17.6	-0.2 pt(s)
Part pension DD (%) : [2000,Inf]	21.9	22.1	-0.2 pt(s)
Part carrières complètes (%)	67.7	67.7	0 pt(s)
Coef. de prorat. moyen (%)	94.7	94.8	-0.1 pt(s)

Champ > Personnes ayant une prestation dans l'EIR (hors volet invalidité) fin 2016.

Source > Appariement EIR + RI-APA-ASH et RI-APA-ASH complètes, DREES.

Comme on l'a déjà souligné, la base appariée EIR + RI-APA-ASH vise en premier lieu à permettre des études économétriques sur les relations entre caractéristiques de retraite et de perte d'autonomie, et pas à calculer de statistiques officielles (agrégats). Pour cette raison, il n'est pas forcément nécessaire que la base coïncide *exactement* avec les données de cadrage. En pratique, comme l'analyse réalisée montre que, pour les indicateurs considérés, les résultats restent proches de ceux tirés des données sources (pour les RI-APA-ASH et plus encore pour l'EIR), **on ne réalise pas de calage sur marge. Les pondérations à utiliser sont donc les pondérations "initiales", qui corrigent le schéma d'échantillonnage des données et se recalent sur la population des 60 ans plus vivant en France fin 2016 ou fin 2017.** Elles correspondent aux variables *pond_recalpop_agequinAAAA* (pour AAAA = 2016 ou 2017 si l'on souhaite être représentatif de la population ventilée par âge quinquennal fin 2016 ou fin 2017) et *pond_recalpop_gen*, selon que l'on travaille sur la population globale ventilée par grandes classes d'âge (*pond_recalpop_agequinAAAA*) ou qu'on réalise des analyses par génération (*pond_recalpop_gen*).

■ LISTE DES VARIABLES

Cette partie détaille les différentes variables retenues dans la table appariée qui sera diffusée au *Centre d'accès sécurisé aux données* (CASD). Ces variables sont distinguées selon la base dont elles sont originaires.

La présentation est ici succincte, dans la mesure où la plupart des variables sont issues des données sources qui ont été appariées. On se reportera donc aux documents méthodologiques de l'EIR et des RI-APA-ASH, et le cas échéant de l'EDP, pour davantage de précisions.

Identifiants et pondérations

nom_variable	type	descriptif
pond_corrirage_gen	num.	Pondération correctrice du plan de sondage, permettant d'être représentatif des générations échantillonnées
pond_corrirage_agequin2016	num.	Pondération correctrice du plan de sondage, permettant d'être représentatif de la population par tranches d'âge quinquennal fin 2016
pond_corrirage_agequin2017	num.	Pondération correctrice du plan de sondage, permettant d'être représentatif de la population par tranches d'âge quinquennal fin 2017
pond_recalpop_gen	num.	Pondération recalée sur la population Insee, permettant d'être représentatif des générations échantillonnées
pond_recalpop_agequin2016	num.	Pondération recalée sur la population Insee, permettant d'être représentatif de la population par tranches d'âge quinquennal fin 2016
pond_recalpop_agequin2017	num.	Pondération recalée sur la population Insee, permettant d'être représentatif de la population par tranches d'âge quinquennal fin 2017
pond_corrirage_ed-pgen	num.	Pondération correctrice du plan de sondage, permettant d'être représentatif des générations échantillonnées (champ restreint aux jours EDP)
pond_corrirage_ed-pagequin2016	num.	Pondération correctrice du plan de sondage, permettant d'être représentatif de la population par tranches d'âge quinquennal fin 2016 (champ restreint aux jours EDP)
pond_corrirage_ed-pagequin2017	num.	Pondération correctrice du plan de sondage, permettant d'être représentatif de la population par tranches d'âge quinquennal fin 2017 (champ restreint aux jours EDP)
pond_recalpop_ed-pgen	num.	Pondération recalée sur la population Insee, permettant d'être représentatif des générations échantillonnées (champ restreint aux jours EDP)
pond_recalpop_ed-pagequin2016	num.	Pondération correctrice du plan de sondage, permettant d'être représentatif de la population par tranches d'âge quinquennal fin 2016 (champ restreint aux jours EDP)
pond_recalpop_ed-pagequin2017	num.	Pondération correctrice du plan de sondage, permettant d'être représentatif de la population par tranches d'âge quinquennal fin 2017 (champ restreint aux jours EDP)
age	num.	Âge au 31/12/2016
agequin2016	factor	Tranche d'âge quinquennal au 31/12/2016
agequin2017	factor	Tranche d'âge quinquennal au 31/12/2017
champ_resid	char.	Champ et source de l'observation de la résidence en France
proba_champ_edp	num.	Probabilité de présence dans le champ de l'EDP
source_appariement	char.	Méthode d'appariement entre l'EIR et les RI-APA-ASH

Variables issues de RI-APA-ASH

Les variables retenues dans les RI-APA-ASH correspondent pour l'essentiel aux dates d'ouverture et de fermeture des droits, au GIR, à la base ressource (pour le calcul du ticket modérateur de l'APA) et aux paiements d'aide humaine. On a également inclus quelques variables issues du répertoire FINESS, pour les bénéficiaires de l'APA en établissement dont le FINESS de l'établissement de résidence a été fourni par le département.

Pour les autres informations, notamment le détail des aides non-humaines (aides techniques, aides aux aidants, etc.) et les axes ayant servi au calcul du GIR, les variables doivent être récupérées en réappariant la base avec les RI-APA-ASH sur la base de l'identifiant *ID*.

nom_variable	type	descriptif
ID_APAD	char.	Identifiant (n° d'ordre) de l'individu dans les RI-APA-ASH pour les informations relatives à l'APA à domicile
ID_APAE	char.	Identifiant (n° d'ordre) de l'individu dans les RI-APA-ASH pour les informations relatives à l'APA en établissement
ID_ASH	char.	Identifiant (n° d'ordre) de l'individu dans les RI-APA-ASH pour les informations relatives à l'ASH
qualiteCSNS_ApaAsh	char.	Indicateur de qualité du CSNS dans les RI-APA-ASH
source_apa	char.	Fichier source des informations sur l'APA
source_ash	char.	Fichier source des informations sur l'ASH
presta_apad_fin2016	logi-	Indicatrice de perception de l'APA à domicile au 31/12/2016
presta_apae_fin2016	logi-	Indicatrice de perception de l'APA en établissement au 31/12/2016
presta_ash_fin2016	logi-	Indicatrice de perception de l'ASH au 31/12/2016
presta_apaash_fin2016	logi-	Indicatrice de perception de l'APA ou de l'ASH au 31/12/2016
prestation_fin2016	fac-	Type de prestation perçue d'après les RI-APA-ASH, au 31/12/2016
BENAPA	char.	Bénéficiaire de l'APA en 2017
BENDPDEC	num.	
AXE_COH	char.	Axe "Cohérence"
AXE_ORI	char.	Axe "Orientation"
AXE_TOI	char.	Axe "Toilette"
AXE_HAB	char.	Axe "Habillage"
AXE_ALI	char.	Axe "Alimentation"
AXE_ELI	char.	Axe "Élimination"
AXE_TRANS	char.	Axe "Transferts"
AXE_DEP	char.	Axe "Déplacement à l'intérieur"
AXE_EXT	char.	Axe "Déplacement à l'extérieur"
AXE_Com	char.	Axe "Communication à distance"
sequence_TY-	char.	Séquence des types d'APA (Domicile ou Établissement) lors des évaluations successives
sequence_GIR_eval	char.	Séquence des GIR lors des évaluations successives
DATEPREDEEAPA	Date	Date de première demande APA
DATEOUVAPA	Date	Date ouverture des droits APA
TYPAPA	char.	Type d'APA
TYPAPA_dec2016	char.	Type d'APA (domicile ou établissement) au 31/12/2016
GIRFIN_APA	char.	Gir
GIR_dec2016	char.	GIR au 31/12/2016
APA	num.	Montant mensuel du plan d'aide APA notifié
NOTPB	num.	Montant mensuel de la participation financière notifiée du bénéficiaire APA
NOTHPB	num.	Montant mensuel de l'APA notifié relevant du conseil départemental
AIDEHUM	char.	Aide humaine dans le plan d'aide APA notifié
DATE_APAD	Date	Date d'ouverture des droits à l'APA à domicile
FERM_APA_D	Date	Date de fermeture des droits à l'APA à domicile
DATE_APAE	Date	Date d'ouverture des droits à l'APA en établissement
FERM_APA_E	Date	Date de fermeture des droits à l'APA en établissement

