



mai 2025

GUIDE D'EXPLOITATION

Échantillon interrégimes de retraités 2020

Pierre Cheloudko, Mélissa M'Piayi, Anam Mohammad

TABLE DES MATIÈRES

■ INTRODUCTION : PRÉSENTATION DE L'EIR	6
Présentation générale	6
Évolution du champ de l'enquête depuis 1988	6
Travaux réalisés avec l'EIR 2020	7
Fondements juridiques	8
Déroulement de la production	9
■ PLAN DE SONDAGE	10
Construction d'un échantillon anonyme via le principe de « double aveugle »	10
Étape 1 : mise en place des règles de sélection	10
Étape 2 : tirage de l'échantillon par l'Insee	10
Étape 3 : envoi de la liste des NIR aux caisses	11
Étape 4 : création des bases de données anonymes	11
Description détaillée du plan de sondage de l'EIR 2020	11
Sélection par génération (année de naissance)	11
Sélection par jour de naissance	11
Évolutions du plan de sondage par rapport à l'EIR 2016	14
De l'échantillon RNIPP à l'échantillon EIR	15
Échantillon RNIPP	15
Échantillon EIR	15
Échantillons EIR 2020 vs. EIR 2016 : comparaison des tailles d'échantillon	17
■ COLLECTE, CONTRÔLES ET REDRESSEMENTS	18
Collecte des données	18
Organismes de retraite participants	18
Variables demandées	18
Contrôles et redressements par caisse	19
Redressements globaux	20
■ PONDÉRATIONS	21
Les pondérations initiales	22
Pour les générations 1915 à 1928, qui représentent aussi les générations adjacentes	22
Pour les générations 1930 à 2000	23
Le calage sur marge	24
La correction de la mortalité différentielle	28
■ POST-TRAITEMENTS	30
Equivalent carrière complète (EQCC)	30
Principe général	30
Calcul de RetrPleine : Le montant de « retraite pleine »	31
Imputation du lieu de résidence	32
Imputation du nombre d'enfant des femmes	32
Principe de l'imputation	32
Résultats de l'imputation	33
■ CRÉATION DE NOUVELLES VARIABLES ET DES BASES FINALES AVANT, INDIV ET DDIR	35
Description des tables	35
Calculs des âges et des tranches d'âges quinquennaux : les mois de naissance inconnus (variable AGEQUIN_AN)	36

Détermination du régime principal d'affiliation (variable CC1)	36
Variable synthétique d'affiliation (variable POPREV).....	36
Variables indicatrices sur la « carrière complète » (variables CARCOMP_XX)	37
Variables de durée et d'exhaustivité des carrières	37
La variable DUREEVALIDEE.....	37
La variable DUREECOT	38
La variable ECARTDUREE	38
La variable EXHAU.....	39
La variable NONEXHAU.....	40
Calcul de la pension nette et du taux de prélèvement sociaux (variables PSOC et TXPSOC).....	40
■ INFORMATIONS COMPLÉMENTAIRES SUR CERTAINES VARIABLES	41
Variables relatives au minimum de pension.....	41
Taux de liquidation de la pension (TAUX).....	41
Régimes général et alignés	41
Régimes de la Fonction publique (SRE, CNRACL, FSPOEIE)	42
Régimes de la CNAVPL	43
Régimes spéciaux (SNCF, RATP, Banque de France, CRPCEN).....	43
Régimes complémentaires	43
Nombre de points pour les droits directs (variables du type NPOINDD)	43
Majoration de pension pour les parents de 3 enfants ou plus	43
Variables relatives au minimum vieillesse.....	44
Retraite progressive (TYPDD=26)	44
ANNEXE 1. LISTE DES CODES CAISSES UTILISÉS – VARIABLES CC ET CCGP	45
ANNEXE 2. LES APPARIEMENTS AVEC L'EIR	48
Appariement avec le Panel tous salariés et l'Échantillon démographique permanent.....	48
Appariement avec les données fiscales.....	48
ANNEXE 3. DESSINS DE FICHIERS DES TABLES DE L'EIR 2020 ET NOMENCLATURE DE CERTAINES VARIABLES	49

PRÉAMBULE

L'**Échantillon Interrégimes de Retraités** (EIR) est une base de données d'une très grande richesse sur les retraités et les retraites du système français ainsi que les bénéficiaires d'une pension d'invalidité. À ce titre, il s'agit d'une source de référence pour les études statistiques du système de retraite français.

Cette base est issue d'un appariement de données administratives provenant de quasiment tous les régimes publics obligatoires français. Le système de retraite français étant très fragmenté, cette quasi-exhaustivité est l'un des points clefs de l'EIR.

Outre la très large couverture, en termes de régime, la richesse de l'EIR tient également dans le détail des informations disponibles ainsi que son très grand nombre d'individus. En effet, les données sont disponibles pour environ 1,4 million de retraités, sur environ 70 générations. Pour chaque assuré, des informations détaillées et anonymes sont disponibles, par type d'avantage (droit direct ou dérivé, majorations familiales, etc.), sur le montant de la pension et les conditions de liquidation (type de départ, date de départ, taux de la pension, etc.).

L'EIR 2020 constitue la neuvième vague de cette opération. Par rapport aux précédentes vagues, l'échantillon a été enrichi par l'ajout de nouvelles générations et de nouvelles variables par rapport à l'EIR 2016, qui lui-même avait été élargi (cf. Évolution du champ de l'enquête depuis 1988, p.6).

Consciente des enjeux sur la diffusion des données produites par le système statistique public, la Drees a souhaité que l'EIR soit diffusé le plus largement possible, tout en respectant le respect de la confidentialité des données à caractère personnel. Pour cet EIR, deux supports de diffusion ont été créés, plus agrégés et anonymes, pour une diffusion plus large :

- **L'EIR « complet »** (fichier pseudonymisé, mais à risque de réidentification non négligeable du fait de la précision des variables) est disponible via le CASD.
- **L'EIR-FPR** (fichier de production et de recherche) contient des données individuelles, avec un détail important de variables. Il n'est accessible qu'aux organismes de recherche (listés à l'annexe 2 de l'avis du comité du secret statistique relatif à l'accès aux fichiers de production recherche) et aux organismes habilités (listés dans l'annexe 3 du même avis du CSS) qui ont une mission de service public après passage par le comité du secret statistique (nécessité d'un avis favorable du CSS et d'une dérogation du service des Archives). L'EIR-FRP est pseudonymisé et, afin de réduire sensiblement le risque de réidentification, contient une information moins détaillée que l'EIR complet.

Le présent guide vise à présenter la constitution des fichiers de l'EIR 2020, les redressements réalisés sur les variables collectées ainsi que la création de variables complémentaires. Il revient également sur l'élaboration des différentes pondérations ainsi que le traitement spécifique des personnes invalides. Enfin, il présente la construction de la version pseudonymisée l'EIR-FRP.

En cas de questions, remarques ou suggestions, les utilisateurs peuvent contacter l'adresse :

drees-bret@sante.gouv.fr

■ INTRODUCTION : PRÉSENTATION DE L'EIR

Cette première partie présente l'historique de l'enquête ainsi que les principes de constitution des fichiers de l'EIR (champ, échantillon, organismes participants et fondements juridiques) et les différentes tables issues du fichier d'étude.

Présentation générale

La pension de retraite versée à un individu peut provenir de plusieurs régimes de retraite différents s'il a changé de régime en cours de carrière (par exemple s'il a été salarié puis a exercé une profession libérale), ou même du simple fait que des droits sont acquis à la fois dans des régimes de base et complémentaires. En outre, une personne peut bénéficier de droits directs et/ou de droits dérivés.

L'Échantillon Interrégimes de Retraités (EIR) a pour but de rapprocher les données des différents régimes français légalement obligatoires. Ce rapprochement, individu par individu, des montants en provenance des différents régimes permet de reconstituer la pension globale de chaque personne, ainsi que les éléments caractéristiques de cette retraite (durées validées, dates de liquidation, etc.).

L'EIR est un panel, les personnes appartenant à l'échantillon d'une vague sont donc re-sélectionnées lors de la vague suivante, hormis les individus décédés entre deux vagues et une partie des personnes appartenant à des générations ponctuellement surreprésentées à la vague précédente. Ceci permet d'observer finement l'évolution des retraites entre deux vagues, en distinguant l'effet de différents facteurs (revalorisations, acquisition de nouveaux droits, arrivée de nouveaux liquidants, disparition des décédés...).