MOTIF_FERM_APA	char.	Motif de sortie des droits à l'APA
type_redressement_date	char.	Type de redressement(s) réalisé(s) sur les variables des dates relatives à l'APA ou l'ASH
DATE_premevalAPAD	Date	Date de la première évaluation pour l'APA à domicile
DATE_premevalAPAE	Date	Date de la première évaluation pour l'APA en établissement
DATE_premevalAPA	Date	Date de la première évaluation pour l'APA (domicile ou établissement)
DATE_premevalGIR12	Date	Date de la première évaluation pour l'APA en GIR 1 ou 2 (domicile ou établissement)
DATE_derevalAPA	Date	Date de la dernière évaluation pour l'APA (domicile ou établissement)
RETRAITE_APA	num.	Dont pension de retraite
RESSOURC_APA	num.	Ressources au sens de l'APA
RESSOURC_APA_AN	num.	Année de l'évaluation des ressources
AIDEHUM_PRESTA	char.	Aide humaine : Prestataire
AIDEHUM_MAND	char.	Aide humaine : Mandataire
AIDEHUM_GRE	char.	Aide humaine : Gré à gré ou emploi direct
AIDEHUM_HEURE	num.	Nombre d'heures d'aide humaine notifié
AIDEHUM_NOT	num.	Montant mensuel d'aide humaine
DECHPB_APA	num.	Montant total de l'APA versé par le conseil départemental
DECHUM_APA	num.	Montant de l'APA versé par le CD pour l'aide humaine
DECNONHUM_APA	num.	Montant de l'APA versé par le CD, hors aide humaine
BENASH	char.	Bénéficiaire de l'ASH en 2017
DATE_ETAB_ASH	Date	Date d'entrée en établissement ASH
DATE_FERM_ASH	Date	Date de fermeture des droits à l'ASH
DATEOUVASH	Date	Date ouverture des droits ASH
HANDL_ASH	char.	Reconnaissance handicapé des bénéficiaires de l'ASH
DECPB_ASH	num.	ASH : Montant de la participation du bénéficiaire
DECHPB_ASH	num.	Montant ASH versé par le conseil départemental
MOTIF_FERM_ASH	char.	Motif de sortie de l'ASH
JURI	char.	Protection juridique
NATJURI	char.	Nature de la protection juridique
FINESS	char.	Numéro Finess de l'établissement
libde	char.	Libellé de la discipline d'équipement de l'établissement d'hébergement d'après le répertoire FINESS (si disponible)
libclient	char.	Libellé du code clientèle de l'établissement d'hébergement d'après le répertoire FINESS (si disponible)

Variables issues de l'EIR

Les variables de l'EIR proviennent pour certaines de la table de diffusion de la vague 2016 (base *INDIV16*) et pour d'autres de la base "générationnelle", construite par empiement et synthèse de toutes les vagues de l'EIR (table *INDIVGEN16*). Des redressements supplémentaires sont réalisés dans cette dernière base, dans une optique d'homogénéisation entre toutes les vagues, si bien que certaines variables peuvent prendre des valeurs différentes, pour un individu donné, de celles renseignées dans la table relative à la seule vague 2016. Quelques variables de synthèse (nombre d'enfants, âge de début et de fin de carrière notamment) ont en outre été construites à partir des informations de base de l'EIR : dans certains cas, des redressements et imputations ont été nécessaires, et on se reportera donc à la partie précédente pour le détail de ces traitements.

nom_variable	type	descriptif
newnordre	char.	Identifiant (n° d'ordre) de l'individu dans l'EIR
vagues_echinsee	char.	Liste des vagues de présence dans l'échantillon de l'EIR
vagues_eir	char.	Liste des vagues d'observation dans l'EIR
ai	num.	Année de naissance (variable Insee)
naiss	char.	Zone de naissance
resid	factor	Zone de résidence
sexe	char.	Sexe de l'individu (renseigné par le régime principal)
andeces	num.	Année de décès
moisdeces	num.	Mois de décès
dr	char.	Département de résidence (renseigné par le régime principal) :
type_droit_eir	factor	Type de droit perçu dans l'EIR
mtot1	num.	Montant mensuel brut total de l'ensemble des avantages principaux de droit direct
mtot2	num.	Montant mensuel brut total de l'ensemble des avantages principaux de droit dérivé
mtot5	num.	Montant mensuel brut total de l'ensemble des bonifications ou majorations pour nombre d'enfants
mtot4	num.	Montant mensuel brut total de l'ensemble des majorations pour tierce personne
mtot6	num.	Montant mensuel brut total de l'ensemble des allocations supplémentaire du FSV ou FSI
mtot7	num.	Montant mensuel brut total de l'ensemble des majorations au titre de l'article L.814-2
mtot8	num.	Montant mensuel brut total de l'ensemble des allocations ASPA ou ASI
mv	logical	Indicatrice de perception du minimum vieillesse
cc1	char.	Régime principal de liquidation
cc2	char.	Deuxième régime de base le plus important (par durée validée dans le régime)
ageliq1	num.	Âge d'entrée en jouissance du droit propre de retraite dans le régime de base principal
entddaamin	integer	Année d'entrée en jouissance minimale d'un droit direct
entdraamin	integer	Année d'entrée en jouissance minimale d'un droit dérivé
ageatteintep1	num.	Âge d'atteinte du taux plein dans le régime de base principal
CondLiq1	factor	Condition de liquidation dans le régime de base principal
ind_incapmax	char.	Indicatrice de liquidation au titre de l'invalidité, de l'inaptitude ou du handicap dans au moins un régime de base
carcomp_prorat	num.	Indicatrice permettant de repérer les individus ayant effectué une carrière complète et pour lesquels l'EIR 2016 a pu capter la totalité ou presque des composantes monétaires de la pension de droit direct
sumcoefprorat	num.	Somme des coefficients de proratisation dans les régimes de base
CS1	char.	Catégorie socioprofessionnelle de l'individu EDP durant l'année à 1 chiffre
CS2	char.	Catégorie socioprofessionnelle de l'individu EDP durant l'année à 2 chiffres
CS1bis	char.	Catégorie socio-professionnelle à 1 position, avec distinction qualifié/non-qualifié pour les employés et ouvriers
PCS4	char.	Catégorie socioprofessionnelle de l'individu EDP durant l'année à 4 chiffres (depuis 1993)
cc1_reg	char.	Régime de retraite de base principal en cours de carrière (regroupé)
nb_enfants	char.	Nombre d'enfants calculé d'après les informations administratives des caisses de retraite
agedebcotis	char.	Âge de première cotisation d'un trimestre pour la retraite
agefincotis	char.	Âge de dernière cotisation d'un trimestre pour la retraite

Variables issues de l'enrichissement de l'EIR avec certaines données fiscales

nom_variable	type	descriptif
zimpof	num.	Montant d'impôt payé
revimp	num.	Revenu imposable
depindiv_etab_fisc	num.	Dépenses d'hébergement en établissement déclarées par l'individu
depfoyer_etab_fisc	num.	Dépenses d'hébergement en établissement déclarées par le foyer fiscal
inval80_fisc	num.	Indicatrice d'invalidité à 80 % donnant droit à une demi-part fiscale
nbfoy_inval80_fisc	num.	Nombre de personnes déclarant une invalidité à 80 % dans le foyer
depfoy_saldom_fisc	num.	Dépenses du foyer pour l'emploi d'un salarié à domicile
depfoy_equip_paph	num.	Dépenses d'équipements spécialement conçus pour les personnes âgées ou handicapées
rbg	num.	Revenu brut global déclaré au fisc
absent_th	num.	Indicatrice d'absence du fichier de la taxe d'habitation
statut_matri_fisc	char.	Statut matrimonial pour le fisc
persfip	char.	Place de la personne dans la déclaration fiscale

Variables issues de l'EDP

Les variables issues de l'échantillon démographique permanent proviennent toutes des données fiscales appariées à cet échantillon, elle-même récupérées dans les bases FIDELI et FILOSOFI de l'Insee.

Elles sont toutes relatives à l'année 2016 (déclaration à l'impôt sur le revenu réalisée en 2017).

nom_variable	type	descriptif
ID_DIFF_2020	char.	Identifiant (n° d'ordre) de l'individu dans l'EDP 2020
CIDECI	char.	Situation conjugale
COM_SCAT1	char.	Première sous-catégorie de communauté la plus importante à l'adresse
COM_SCAT_red_2016	char.	Première catégorie de commune à l'adresse en 2016 (redressée)
COM_SCAT_red_2017	char.	Première catégorie de commune à l'adresse en 2017 (redressée)
type_revenu_edp	fac-	Type de revenu perçu d'après l'EDP
YSALI	num.	Salaires perçus
YCHOI	num.	Allocation chômage
YRSTI	num.	Retraites et pensions
YALRI	num.	Pensions alimentaires perçues
YNSAL	num.	Revenus non-salariés
ZOXYZD	char.	Code du dernier événement
ZSALF	num.	Somme des salaires des membres des foyers fiscaux de l'individu
ZCHOF	num.	Somme des allocations chômage des membres des foyers fiscaux de l'individu
ZRSTF	num.	Somme des retraites des membres des foyers fiscaux de l'individu
ZALRF	num.	Somme des pensions alimentaires des membres des foyers fiscaux de l'individu
ZNSALF	num.	Somme des revenus non-salariés des membres des foyers fiscaux de l'individu
ZSALM	num.	Salaire au sens strict du ménage
ZCHOM	num.	Allocation de chômage et de préretraite du ménage

ZRETM	num.	Retraites et pensions du ménage
ZRSTM	num.	Retraites au sens strict du ménage
ZALRM	num.	Pensions alimentaires reçues du ménage
ZETRM	num.	Revenus de l'étranger du ménage
ZNSAL	num.	Revenus non-salariés du ménage
REV_PRINC	char.	Type de revenu principal
REVDISPM	num.	Revenu disponible monétaire du ménage
REVINIM	num.	Revenu initial calculé après imputation des revenus financiers
REVDECM	num.	Revenu déclaré calculé après imputation des revenus financiers
M_AEEHM	num.	Allocation d'éducation enfant handicapé du ménage
M_MINVM	num.	Minimum vieillesse du ménage
M_RSAM	num.	RSA total du ménage
M_AAH_CAAHM	num.	Allocation adulte handicapé du ménage et Complément AAH
M_INVAL_MEN	num.	Montant des pensions d'invalidité déclarées par le ménage (recalculé)