Évolution du champ de l'enquête depuis 1988

L'EIR s'enrichi au fur et à mesure des vagues de collecte dans le but de représenter au mieux la population des retraités. Ainsi, le champ de l'enquête a fortement évolué depuis la première vague de 1988 (cf. **Erreur ! Source du renvoi introuvable.**).

En effet, les opérations menées en 1988 et 1993 concernaient uniquement les retraités de 65 ans ou plus nés en France métropolitaine. L'échantillon de 1997 a été étendu aux retraités âgés de 55 ans ou plus nés en France métropolitaine. L'EIR a ensuite été élargi aux personnes nées dans les départements d'outre-mer (EIR 2001) puis aux personnes nées à l'étranger ou dans les territoires d'outre-mer (EIR 2004). La vague 2008, est enrichie des retraités âgés de 35 ans ou plus, afin de permettre une meilleure couverture des âges de liquidation et de mieux prendre en compte les possibilités de départs anticipés. L'EIR 2012 avait été complété des pensions d'invalidité (hors celles du régime général versées par la CNAM). Puis pour la vague de 2016, l'échantillon a été enrichi par l'ajout de nouvelles générations et a été élargi aux pensions d'invalidité versées par le régime général (la CNAM), de manière à couvrir en totalité les bénéficiaires de pensions d'invalidité.

Par ailleurs, pour l'EIR 2020, **le champ est étendu à d'autres prestations proches de l'invalidité, qui correspondent à l'incapacité permanente résultant d'un accident du travail (AT) ou d'une maladie professionnelle (MP)**. Les prestations ou indemnités versées au titre d'incapacité temporaire ou qui ne conduisent pas à un retrait total du marché du travail sont exclues. Entrent dans ce champ les rentes viagères d'invalidité dans la fonction publique et les rentes d'incapacité permanentes versées par la Cnam et les fonds spéciaux comme le FIVA, le FCAT et le FCAATA (avec, dans les deux cas, l'information sur l'éventuelle majoration tierce personne – MTP – associée).

L'EIR 2020 vise donc à représenter les personnes qui perçoivent une pension de droit direct ou de droit dérivé (retraite et invalidité y compris incapacité permanente) d'une caisse de retraite âgées de 20 ans ou plus au 31 décembre 2020.

A noter, 3 bases constituent l'EIR, à savoir AVANT, INDIV et DDIR (cf. Création de nouvelles variables et des bases finales AVANT, INDIV et DDIR, p35).

Travaux réalisés avec l'EIR 2020

À la Drees, l'EIR est directement utilisé pour des études (voir *infra*), mais sert également indirectement pour estimer des agrégats annuels. En effet, estimer le nombre de retraités annuel ou la pension moyenne n'est pas aisé, étant donné que le système de retraite français compte de nombreux régimes (plus de 40) et que la plupart des assurés perçoivent des pensions de plusieurs régimes. L'EIR est une source de référence pour estimer le nombre annuel de retraités ou la pension moyenne, puisque par définition elle contient l'ensemble des pensions de l'ensemble des assurés. Néanmoins, comme l'EIR n'est réalisé que tous les 4 ans, une solution alternative a été développée pour les années inter-EIR : un modèle de microsimulation statique, ANCETRE, qui consiste, en première approximation, à créer un mini-EIR annuel, à partir du dernier EIR connu, en vieillissant les individus et en créant de nouvelles liquidations. Un double calage sur marges, sur les effectifs et les montants de pension, est effectué sur les données de l'enquête annuelle auprès des caisses de retraite. Ces agrégats annuels sont présentés dans Le panorama annuel [Les retraités et les retraites](#).



Fondements juridiques

La loi du 9 juillet 1984 institue l'Échantillon Interrégimes de Retraités : elle fait obligation à tous les organismes concernés (caisses de retraite, Insee, Unedic puis Pôle Emploi) de participer à l'opération et de communiquer les données nécessaires à la constitution de l'EIR au Ministère des Affaires Sociales (SESI, puis Drees). Les principaux textes juridiques encadrant l'EIR sont les suivants. Deux décrets (16 janvier, 3 avril 1985) autorisent l'utilisation du RNIPP (Répertoire National d'Identification des Personnes Physiques) dans le cadre de l'enquête. La CNIL (Commission Nationale de l'Informatique et des Libertés) a définitivement validé le principe de l'EIR en autorisant en 1992 le renouvellement périodique (tous les quatre ans) de l'opération de collecte des données.

Le décret du 22 juillet 2003 pose le cadre général du déroulement de l'EIR et de l'EIC. Il crée plusieurs articles du code de la Sécurité sociale relatifs à ces opérations¹. Les modalités liées à chaque vague (générations sélectionnées, modifications des variables, etc.) sont établies par arrêté. Celui relatif à l'EIR 2020 est l'arrêté du 16 avril 2021.

Textes juridiques

[Arrêté du 29 janvier 1993 autorisant la mise en place d'un système automatisé permanent d'information sur les retraites - Légifrance](#)

[Arrêté du 16 février 2017 \(JO du 28 février 2017\) modifiant l'arrêté du 29 janvier 1993 autorisant la mise en place d'un système automatisé permanent d'information sur les retraites](#)

[Décret n° 2015-1570 du 1^{er} décembre 2015 \(JO du 3 décembre 2015\) autorisant la mise en œuvre d'un traitement de données à caractère personnel à partir de l'échantillon interrégimes de retraités \(EIR\) et de données fiscales en vue de la constitution de statistiques et autorisant l'extension aux pensions d'invalidité du champ des échantillons interrégimes de cotisants et de retraités](#)

[Arrêté du 27 mars 2013 \(JO du 29 mai 2013\) modifiant l'arrêté du 29 janvier 1993 autorisant la mise en place d'un système automatisé permanent d'information sur les retraites](#)

[Arrêté du 20 mars 2009 \(JO du 1^{er} avril 2009\) modifiant l'arrêté du 29 janvier 1993 autorisant la mise en place d'un système automatisé permanent d'information sur les retraites](#)

[Décret n°2003-686 du 22 juillet 2003 \(JO du 29 juillet 2003\) relatif à l'échantillon interrégimes de cotisants et l'échantillon interrégimes de retraités et modifiant le code de la sécurité sociale](#)

[Arrêté du 9 novembre 2001 \(JO du 17 novembre 2001\) modifiant l'arrêté du 29 janvier 1993 autorisant la mise en place d'un système automatisé permanent d'information sur les retraites](#) : élargissement de l'échantillon interrégimes de retraités pour une meilleure couverture des générations âgées et pour l'introduction des retraités nés dans les DOM

[Arrêté du 29 septembre 1998 \(JO du 30 octobre 1998\) modifiant l'arrêté du 29 janvier 1993 autorisant la mise en place d'un système automatisé permanent d'information sur les retraites](#) : élargissement de l'échantillon interrégimes de retraités afin de couvrir le champ des retraités de 55 ans et plus

[Arrêté du 29 janvier 1993 \(JO du 9 février 1993\) autorisant la mise en place d'un système automatisé permanent d'information sur les retraites](#) : autorise le renouvellement systématique de l'échantillon interrégimes de retraités, avec une périodicité de quatre ans, dans les conditions décrites par les arrêtés précédents.

[Arrêté du 26 décembre 1989 \(JO du 11 janvier 1990\) modifiant l'arrêté du 17 mars 1988 autorisant la création d'un système automatisé d'information sur les retraités](#) : enrichissement de l'échantillon interrégimes de retraités à partir des déclarations annuelles de salaires, des fichiers de paie de la Fonction Publique (tous deux gérés par l'Insee) et du fichier national des allocataires de l'Unedic (objectif : étude des taux de remplacement, notamment)

[Arrêté du 17 mars 1988 \(JO du 20 mars 1988\) autorisant la création d'un système automatisé d'information sur les retraites](#) : autorise la mise en place d'un « traitement automatisé d'informations pour connaître le nombre de retraités et la distribution statistique des retraites en France ».

¹ [Articles R161-59 à R161-69 du code de la Sécurité sociale](#)

Arrêté du 24 avril 1986 relatif à la mise en place, à titre expérimental, d'un système d'information statistique sur les retraites

Loi n° 84-575 du 9 juillet 1984 portant diverses dispositions d'ordre social (JO du 10 juillet 1984) : obligation faite aux organismes de retraites de transmettre les données nécessaires à la constitution de l'échantillon interrégimes de retraités au ministère compétent.