PREMIERS RÉSULTATS

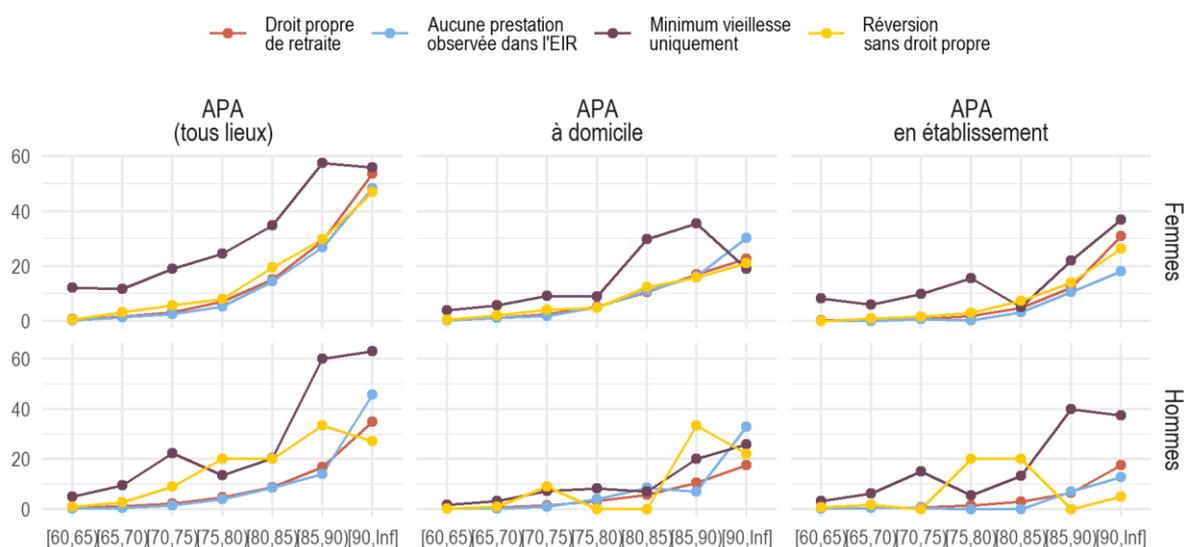
Cette partie présente, à titre purement illustratif, quelques premiers résultats sur les disparités de prévalences des prestations d'autonomie selon les caractéristiques de retraite. Elle ne prétend bien sûr pas constituer une analyse complète.

Prévalences des prestations d'autonomie selon les caractéristiques des retraités

La prévalence de la perte d'autonomie, au sens du recours aux prestations, apparaît plus élevée à tous âges parmi les personnes qui perçoivent uniquement le minimum vieillesse (figure 32). Ce champ inclut des personnes qui n'ont jamais travaillé et n'ont pas de conjoint, et compte une forte proportion de personnes handicapées âgées. Sans être aussi élevées que pour ces personnes, les prévalences de personnes qui perçoivent uniquement une pension de réversion apparaissent également un peu plus hautes que celles des retraités de droit direct.

Ces résultats soulignent le risque de perte d'autonomie significativement plus élevé des non-retraités parmi les personnes âgées, et justifie le choix qui a été réalisé ici de retenir toutes les personnes *échantillonnées dans l'EIR*, et pas uniquement celles observées comme retraitées dans cette source, pour construire la base appariée EIR + RI-APA.

Figure 32 Prévalences des prestations selon le type de droit perçu



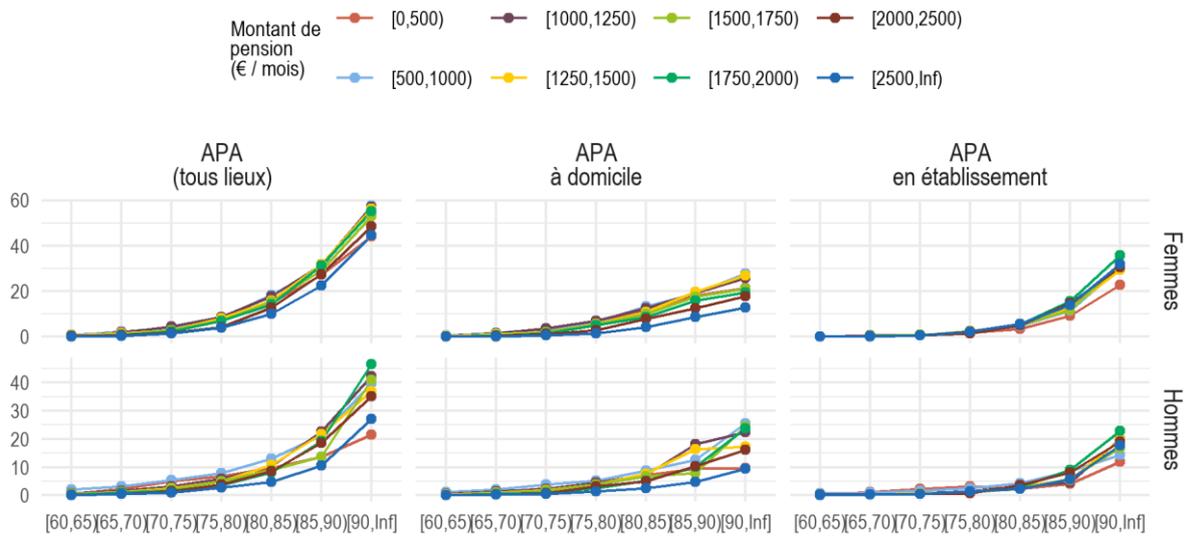
Champ > Personnes percevant une retraite ou le minimum vieillesse dans l'EIR, âgée de plus de 60 ans et résidente en France.

Source > Appariement EIR + RI-APA-ASH, DREES.

Parmi les retraités, le gradient observé pour la mortalité selon le niveau de pension s'observe également pour la perte d'autonomie (au sens du recours à l'APA). Quels que soient le sexe et l'âge, les prévalences sont généralement d'autant plus élevées que le montant de pension est bas (figure 33). Parmi les femmes, la prévalence de l'APA à domicile à partir de 90 ans est par exemple deux fois plus élevée pour celles qui touchent moins de 1 000 euros de retraite par mois que pour celles qui touchent plus de 2 500 euros. Les prévalences apparaissent un peu plus bruitées pour les hommes après 85 ans, mais cela s'explique sans doute par les plus faibles effectifs de retraités survivants parmi les hommes à ces âges.

Soulignons toutefois que c'est surtout pour le recours à l'APA à domicile que les différences selon le niveau de pension sont nettes ; pour l'APA en établissement, les écarts sont beaucoup moins marqués, et les prévalences à âge donné n'apparaissent pas toujours strictement plus basses pour les retraités à pension élevée.

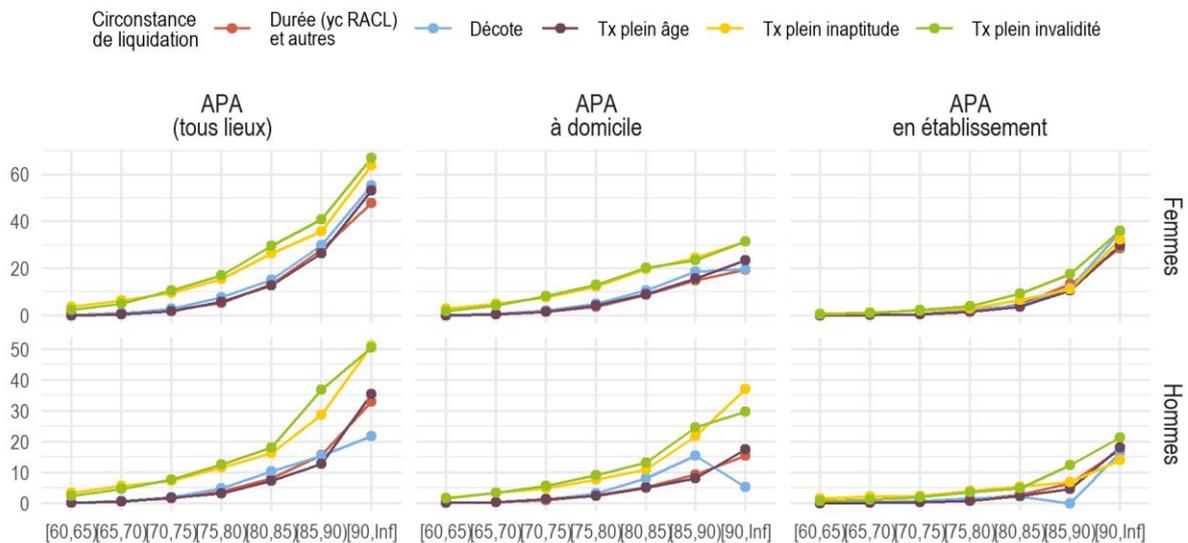
Figure 33 Prévalences des prestations selon le type de droit perçu



Champ > Personnes percevant une retraite ou le minimum vieillesse dans l'EIR, âgée de plus de 60 ans et résidente en France.
Source > Appariement EIR + RI-APA-ASH, DREES.

Les prévalences des prestations d'autonomie varient également fortement selon les circonstances de départ à la retraite. Elles sont significativement plus élevées, à tous les âges, pour les retraités qui ont obtenu le taux plein au titre de l'inaptitude ou de l'ex-invalidité (figure 34). Elles apparaissent également un peu plus élevées pour les retraités partis avec une décote (du moins avant 60), tandis qu'elles sont très proches des retraités ayant obtenu le taux plein au titre de la durée pour ceux ayant eu le taux plein par l'âge (c'est-à-dire les retraités à carrière incomplète, qui ont dû attendre d'avoir atteint l'âge d'annulation de la décote de 65 ou 67 ans pour obtenir le taux plein). Cette comparaison semble indiquer que, parmi les retraités à carrière incomplète, le choix de partir plus tôt avec une décote plutôt que d'attendre quelques années pour l'éviter serait lié avec un état de santé dégradé, qui se traduirait par une plus forte probabilité d'être dépendant aux grands âges.