Déroulement de la production

L'Échantillon Interrégimes de Retraités (EIR) est collecté tous les 4 ans par la Drees. Il s'agit de la neuvième vague de l'opération : la première a eu lieu en 1988, l'expérience a été reconduite en 1993, 1997, 2001, 2005, 2008, 2012 et 2016.

L'EIR est une opération qui se déroule sur une plage temporelle longue : entre les premiers travaux et la diffusion des bases s'écoulent au mieux deux années (voire trois).

La production de l'EIR se déroule en 6 phases principales qui sont mentionnées dans le **Schéma 1** ci-après.



Schéma 1 • Les 6 phases de la production de l'EIR

À l'issue de ces étapes de productions, on obtient 3 bases de données AVANT20, INDIV20 et DDIR20 (cf. Description des tables, p.35).

Chacune de ces étapes sont détaillées dans les parties suivantes.

■ PLAN DE SONDAGE

Construction d'un échantillon anonyme via le principe de « double aveugle »

Pour respecter les obligations de la loi « Informatique et Libertés », la collecte des informations suit la technique dite de « **double aveugle** » destinées à garantir l'anonymat des informations recueillies. Ni l'Insee ni la Drees ne possède la totalité des informations : NIR, numéro d'ordre, ensemble des données sur les retraites perçues.

Cette technique se réalise en 4 étapes décrites ci-après et dans le **Schéma 2**.

Etape 1 : mise en place des règles de sélection

Dans un premier temps, la Drees détermine les règles de sélection de l'échantillon, sur la base **des années et jours de naissance**. Le tirage est réalisé pour des personnes dont le mois de naissance est connu et des personnes dont le mois de naissance est inconnu.

Pour les personnes dont le jour de naissance est connu, la Drees sélectionne des jours de naissance dits « **jours EIR** » qui changent à chaque vague de l'EIR et qui sont positionnés au début de chaque trimestre. L'ensemble des individus nés ces jours sont sélectionnés dans l'échantillon. Le nombre de jour sélectionnés varie avec la génération. Pour les personnes dont le mois de naissance est inconnu dans le RNIPP, la sélection se fait à partir des valeurs de la « clé » du NIR.

Pour rappel, l'EIR est un panel, les mêmes individus sont réinterrogés d'une vague à l'autre, hormis les individus décédés entre deux vagues et une partie des personnes appartenant à des générations ponctuellement surreprésentées à la vague précédente.

Au final, **l'échantillon est conçu pour représenter les personnes âgées de 20 ans ou plus, vivants au 1er janvier 2020.**

Etape 2 : tirage de l'échantillon par l'Insee

Ensuite la Drees demande à l'Insee de tirer un échantillon de personnes correspondant à ces critères dans le **Répertoire National d'Identification des Personnes Physiques (RNIPP)**² pour les personnes nées en France métropolitaine, répertoires tenus par l'Insee et dans les départements d'outre-mer et dans le répertoire de la **Section Hors Métropole (SHM)**, pour les personnes nées à l'étranger et dans les collectivités d'outre-mer COM, section gérée par la Cnav pour le compte de l'Insee. L'EIR 2020 comprend donc des personnes nées en France ou à l'étranger, quel que soit leur pays de résidence, c'est-à-dire :

- les personnes inscrites à l'état civil et donc nées ou décédées en France
- les personnes inscrites dans un régime de sécurité sociale, c'est-à-dire les étrangers ayant travaillé ou s'étant fait soigner en France.

A partir de ces individus l'Insee établit la liste des individus échantillonnés pour l'EIR 2020.

² Il s'agit d'un répertoire tenu par l'Insee recensant les personnes vivantes et décédées nées en France et les français nés à l'étranger.

Étape 3 : envoi de la liste des NIR aux caisses

L'Insee envoie cette liste aux différents organismes d'assurance vieillesse. Les informations transmises sont un numéro d'ordre (identifiant) anonyme, le numéro d'identification au répertoire (NIR), le nom patronyme, le (ou les) prénom(s), le jour et la commune de naissance, etc.

À l'aide des éléments d'identification contenus dans cette liste, les organismes d'assurance vieillesse sélectionnent les individus de l'échantillon qui sont affiliés à leur régime comme retraités ainsi que ceux qui l'étaient mais qui sont décédés entre deux vagues de l'EIR.