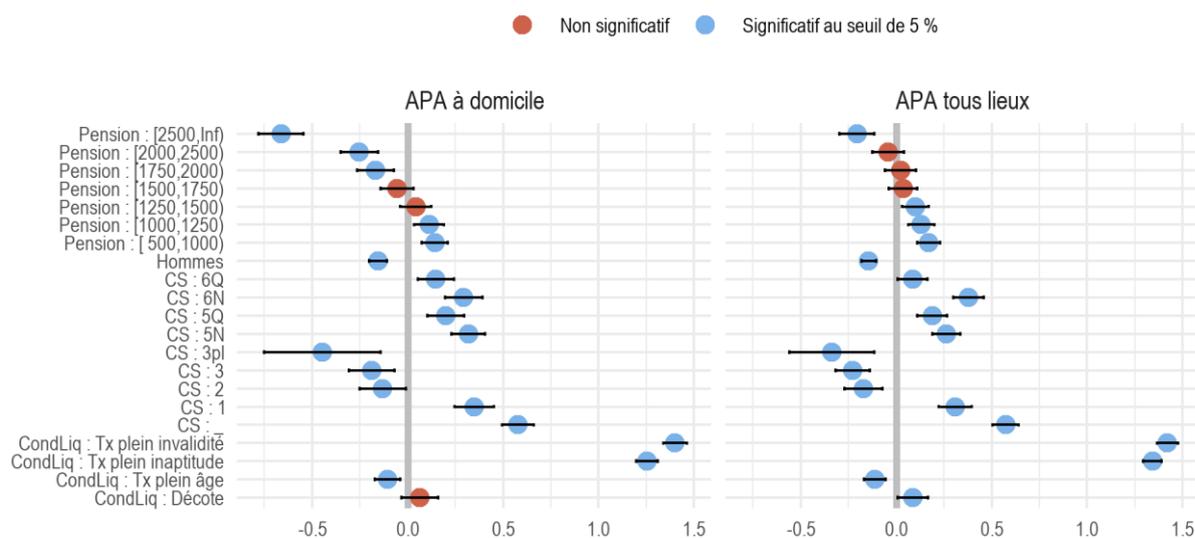
Figure 34 Prévalences des prestations selon le type de droit perçu



Champ > Personnes percevant une retraite ou le minimum vieillesse dans l'EIR, âgée de plus de 60 ans et résidente en France.
Source > Appariement EIR + RI-APA-ASH, DREES.

En complément des ventilations critère par critère présentées dans les graphiques précédent, les différences de probabilité de bénéficiaire de l'APA entre catégories peuvent être illustrées dans une analyse multivariée. Par exemple, on régresse, pour le graphique ci-après, la probabilité d'être dépendant au sens de l'APA à domicile, ou respectivement de l'APA tous lieux confondus, sur diverses variables, de sexe, d'âge et de retraite, sur le champ des bénéficiaires d'une retraite de droit propre. Les probabilités plus élevées d'avoir l'APA pour les personnes à inaptitudes ou ex-invalides, celles à faible niveau de pension ou pour les catégories socioprofessionnelles les moins qualifiées sont confirmées toutes choses égales par ailleurs ; le fait d'avoir liquidé sa retraite avec une décote n'est en revanche pas statistiquement significatif (figure 35).

Figure 35 Coefficients de la régression logit du fait d'être dépendant au sens de l'APA



Note > Pour plus de lisibilité, la constante du modèle et les variables d'âge, d'âge au carré et d'âge au cube ne sont pas représentées. Les modalités de référence : professions intermédiaires (CS = 4) pour la catégorie socio-professionnelle, départ au taux plein au titre de la durée (y compris retraite anticipée pour carrière longue) pour les circonstances de liquidation (CondLiq), moins de 500 euros par mois pour le montant de pension.

Champ > Personnes percevant une retraite de droit propre, âgées de plus de 60 ans et résidentes en France, pour lesquelles la catégorie socioprofessionnelle à 1 chiffre (avec distinction qualifiés [Q] / non-qualifiés [N] pour les employés et les ouvriers) a pu être calculée.

Source > Appariement EIR + RI-APA-ASH, DREES.

Durées de retraite passées en perte d'autonomie

À partir des prévalences de la perte d'autonomie calculées par catégories de retraités, il est aisé d'estimer, par la méthode de Sullivan, des espérances de vie avec et sans perte d'autonomie (encadré). Ce calcul nécessite d'avoir également calculé au préalable des taux de mortalité différenciés selon les mêmes catégories, ce qui ne pose pas de problème puisque le suivi de mortalité prévu dans l'EIR permet d'estimer ces taux.

En pratique, les durées espérées de retraite des diverses catégories d'assurés sont calculées en pondérant les espérances de vie à chaque âge (calculées à partir des coefficients de mortalité par catégorie) par la proportion de personnes partant à la retraite à ces âges. Par exemple, pour une catégorie où 80 % des affiliés partent à la retraite à 60 ans et 20 % à 65 ans, la durée espérée de retraite est égale à 80 % de l'espérance de vie de la catégorie à 60 ans et 20 % de son espérance de vie à 65 ans. Le calcul est réalisé à partir des âges arrondis à l'entier inférieur (notion d'âge exact), mais une correction est appliquée pour tenir compte de l'âge de départ au mois près parmi tous les affiliés partis au même âge (arrondi). Les durées espérées de retraite passées en ou sans perte d'autonomie sont calculées de la même façon à partir des espérances de vie avec et sans perte d'autonomie (elles-mêmes calculées à partir des coefficients de mortalité et des prévalences de la perte d'autonomie par catégorie, en appliquant la méthode de Sullivan).

On fait ici le calcul en utilisant les coefficients de mortalité calculés en moyenne sur la période 2017-2021¹⁴, les âges de départ à la retraite observés en 2016 (en ayant normalisé au préalable le poids de chaque génération de façon à ce que leur taille soit la même à 55 ans), et les prévalences de la perte d'autonomie observés en moyenne fin 2016 et fin 2017. Par convention, les personnes qui ne perçoivent aucun droit direct de retraite sont supposées commencer leur "période de retraite" à 62 ans.

La durée espérée passée en retraite et en tant que bénéficiaire d'une aide à l'autonomie est estimée à 1,4 année pour l'APA à domicile, 1,1 année pour l'APA en établissement, et 2,5 années pour l'APA dans son ensemble. Dans les trois cas, les estimations s'avèrent très proches de celles [publiées par la DREES à partir de l'enquête Aide sociale](#) (1,4 année, 1 année et 2,4 années respectivement d'espérance de vie dans l'APA calculée à 60 ans en 2019). Rappelons que, pour l'APA en établissement et pour l'APA globale, cette proximité tient au fait qu'on a imputé des bénéficiaires de l'APA dans la base appariée, afin de se rapprocher des prévalences observées dans les données de cadrage. La proximité s'observe aussi lorsqu'on ventile les durées espérées par sexe : pour l'APA à domicile, l'espérance de vie passée en retraite et dans l'APA est de 1,9 année et 0,8 année pour les femmes et

¹⁴ Ces coefficients différenciés ne sont calculés qu'à partir de 60 ans mais, pour les personnes parties à la retraite avant, on considère que, à tous les âges inférieurs à 60 ans, l'écart par rapport au taux de mortalité moyen dans l'ensemble de la population est égal à l'écart estimé à 60 ans.

pour les hommes respectivement, contre 1,9 et 0,9 année fin 2019 dans l'Études et Résultats publié fin 2021 ; pour l'APA en établissement, ces espérances sont de 1,5 année et 0,6, contre 1,5 et 0,5 année fin 2019.

Il est possible également de calculer la part de personnes bénéficiant d'une aide à l'autonomie au moins une fois au cours de leur période de retraite (pour la génération fictive vivant avec les conditions du moment, considérée par la méthode de Sullivan), et donc la durée moyenne de perception par les bénéficiaires, en divisant l'espérance de durée de vie passée en tant que bénéficiaire par la proportion de bénéficiaires (l'espérance de durée de vie étant, elle, implicitement calculée en incluant tous les non-bénéficiaires avec une durée de 0 année dans les prestations). Cette proportion est elle-même estimée sous l'hypothèse que les personnes âgées ont une seule période continue de perception des aides à l'autonomie au cours de leur vie : le nombre total de bénéficiaires peut alors être calculé en multipliant les taux de nouveaux bénéficiaires à chaque âge (c'est-à-dire les proportions de personnes dont un droit est ouvert en cours d'année parmi les personnes en vie au 1^{er} janvier de l'année) par le nombre de personnes en vie en début d'année.

Sous ces hypothèses, 58 % des personnes âgées (69 % des femmes et 44 % des hommes) bénéficient de l'APA au cours de leur vie. Les durées moyennes de perception de l'APA (domicile ou établissement) pour les personnes qui en sont bénéficiaires au moins une fois dans leur vie sont estimées à 4,8 années pour les femmes et 3,2 années pour les hommes. Ces valeurs apparaissent un peu plus élevées que celles estimées dans le [Dossier de la DREES consacrée à ces durées](#) sur le champ des personnes décédées en 2011 dans les RI-APA de 2011 (durées moyennes de 3,9 années et 2,9 années respectivement). L'écart pourrait s'expliquer pour partie par la différence de date d'observation : en particulier, le poids des personnes ayant de longues durées d'APA est mécaniquement plus élevé en 2017, soit 15 ans après la création de la prestation, qu'en 2011, soit 9 ans après cette création. Les durées moyennes recalculées sur les personnes décédées en 2017 dans l'appariement EIR + RI-APA de 2017 sont respectivement de 4,6 ans pour les femmes et 3,1 ans pour les hommes, soit des valeurs beaucoup plus proches de celles calculées avec la méthode de Sullivan.

Les écarts entre ces estimations peuvent tenir également en partie à la différence de méthode. La méthode de Sullivan considère en effet les conditions de mortalité du moment ; dans un contexte de baisse régulière de la mortalité, le poids des âges les plus élevés est mécaniquement plus grand, car la probabilité d'avoir atteint ces âges est plus élevée qu'elle ne l'était avec les conditions de mortalité plus anciennes. Cet effet apparaît lorsqu'on compare de façon plus systématique les indicateurs calculés sur une génération fictive vivant avec à tous âges les conditions de mortalité et de prévalence de l'APA du moment (méthode de Sullivan) avec les mêmes indicateurs calculés en moyenne sur le champ des personnes de 60 ans et plus décédées en 2017. Les durées moyennes passées dans l'APA parmi l'ensemble des retraités (y compris ceux qui ne sont jamais bénéficiaires, et ont donc une durée nulle), tout comme les durées moyennes de retraite, sont un peu plus courtes sur le champ des personnes décédées en 2017 qu'elles ne le sont avec la méthode de Sullivan, car dans ce dernier cas les individus ont une probabilité plus élevée de survivre jusqu'aux grands âges et d'atteindre les tranches d'âge où le bénéfice de l'APA est le plus fréquent. Pour la même raison, les proportions de personnes ayant été bénéficiaires de l'APA au moins une fois au cours de leur période de retraite apparaissent un peu plus élevées avec la méthode de Sullivan (figure 36).

Figure 36 Calcul de divers indicateurs selon la méthode utilisée : méthode de Sullivan vs. calcul sur le flux de décédés en 2017

indicateur	type	Ensemble	Femmes	Hommes
Espérance de vie à 60 ans	Ensemble	25.6 / 24.3 ans	27.7 / 26.6 ans	23.4 / 21.5 ans
Durée espérée de retraite	Ensemble	23.8 / 23.1 ans	25.5 / 24.5 ans	21.9 / 21.4 ans
Durée espérée dans l'APA parmi l'ensemble des retraités	Tous lieux, dont :	2.5 / 2.1 ans	3.3 / 2.9 ans	1.4 / 1.1 ans
	Domicile	1.4 / 1.2 ans	1.9 / 1.7 ans	0.8 / 0.7 ans
	Établissement	1.1 / 0.9 ans	1.5 / 1.2 ans	0.6 / 0.4 ans
Proportion de bénéficiaires de l'APA au moins une fois pendant la retraite	Tous lieux, dont :	57.6 / 50.9 %	69.4 / 63.2 %	44.4 / 36.6 %
	Domicile	38.4 / 35.3 %	46.9 / 43.5 %	28.8 / 25.8 %
	Établissement	35.9 / 30.4 %	45.4 / 39.4 %	24.7 / 19.9 %
Durée espérée dans l'APA parmi les bénéficiaires	Tous lieux, dont :	4.3 / 4.1 ans	4.8 / 4.6 ans	3.2 / 3.1 ans
	Domicile	3.7 / 3.4 ans	4 / 3.8 ans	2.9 / 2.7 ans
	Établissement	3 / 2.8 ans	3.2 / 3.1 ans	2.3 / 2.1 ans
Part de la durée dans l'APA dans la durée espérée de retraite	Tous lieux, dont :	10.4 / 8.6 %	13.1 / 11.6 %	6.5 / 5.1 %
	Domicile	5.9 / 5.2 %	7.4 / 6.8 %	3.9 / 3.3 %
	Établissement	4.5 / 3.5 %	5.7 / 4.9 %	2.6 / 1.9 %
Âge moyen à l'entrée dans l'APA	Tous lieux, dont :	84.8 / 84.2 ans	85.3 / 85 ans	84.1 / 82.7 ans

	Domicile	83.7 / 83 ans	83.9 / 83.6 ans	83.1 / 81.8 ans
	Établissement	87.5 / 87.1 ans	88.2 / 87.8 ans	86.1 / 85.4 ans

Lecture > La durée espérée de vie dans l'APA pour les femmes est de 3,3 années lorsqu'on la calcule avec la méthode de Sullivan (c'est-à-dire pour une génération fictive ayant à tous âges les conditions de mortalité et de prévalence de l'APA de 2017), mais de 2,9 années lorsqu'on la calcule en moyenne sur le champ des personnes de 60 ans ou plus décédées en 2017.

Source > Appariement EIR + RI-APA-ASH, DREES.

Encadré Calcul des divers indicateurs avec la méthode de Sullivan

La méthode dite « de Sullivan » est une méthode très simple utilisée pour calculer les espérances de vie « en santé » ou « sans incapacité ». Elle combine des données de mortalité issues de tables de mortalité du moment avec des prévalences d'un état de santé ou d'incapacité à une date d'observation donnée. On mobilise ici cette méthode, en la généralisant, pour calculer les divers indicateurs relatifs à la durée, la probabilité d'occurrence et l'âge moyen à l'entrée de la période de bénéfice de l'APA au sein de la période de retraite. Les prévalences mobilisées pour le calcul sont donc non pas celles des divers états de santé, mais celles d'être à la retraite et de bénéficier de l'APA à chaque âge.

L'approche consiste à raisonner sur une cohorte fictive, comptant par exemple 100 000 personnes en vie à un âge donné (souvent 60 ans lorsqu'on s'intéresse aux politiques du vieillissement), et à calculer les nombres de personnes de cette génération dans chaque état (retraite, APA, décès...) à chaque âge après celui-ci. Ce calcul s'appuie sur les diverses probabilités associées à ces états (probabilité de décéder à un âge a : $P_a^{D\acute{e}c}$, probabilité d'être retraité à cet âge a : P_a^{Retr} , probabilité de bénéficier de l'APA pour un retraité : P_a^{APA}), estimées lors d'une étape préalable à partir de sources statistiques (l'appariement de l'EIR avec les RI-APA-ASH dans le cadre de ce document). Le nombre de personnes encore en vie à un âge A correspond par exemple à $100000 \cdot \prod_{a=60}^A (1 - P_a^{D\acute{e}c})$, le nombre de personne décédant à cet âge à $100000 \cdot \prod_{a=60}^{A-1} (1 - P_a^{D\acute{e}c}) \cdot P_A^{D\acute{e}c}$, le nombre de retraités à $100000 \cdot \prod_{a=60}^A (1 - P_a^{D\acute{e}c}) \cdot P_a^{Retr}$, et, enfin, le nombre de retraités bénéficiaires de l'APA à $100000 \cdot \prod_{a=60}^A (1 - P_a^{D\acute{e}c}) \cdot P_a^{Retr} \cdot P_a^{APA}$ (les formules de calcul sont ici simplifiées par souci de lisibilité : elles sont en réalité un peu plus compliquées pour tenir compte du fait que les décès à l'âge A peuvent avoir lieu entre la date anniversaire de l'âge A et la veille de l'anniversaire de l'âge $A+1$). Pour la suite des calculs, on notera $P_A^{Surv} = \prod_{a=60}^A (1 - P_a^{D\acute{e}c})$ la proportion de personnes encore en vie (survivantes) à l'âge A .

À partir des effectifs précédents, les espérances de vie peuvent être calculées en rapportant le nombre d'années de vie dans chaque état sur l'ensemble des âges postérieurs à l'âge initial choisi (60 ans) au nombre de personnes de la génération fictive à cet âge initial. L'espérance de vie à 60 ans correspond ainsi à $\frac{100000 \cdot \sum_{A=60}^{AgeMax} P_A^{Surv}}{100000}$,

la durée espérée de retraite à $\frac{100000 \cdot \sum_{A=60}^{AgeMax} P_A^{Surv} \cdot P_A^{Retr}}{100000}$ et la durée « espérée » de retraite passée dans l'APA à

$\frac{100000 \cdot \sum_{A=60}^{AgeMax} P_A^{Surv} \cdot P_A^{Retr} \cdot P_A^{APA}}{100000}$. Le rapport entre ces deux dernières espérances de vie $\frac{\sum_{A=60}^{AgeMax} P_A^{Surv} \cdot P_A^{Retr} \cdot P_A^{APA}}{\sum_{A=60}^{AgeMax} P_A^{Surv} \cdot P_A^{Retr}}$ corres-

pond par ailleurs à la part de la durée d'APA (en espérance mathématique) dans la durée espérée de retraite.

Les autres indicateurs présentés dans ce document, qui ne relèvent pas de la méthode de Sullivan à proprement parler mais qui s'inspirent de sa logique, s'appuient sur des calculs analogues. Par exemple, si on note $P_a^{NouvRetr}$ la probabilité d'être nouveau retraité à l'âge a et $P_a^{NouvAPA}$ la probabilité d'être nouveau bénéficiaire de l'APA (en étant retraité), c'est-à-dire les proportions parmi les personnes en vie à cet âge qui sont bénéficiaires alors qu'elles ne l'étaient pas un an avant (à l'âge $a-1$), et si l'on suppose en outre que les personnes qui deviennent retraitées ou bénéficiaires de l'APA le restent jusqu'à leur décès (ce qui implique qu'une personne ne peut être qu'une seule fois nouvelle retraitée ou nouvelle bénéficiaire de l'APA au cours de son existence), alors les effectifs de personnes qui liquident une retraite et de personnes retraitées qui bénéficient de l'APA au cours de leur vie sont égales respectivement à $100000 \cdot \sum_{A=60}^{AgeMax} P_A^{Surv} \cdot P_A^{NouvRetr}$ et à $100000 \cdot \sum_{A=60}^{AgeMax} P_A^{Surv} \cdot P_A^{Retr} \cdot P_A^{NouvAPA}$. Le ratio entre ces effectifs et le nombre de personnes initiales dans la génération fictive (soit 100 000 personnes) correspond alors à la proportion de retraités et de bénéficiaires retraités de l'APA dans la génération. La part de bénéficiaires de l'APA parmi les seuls retraités est quant à elle de $\frac{\sum_{A=60}^{AgeMax} P_A^{Surv} \cdot P_A^{Retr} \cdot P_A^{NouvAPA}}{\sum_{A=60}^{AgeMax} P_A^{Surv} \cdot P_A^{NouvRetr}}$.