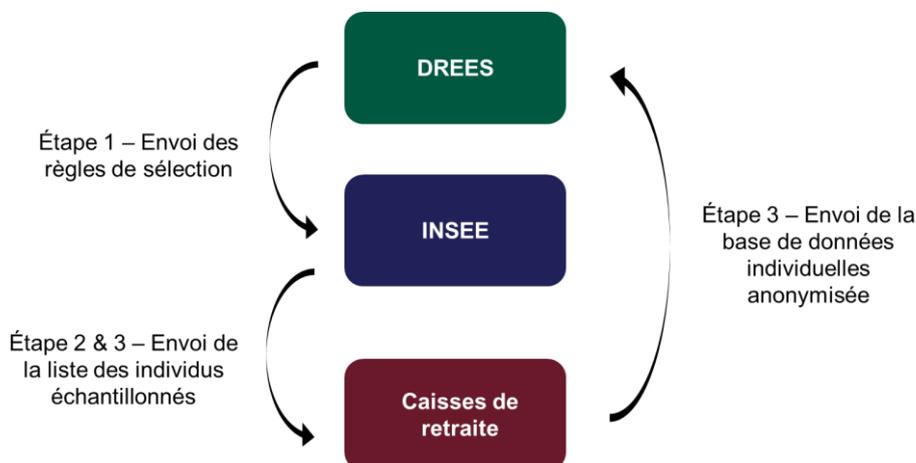


Schéma 2 • Création d'un échantillon anonyme

Étape 4 : création des bases de données anonymes

Enfin les caisses de retraite recueillent pour les personnes échantillonnées identifiées dans leurs données, tous les éléments demandés par la Drees et constitue leur base de données. Celle-ci est ensuite transmise à la Drees.

Dans cette base de données, **les éléments d'identification, en particulier le NIR, sont effacés, de manière à garantir l'anonymat** et à chaque individu un numéro d'ordre anonyme est affecté.

Description détaillée du plan de sondage de l'EIR 2020

Sélection par génération (année de naissance)

Le plan de sondage est stratifié par générations et toutes les générations ne sont pas échantillonnées (*tableau 1*).

- La génération la plus ancienne est constituée des assurés nés en 1915, qui ont 105 ans le 31 décembre 2020 ;
- Entre les générations 1915 et 1930 (inclusive), on sélectionne une génération sur deux ;
- A partir de la génération 1931 (personnes âgées de 89 ans), toutes les générations sont sélectionnées ;
- La génération la plus récente est constituée des assurés nés en 2000, qui ont 20 ans le 31 décembre 2020.

Finalement, 78 générations sont sélectionnées.

Sélection par jour de naissance

Pour chaque génération échantillonnée, les individus de l'EIR sont sélectionnés en fonction de leur date de naissance. La majorité d'entre eux sont nés parmi les premiers jours de janvier, avril, juillet et octobre à raison de 22 jours échantillonnés par générations hormis pour :

- Les générations de 1941 et avant qui ont 79 ans et plus au 31 décembre 2020 qui sont sous-représentées et pour lesquelles seulement 10 jours sont sélectionnés (5 jours pour la génération 1915) : on sélectionne les individus nés les premiers jours d'octobre.

■ Certaines générations sont surreprésentées³ :

- La génération 1950 est surreprésentée dans la strate des jours de naissance connus en sélectionnant davantage de jours de naissance : aux 22 déjà sélectionnés, on en ajoute 14 autres (11 au 24 octobre), de sorte que le taux de sondage est légèrement inférieur à 10 % (36 jours sur 365).
- Pour l'EIR 2020, les générations de 1951 à 1960 sont également surreprésentées : aux 22 jours déjà sélectionnés, on en ajoute 32 autres. En effet, pour ces générations les jours sélectionnés s'étalent sur toute l'année, permettant ainsi d'estimer les effets de certaines réformes qui touchent parfois différemment les individus d'une même génération, comme la réforme des retraites de 2010.

Le **Tableau 1** présente les générations sélectionnées pour la construction de l'échantillon de l'EIR 2020 et le nombre de jours sélectionnés par génération. Au final 825 jours ont été sélectionnés contre 424 en 2016.