Ces proportions permettent notamment de calculer les espérances de vie dans chaque état pour les seuls bénéficiaires. Les espérances de vie à la retraite et dans l'APA calculées avec les formules présentées ci-avant s'interprètent en effet comme des durées espérées dans l'ensemble de la population : il s'agit donc de moyennes qui tiennent compte des personnes qui décèdent avant de partir à la retraite ou de bénéficier de l'APA, et qui comptabilisent donc les durées de retraite ou d'APA par définition égales à 0 de ces personnes. En divisant les durées espérées par les proportions, au sein de la population, de personnes qui liquident effectivement une retraite ou bénéficient de l'APA, on peut toutefois calculer les durées espérées pour les seuls bénéficiaires. Elle

valent respectivement $\frac{\sum_{A=60}^{AgeMax} p_A^{Surv} \cdot p_A^{Retr}}{\sum_{A=60}^{AgeMax} p_A^{Surv} \cdot p_A^{NouvRetr}}$ pour la durée espérée de retraite des retraités, et $\frac{\sum_{A=60}^{AgeMax} p_A^{Surv} \cdot p_A^{Retr} \cdot p_A^{APA}}{\sum_{A=60}^{AgeMax} p_A^{Surv} \cdot p_A^{Retr} \cdot p_A^{NouvAPA}}$ pour la durée espérée d'APA par ses bénéficiaires.

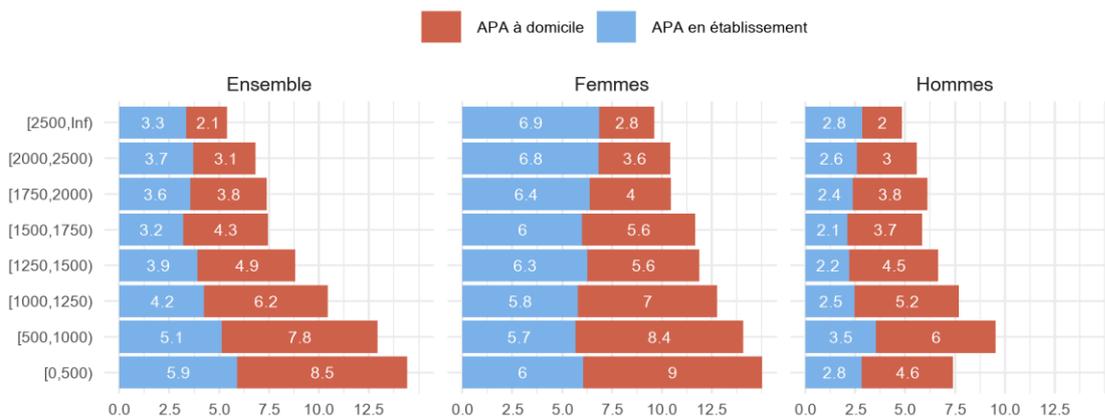
L'âge moyen d'entrée dans l'APA est enfin calculé comme la moyenne de tous les âges à partir de 60 ans, pondérés par le nombre de nouveaux bénéficiaires de l'APA à chaque âge, à savoir : $\frac{\sum_{A=60}^{AgeMax} A \cdot [p_A^{Surv} \cdot p_A^{Retr} \cdot p_A^{NouvAPA}]}{\sum_{A=60}^{AgeMax} p_A^{Surv} \cdot p_A^{Retr} \cdot p_A^{NouvAPA}}$.

Signalons que, pour le présent document, ce n'est pas exactement le calcul selon la méthode de Sullivan, décrite dans cet encadré, qui a été mobilisé pour les durées espérées de retraite, mais un calcul simplifié très proche. En pratique, ces durées ont été calculées en pondérant les espérances de vie à chaque âge par la proportion de personnes partant à la retraite à ces âges. Par exemple, pour une catégorie où 80 % des affiliés partent à la retraite à 60 ans et 20 % à 65 ans, la durée espérée de retraite est égale à 80 % de l'espérance de vie de la catégorie à 60 ans et 20 % de son espérance de vie à 65 ans. Le calcul est réalisé à partir des âges arrondis à l'entier inférieur, mais une correction est appliquée pour tenir compte de l'âge de départ au mois près parmi tous les affiliés partis au même âge. Les durées espérées de retraite passées en ou sans perte d'autonomie sont calculées de la même façon à partir des espérances de vie avec et sans perte d'autonomie.

On présente également ci-après, toujours à titre d'illustration, les parts de la durée espérée de retraite passée en perte d'autonomie ventilées selon le niveau de pension de droit direct (par tranches de montants). Les retraités à plus basses pensions passent près de 15 % de leur durée de retraite en étant en perte d'autonomie, au sens du recours aux aides publiques (APA ou réduction d'impôt en établissement). Cette proportion décroît progressivement avec le montant de retraite, et la part est d'environ 5 % pour les retraités touchant plus de 2 500 euros de retraite par mois (figure 37). Cette relation entre durée en perte d'autonomie et montant de retraite tient pour partie à un effet de composition par sexe, les femmes ayant en moyenne des pensions plus faibles mais des durées en perte d'autonomie plus élevées. La relation s'observe également parmi les femmes et parmi les hommes, mais les écarts sont plus réduits. C'est surtout pour l'APA à domicile que la durée de perception (en % de la durée de retraite globale) diminue avec le niveau de pension ; les personnes à plus faible retraite (tous sexes confondus) passent une part de leur durée de retraite plus de quatre fois plus longue en tant que bénéficiaire à domicile que ceux aux pensions les plus élevées. Les différences sont en revanche moins marquées en établissement ; elles apparaissent essentiellement lorsqu'on considère la population tous sexes confondus, du fait de l'effet de composition genré, mais sont beaucoup plus faibles au sens de chaque sexe pris séparément.

Il est important de rappeler que la perte d'autonomie est appréciée ici par le biais des aides publiques. Les disparités peuvent donc s'expliquer par des écarts de niveaux d'incapacité entre catégories, mais aussi par des différences de comportement de recours aux aides. Le ticket modérateur de l'APA à domicile (c'est-à-dire la part du montant de l'aide qui est laissée à la charge du bénéficiaire) varie par exemple fortement selon les revenus des personnes : les plus aisés ne perçoivent en pratique que 10 % du montant complet du plan d'aide, le reste étant laissé à leur charge. Cette modulation pourrait expliquer un recours moindre des plus aisés à l'APA, ce qui pourrait à son tour expliquer une partie de la diminution avec le niveau de pension de la durée de retraite moyenne passée en tant que bénéficiaire de l'APA à domicile.

Figure 37 Part de la durée espérée de retraite passée en tant que bénéficiaire d'aides à l'autonomie, en %



Note > Les calculs sont réalisés à partir des âges de départ à la retraite observés en 2016 (en normalisant la taille de chaque génération), des prévalences de l'APA fin 2017, et des conditions de mortalité moyennes sur la période 2017-2021.

Champ > Personnes percevant une retraite de droit propre de retraite, résidentes en France.

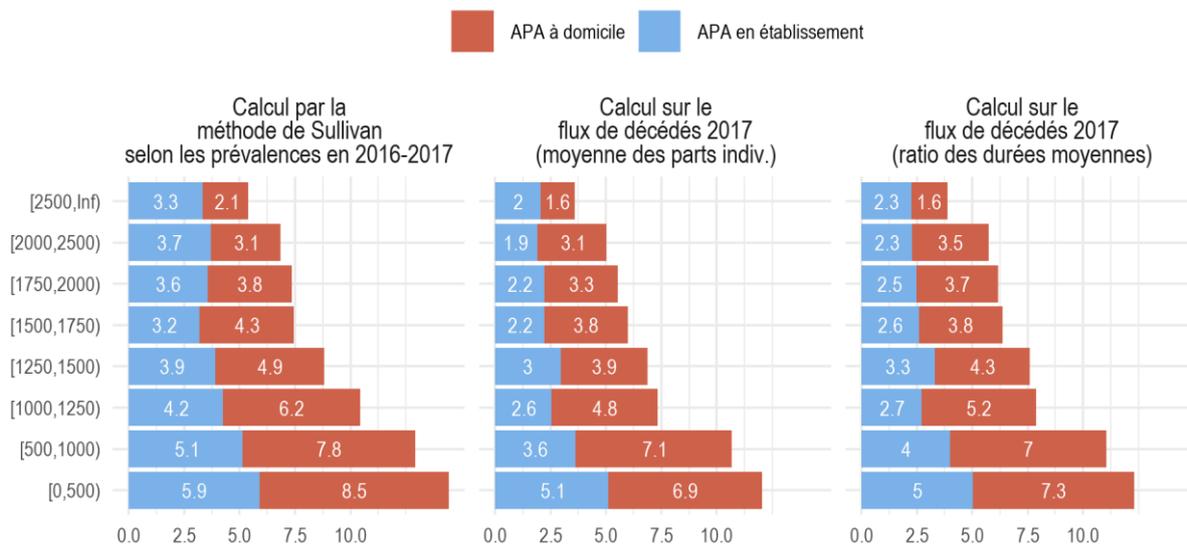
Source > Appariement EIR + RI-APA-ASH, DREES.

La base appariée EIR + RI-APA-ASH permet aussi de calculer, à titre d'alternative, les durées *effectivement* passées en tant que bénéficiaire des prestations, et leur part dans la durée de retraite totale, pour le flux des personnes

qui décèdent en 2017. Le flux de décédés d'une année donnée est cependant sensible aux effets de composition démographique. En particulier, le *baby-boom* implique un poids plus élevé des générations nées à partir de 1946 dans la population, donc une surpondération parmi les décès des personnes ayant eu les durées de retraite les plus courtes et n'étant pas encore entrées en perte d'autonomie. Pour corriger cet effet de composition, les pondérations ont été normalisées de façon à ce que chaque génération soit ramenée à une taille identique à 55 ans.

Le calcul des durées moyennes passées en tant que bénéficiaire des prestations (en % de la durée de retraite) sur le flux de décédés plutôt que par un calcul d'espérance de vie conduit à des résultats davantage bruités, du fait de l'échantillon plus petit. Les deux méthodes conduisent aux mêmes résultats qualitatifs concernant les différences selon le montant de pension entre les durées (relatives) passées dans l'APA, à savoir des durées d'autant plus courtes que la pension est élevée, l'écart étant porté pour l'essentiel par l'APA à domicile (figure 38). Les durées relatives paraissent plus courtes lorsqu'elles sont calculées sur le flux de décédés, mais l'écart tient *a priori* à la différence mécanique induite par la méthode, déjà évoqué précédemment, qui fait que toutes les durées dans l'absolu (durée moyenne dans l'APA et durée moyenne de retraite) sont un peu plus courtes lorsqu'elles sont calculées sur un flux de décédés, dans le cadre d'une mortalité diminuant tendanciellement au cours du temps. En effet, la composition de ce flux dépend aussi des conditions passées de mortalité (malgré la normalisation du poids de chaque génération), contrairement à l'espérance de vie "instantanée", qui ne dépend que des conditions de mortalité du moment. En particulier, si la mortalité diminue à chaque âge, le poids des personnes les plus âgées (donc des durées de retraite et d'APA les plus longues) est plus faible dans le flux de décédés, puisque la probabilité d'avoir atteint ces âges élevés est elle-même plus faible avec les conditions de mortalité du passé qu'elle ne l'est avec les conditions de mortalité du moment.

Figure 38 Part de la durée de retraite passée en tant que bénéficiaire d'aides à l'autonomie, en %



Note > Les durées présentées correspondent aux durées espérées pour le calcul d'après la méthode de Sullivan (espérance mathématique compte tenu des probabilités de décéder de bénéficier de l'APA à chaque âge) et aux durées effectivement réalisées pour le calcul sur le flux des personnes décédées en 2017.

Champ > Personnes percevant une retraite de droit propre de retraite, résidentes en France.

Source > Appariement EIR + RI-APA-ASH, DREES.

■ PRISE EN MAIN DE LA BASE : PRINCIPAUX POINTS D'ATTENTION

Cette dernière partie résume les principaux points à avoir à l'esprit pour une bonne utilisation de la base appariée EIR - RI-APA-ASH. Elle peut être lue en premier, en guise d'introduction à l'utilisation de cette base.

Accès aux données

Les données sont accessibles uniquement via le [centre d'accès sécurisé aux données \(CASD\)](#). La demande d'accès doit être adressée au comité du secret statistique au moyen du [portail d'accès dédié](#). La source est intitulée "Échantillon inter-régimes de retraités apparié avec les données des Remontées individuelles sur les bénéficiaires de l'allocation personnalisée d'autonomie (APA) et de l'aide sociale à l'hébergement (ASH) (RI-APA-ASH) (EIR apparié RI-APA-ASH)" dans le [référentiel des sources](#).

La base appariée EIR - RI-APA-ASH **ne contient qu'une partie des variables d'origine disponibles dans ses données-sources**, mais il est possible de récupérer d'autres informations en la ré-appariant à ces données-sources grâce aux identifiants pseudonymisés. Pour des travaux nécessitant des informations qui n'ont pas été retenues dans la base appariée (cf. partie "Liste des variables"), il est donc **nécessaire de demander également l'accès aux données-sources** par le portail d'accès dédié :

- *Échantillon inter-régimes de retraités (EIR)* (pour davantage d'informations sur les retraites et pour le détail régime par régime),
- *Échantillon inter-régimes de cotisants (EIC)* (pour davantage d'informations sur les droits acquis en cours de carrière – disponibles pour certaines générations seulement),
- *Échantillon inter-régimes de retraités apparié avec les données issues du panel tous actifs (Insee-Drees) (EIR apparié panel tous actifs)* (pour les informations sur le détail des emplois exercés après 50 ans : salaires, conditions d'emploi, secteur d'activité, etc.)
- *Échantillon inter-régimes de retraités apparié avec les données issues du panel EDP (Insee-Drees) (EIR apparié EDP)* (pour des informations complémentaires issues de l'EDP : diplôme, années des mariages...)
- *Remontées individuelles sur les bénéficiaires de l'allocation personnalisée d'autonomie (APA) et de l'aide sociale à l'hébergement (ASH) (RI-APA-ASH)* (pour des informations complémentaires sur les prestations d'APA et d'ASH)

Informations disponibles

Les informations disponibles sont les suivantes :

- *Informations de l'EIR* : caractéristiques sociodémographiques (sexe, année et lieu de naissance, département de résidence), montants mensuels de retraite (droit direct, réversion, MTP...), régime de retraite principal et secondaire en cours de carrière, âge et circonstances de départ à la retraite, durée de carrière, catégorie socioprofessionnelle. Un descriptif détaillé de ces variables est disponible dans le [guide utilisateur de l'EIR de 2016](#), téléchargeable sur la [page internet de présentation de l'EIR](#).
- *Informations des RI-APA-ASH* : dates d'ouverture et de fermeture éventuelle des droits aux prestations, dates des première et dernière évaluation du GIR, GIR et axes ayant servi à coder le GIR, montant d'aide et ventilation du montant entre la part relevant du conseil départemental et le ticket modérateur restant à la charge du bénéficiaire, nombre d'heures d'aide humaine, ressources prises en compte pour le calcul du ticket modérateur. La [présentation des RI-APA-ASH](#) et le [guide utilisateur des données](#), avec le détail des diverses variables reprises dans la base appariée, sont également disponibles sur le site de la DREES.
- *Informations issues des données fiscales appariées à l'EIR 2016* : dépenses d'hébergement en établissement de la personne ou du foyer dans son ensemble, invalidité à 80 % donnant droit à une demi-part fiscale et nombre de personnes invalides à 80 % dans le foyer, dépenses du foyer pour l'emploi d'un salarié à domicile, dépenses d'équipements spécialement conçus pour les personnes âgées ou handicapées. On se reportera à la [sous-section correspondante ci-avant](#) pour plus d'information.
- *Informations issues de l'EDP apparié à l'EIR* : ressources (retraite, salaires, allocations chômage, etc.) de l'individu, de l'ensemble des personnes de son ou ses foyers fiscaux, et de l'ensemble des personnes de son ménage, revenu disponible et minima sociaux perçus par le ménage, type de communauté (pour les

personnes vivant en communauté). Le détail des variables de l'EDP est présenté sur le [site du CASD](#). Une version est également disponible sur la [page internet de l'Ined consacrée à l'utilisation de l'EDP](#).

Pour enrichir la base appariée avec des informations complémentaires issues des bases sources, il faut ré-apparier en utilisant les identifiants pseudonymisés suivant :

- *NOIND* pour l'EIR,
- *ID_APAD*, *ID_APAE* et *ID_AHS* pour les RI-APA-ASH. Ces trois identifiants correspondent à la variable *ID* dans les RI-APA-ASH. La chaîne de caractère "*clone*" ajoutée au début de la variable pour les individus imputés par *cold deck* (voir ci-après) doit être préalablement supprimée.
- *ID_DIFF_2020* pour l'EDP (correspondant à l'identifiant *ID_DIFF* dans la base 2020 de l'EDP)

Champ et pondérations à utiliser

Le champ de la base correspond à **l'ensemble des personnes en vie et âgées de 59 ans ou plus au 31 décembre 2016, résidentes en France et dont le jour de naissance fait partie des jours échantillonnés dans l'EIR**. Le champ n'est pas restreint aux bénéficiaires de prestations : la base inclut donc aussi des personnes qui ne sont pas retraitées et qui ne bénéficient pas de l'APA ou de l'ASH. Les pondérations disponibles dans la base visent à la rendre représentative de l'ensemble des personnes de 60 ans et plus résidentes en France (l'inclusion dans la base des personnes ayant 59 ans fin 2016 se justifiant uniquement par le fait de couvrir toutes les personnes de 60 ans ou plus fin 2017 : cf. ci-après).