Concrètement, l'échantillon est encore stratifié :

- La première strate est constituée des personnes du RNIPP pour lesquelles le jour de naissance est connu. Pour ces personnes, on sélectionne l'ensemble des individus du RNIPP nés certains jours (cf. **Tableau 1**).
- La seconde strate est constituée des personnes du RNIPP pour lesquelles le jour de naissance n'est pas connu. Il s'agit exclusivement de personnes nées à l'étranger. Afin d'avoir une meilleure couverture de ces retraités nés à l'étranger, le tirage inclut donc une partie des individus dont le mois de naissance est inconnu, pour toutes les générations interrogées. De façon à avoir un échantillonnage similaire de ces individus, on retient uniquement les individus dont les 2 derniers chiffres du NIR (« clé de NIR ») appartiennent à la liste suivante : 06, 16, 27, 30, 43, 60, 70, 79 et 89 (le nombre de clés de NIR retenu pour une génération dépend du taux de sondage parmi les mois connus pour la génération concernée)⁴.

³ Les raisons pour lesquelles ces générations sont surreprésentées sont détaillées dans la partie Évolutions du plan de sondage par rapport à l'EIR 2016 p.4.

⁴ Pour l'EIR 2008, l'intégralité des individus nés à l'étranger dont le mois de naissance était inconnu avait été ajouté à l'échantillon et aucun pour l'EIR 2004. Pour l'EIR 2012, les individus dont les 2 derniers chiffres du NIR appartiennent à la liste suivante : 06, 27, et 79 avaient été inclus.

Tableau 1 • Jours de naissance retenus pour chaque génération sélectionnée pour le tirage de l'échantillon, pour les personnes dont le mois de naissance est connu

Génération	EIR 2016		EIR 2020	
	Jours sélectionnés	N	Jours sélectionnés	N
1915	1-5 octobre	5	1-5 octobre	5
1918	1-10 octobre	10	1-10 octobre	10
1920	1-3 octobre	3	1-10 octobre	10
1922	1-10 octobre	10	1-10 octobre	10
1924	1-3 octobre	3	1-10 octobre	10
1926	1-10 octobre	10	1-10 octobre	10
1928	1-3 octobre	3	1-10 octobre	10
1930	1-6 octobre	6	1-10 octobre	10
1931	-	0	1-10 octobre	10
1932	1-6 octobre	6	1-10 octobre	10
1933	-	0	1-10 octobre	10
1934	1-6 octobre	6	1-10 octobre	10
1935	-	0	1-10 octobre	10
1936	1-6 octobre	6	1-10 octobre	10
1937	-	0	1-10 octobre	10
1938	1-6 octobre	6	1-10 octobre	10
1939	-	0	1-10 octobre	10
1940	1-6 octobre	6	1-10 octobre	10
1941	-	0	1-10 octobre	10
1942	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1943	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1944	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1945	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1946 à 1949	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1950	2-5 janvier, 1-4 avril, 1-4 juillet, 1-24 oct	36	2-5 janvier, 1-4 avril, 1-4 juillet, 1-24 oct	36
1951 à 1960	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, du 1-4 pour tous les mois sauf janvier et octobre, 1-10 octobre	54
1961 à 1964	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1965	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1966	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1967	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1968	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1969	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1970	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1971	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1972	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1973	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1974	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1975	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1976	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1977	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1978	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1979	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1980	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1981	-	0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1982	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22
1983 à 2000		0	2-5 janvier, 1-4 avril, 1-4 juillet, 1-10 octobre	22

Source : EIR 2020

Note : N est le nombre de jours sélectionnés.

Évolutions du plan de sondage par rapport à l'EIR 2016

Pour la vague 2020, la Drees a élargi le tirage de l'échantillon à la fois la plage des générations ciblées mais aussi les dates concernées au sein des générations, par rapport à l'EIR 2016. On peut le voir dans le Tableau 1 dans lequel sont indiqués les générations sélectionnées et les jours par génération sélectionnés dans l'EIR 2020 et dans l'EIR 2016. Nous notons que :

- Les générations de 1931 à 1945 sont toutes échantillonnées alors que pour la vague 2016 seulement une génération sur deux l'était. Idem pour les générations de 1965 à 1982 ;
- Le nombre de jours échantillonnés pour les générations 1920, 1924, 1928 a augmenté, passant de 3 jours sélectionnés à 10. Idem pour les générations 1930, 1932, 1934, 1936, 1938 et 1940, avec un nombre de jours sélectionné passant de 6 à 10 ;
- Les générations de 1951 à 1960 sont surreprésentées avec 54 jours échantillonnés contre 22 pour la vague 2016 ;
- Les générations de 1983 à 2000 ont été ajoutées.