Plusieurs variables de pondérations sont disponibles, correspondant à divers sous-champs pertinents pour l'analyse. **Il est donc indispensable, préalablement à tout traitement, de définir le sous-champ souhaité et de choisir la pondération adéquate en conséquence**. Les choix reposent sur les trois questions suivantes :

- *L'analyse cherche-t-elle à être représentative des seules générations et âges observés dans l'EIR (pondération "pond_recalpop_gen") ou bien de l'ensemble de la population des 60 ans et plus ? (pondérations "pond_recalpop_agequin")* : L'EIR ne couvre en effet pas toutes les générations ayant plus de 60 ans. L'utilisation de la pondération par génération (*pond_recalpop_gen*) conduit donc à une population totale inférieure à l'ensemble des personnes de 60 ans ou plus (et, par conséquent, des nombres de retraités ou de bénéficiaires de l'APA eux-aussi plus bas que les effectifs totaux d'après les données de cadrage). Pour être représentatif de l'ensemble de la population et pour que les totaux soient bien cohérents avec les données de cadrage, il est nécessaire d'utiliser les pondérations par âge quinquennal (*pond_recalpop_agequin*). Dans ce cas, seule la variable d'âge quinquennal (*agequin*) reste pertinente ; les variables d'année de naissance (*ai*) et d'âge fin (*age*) sont en revanche faussées, puisque certaines générations ont vu leur poids augmenté afin de "représenter" les générations adjacentes, absentes de l'EIR.
- *La population de référence est-elle celle des 60 ans et plus au 31 décembre 2016 ou au 31 décembre 2017 ? (pondérations "pond_recalpop_agequin2016" et "pond_recalpop_agequin2017" respectivement)* : Dans le premier cas, la génération née en 1957 (qui a moins de 60 ans fin 2016) a une pondération nulle. Dans le second, la pondération est mise à 0 pour les personnes décédées en cours d'année 2017. En outre, le découpage par âge quinquennal n'est pas le même fin 2016 et fin 2017, si bien que l'extrapolation des générations observées de façon à être représentatives des générations adjacentes absentes de l'EIR varie selon la date de référence retenues. Les pondérations sont recalées sur les populations par âge diffusées par l'Insee aux premiers janvier 2017 et 2018 respectivement (supposées égales aux populations aux 31 décembre 2016 et 2017). Ces pondérations aux deux dates ne sont fournies que pour les jeux de pondérations représentatives de la population totale des 60 ans plus par âge quinquennal. Pour les pondérations représentatives des générations échantillonnées dans l'EIR, une seule pondération est en revanche disponible, calée sur la population par âge fin en vie au 31 décembre 2016.
- *L'analyse retient-elle tous les individus du champ de l'EIR (pondérations "pond_recalpop_...") ou uniquement ceux qui sont en outre dans le champ de l'EDP ? (pondérations "pond_recalpop_edp...")* : Une partie seulement des individus échantillonnés dans l'EIR sont dans le champ de l'échantillon démographique permanent (EDP). La restriction à ce dernier champ diminue le nombre d'observations disponibles, mais elle peut être malgré tout intéressante, car elle permet de disposer d'informations supplémentaires issues de l'EDP, notamment celles sur les ressources autres que la retraite et sur les revenus de l'ensemble des personnes du ménage.

Une fois le champ d'analyse défini, la base doit être filtrée pour ne retenir que les individus pour lesquels la pondération correspondante prend des valeurs non nulles. Par exemple, le filtre *pond_recalpop_edpagequin2017>0* doit être appliqué si on veut travailler sur une base représentative de l'ensemble des personnes de 60 ans et plus en vie fin 2017, présentes dans le champ de l'EDP.

Des variables de pondération intitulées "*pond_corttirage_...*" sont également disponibles dans la base de données. Celles-ci sont fournies à titre d'information, et ne doivent *a priori* pas être utilisées. Elles correspondent aux poids

calculés de façon à corriger le plan de sondage de l'EIR, mais **avant** recalage sur les populations de référence par âge diffusées par l'Insee.

Toutes les pondérations (par génération ou par âge quinquennal) sont calculées de façon à être représentatives de la France entière. **La représentativité n'est en revanche pas assurée à un niveau géographique plus fin (région ou département)**. En particulier, la base contient un repérage du département de résidence (variable *di*), mais cela ne garantit pas pour autant qu'elle soit représentative de la population de chaque département en particulier. Pour les départements les moins peuplés, le taux de sondage de l'EIR est en particulier sans doute trop faible pour permettre une représentativité satisfaisante.

Repérage des populations d'intérêt

Les populations d'intérêt pour les analyses correspondent notamment aux retraités (de droit direct ou dérivé) et/ou aux bénéficiaires de l'APA ou de l'ASH. Ces populations peuvent être repérées grâce aux variables suivantes :

- Les retraités peuvent être repérés par le fait de bénéficier d'un montant de pension positif en décembre 2016 : filtre *mtot1>0* pour les retraités de droit direct, et filtre *mtot2>0* pour les retraités de droit dérivé. Un filtre similaire peut être appliqué pour repérer les bénéficiaires de la majoration tierce personne (MTP), par la condition *mtot4>0*. Par ailleurs, les allocataires du minimum vieillesse observés dans l'EIR peuvent être repérés par la condition *mv==TRUE*. De façon alternative, la variable de synthèse *type_droit_eir* résume les principaux statuts relatifs à la retraite et au minimum vieillesse (cette variable permettant en outre de repérer les bénéficiaires d'une pension d'invalidité fin 2016).
- Les bénéficiaires de l'APA ou de l'ASH peuvent être repérés au moyen des dates d'ouverture et de fermeture des droits à ces prestations (variables *DATE_APAX* et *FERM_APA_x*, où *x = D ou E* pour le domicile et les établissements, en ce qui concerne respectivement l'ouverture et la fermeture des droits à l'APA, et *DATEOUVASH* et *DATE_FERM_ASH* en ce qui concerne l'ASH). À titre d'exemple, le fait d'être bénéficiaire de l'APA à domicile fin 2016 peut être repéré par la condition (*!is.na(DATE_APAD) & DATE_APAD<="2016-12-31" & (is.na(FERM_APA_D) | FERM_APA_D>"2016-12-31")*). Des variables *presta_xxx_fin2016* (où *xxx = apad, apae, ash* ou *apaash*) ont par ailleurs été déjà construites et permettent de repérer les bénéficiaires des diverses prestations au 31 décembre 2016 (par exemple *presta_apad_fin2016==TRUE* pour les personnes ayant un droit ouvert à l'APA à domicile). Comme pour le statut vis-à-vis de la retraite, une variable de synthèse *prestation_fin2016* résume enfin le statut de perception de l'APA et/ou de l'ASH fin 2016. Soulignons que toutes ces variables et filtres permettent de repérer les bénéficiaires de la prestation au titre d'un droit ouvert, champ qui peut être un peu plus large que celui des personnes effectivement payées au titre d'un mois données, sur lequel portent la plupart des statistiques de cadrage relatives à l'APA. Par ailleurs, les bénéficiaires de l'ASH peuvent inclure des personnes qui bénéficient de cette prestation au titre du handicap, et non de l'âge ; la condition supplémentaire *HANDI_ASH != "1"* peut être ajoutée pour se restreindre aux seuls bénéficiaires au titre de l'âge (voire *HANDI_ASH == "0"* pour éviter les cas où des modalités inconnues soient incluses, parmi lesquelles certains bénéficiaires pourraient relever de l'ASH au titre du handicap).

Repérage des informations imputées

Les données-sources de la base appariée EIR - RI-APA-ASH sont des bases de données statistiques qui ont déjà fait l'objet de divers redressements et imputations de la part de la DREES. Des imputations complémentaires ont cependant été réalisées spécifiquement sur la base appariée, en vue des usages spécifiques à cette base.

Pour certains usages, il peut toutefois être nécessaire de repérer ces données imputées, car il est parfois pertinent de travailler sur les seules informations "brutes". Les principales imputations et la façon de les repérer sont les suivantes :

- Une partie des bénéficiaires de l'APA a été imputée pour pallier la couverture incomplète par les RI-APA-ASH de la population des bénéficiaires de l'APA en établissement (elle-même liée au fait que les départements ne disposent généralement pas d'informations individuelles sur les bénéficiaires résident en établissement sous dotation globale). Cette imputation a été réalisée par *cold deck*, c'est-à-dire en cherchant, pour chaque individu de la base appariée qui n'est pas observé comme bénéficiaire de l'APA en établissement dans les RI-APA-ASH mais dont on suppose qu'il est malgré tout bénéficiaire de cette prestation, un "jumeau" dont les caractéristiques d'APA seront recopiées¹⁵. Les individus pour lesquels les caractéristiques d'APA en établissement ont été imputées peuvent être repérés par le fait que l'identifiant

¹⁵ La méthode de *cold deck*, utilisée ici, consiste à choisir les jumeaux dans une base externe (ici les RI-APA-ASH sur le champ des individus qui ne sont pas dans l'EIR) ; à l'inverse, la méthode de *hot deck* consiste à choisir les jumeaux parmi les autres individus de la même base.

pseudonymisé *ID_APAE* commence par la chaîne de caractère “*clone*” (condition *grep*(“*^clone*”,*ID_APAE*)). De même, les individus pour lesquels ce sont des caractéristiques d’APA à domicile qui ont été imputées peuvent être repérés par une condition similaire sur l’identifiant pseudonymisé *ID_APAD* (ces imputations sont nettement plus rares : elles concernent uniquement des périodes d’APA à domicile s’achevant avant le 31 décembre 2016, pour des individus ayant enchaîné une APA à domicile puis APA en établissement). Les imputations de caractéristiques d’APA peuvent également être repérées au moyen des modalités de la variable de synthèse *source_apa*. **Pour des analyses économétriques qu’on voudrait mener sur les seules données brutes et non sur les données imputées, il faut au préalable redresser à valeur manquante toutes les variables relatives à l’APA en établissement (resp. à domicile) pour les individus “clonés” pour l’APA-E (resp. l’APA-D).**

- Les dates d’ouverture ou de fermeture des droits à l’APA ont également été redressées pour un nombre non négligeable d’individus, pour les mettre en cohérence entre elles ou pour imputées des valeurs manquantes. La variable *type_redressement_date* indique, pour chaque individu concerné, le ou les redressements qui ont été réalisés (par exemple, “*FERM_APA_D = DATEDECES*” signale que la valeur de la variable *FERM_APA_D* a été remplacée par la date de décès *DATEDECES*).

DREES MÉTHODES
N° 20 • février 2025

L'appariement
de l'EIR de 2016
et des RI-APA-ASH
de 2017

Directeur de la publication
Fabrice LENGART

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