Premièrement, le fait de sélectionner toutes les générations :

- De 1930 à 1968 soit celles des individus ayant entre 52 ans et 90 ans au 31 décembre 2020, permet une meilleure couverture des générations entièrement parties en retraite.
- Des générations les plus jeunes, à savoir celles de 1969 à 2000, permet à la Drees de suivre avec plus de précision les bénéficiaires d'invalidité et d'incapacité permanente. En effet, au-delà de la retraite, l'EIR porte également sur l'invalidité au sens large, englobant les pensions et rentes d'invalidité ainsi que les rentes d'incapacité permanente. Si l'invalidité au sens strict (pension et rentes d'invalidité) concerne surtout des personnes de 50 ans et plus, l'incapacité permanente concerne des personnes de tous âges. L'extension de l'EIR 2020 à l'incapacité permanente nécessite donc d'étendre l'interrogation à l'ensemble des âges ; on aurait pu se contenter d'interroger une génération sur deux, comme pour les plus anciennes générations, mais cette solution pose des difficultés méthodologiques (calcul des poids), et peut limiter les capacités d'analyse en cas d'évolution des règles sur une génération donnée.

Deuxièmement, augmenter les nombres de jours échantillonnés pour les générations de 1951 à 1960 permet d'améliorer la qualité des résultats des études sur les taux de remplacement et les âges de liquidation en termes de représentativité. En effet, le détail par trimestre s'est avéré insuffisant pour plusieurs réformes passées ou étudiées. Les générations 1952 et 1953 auront alors atteint l'âge d'annulation automatique de la décote en 2018, 2019 ou 2020 ; la réforme des retraites de 2010 relève progressivement cet âge de 65 à 67 ans.

Troisièmement, l'augmentation à la fois du nombre de générations sélectionnées mais aussi du nombre de jour sélectionnés pour les générations les plus anciennes (antérieures à 1930), permet de répondre à des demandes récurrentes sur la mortalité différentielle qui nécessite des tailles d'échantillon suffisantes.

Enfin, nous notons que la surreprésentation de la génération 1950 est conservée pour exploiter la dimension panel de l'EIR.

Globalement, ce tirage permet de se rapprocher de la construction des panels gérés par l'Insee à savoir, le Panel tous salariés, ainsi que l'Échantillon démographique permanent. De plus, il offre une meilleure représentativité des retraités et de leurs droits. Cette augmentation de la taille de l'échantillon répond notamment à une demande d'utilisateurs, en particulier des chercheurs. En effet, certaines études ne pouvaient être réalisées avec une précision suffisante avec les EIR passés du fait d'une taille d'échantillon trop faible.

Une dernière modification à noter est le fait que l'échantillon est constitué des **personnes vivantes au 1^{er} janvier 2020 et non au 31 décembre** contrairement aux vagues précédentes. Cette évolution importante nous permet de disposer des dernières informations de retraite des personnes décédées entre le 1^{er} janvier et le 31 décembre 2020.

Au final, l'échantillon est conçu pour représenter les personnes âgées de 20 ans ou plus, vivants au 1er janvier 2020, soit une population plus jeune que l'édition 2016.

d'une caisse et reconnue comme décédée est exclue du fichier, même si elle apparaît encore vivante dans une autre caisse. ;

- Certains assurés ne sont pas bénéficiaires d'une pension de retraite en rente. Ils peuvent être bénéficiaire d'un VFU ou de pensions d'invalidité, on retrouvera ces derniers dans la base relative à l'invalidité.

Globalement, l'échantillon comprend un peu plus de femmes que d'hommes. En effet, elles représentent 53 % de l'échantillon EIR (cf. **Figure 3**). La génération 1951, surreprésentée, représente 8,2 % de l'échantillon, et toutes les générations surreprésentées, soit de 1950 à 1960, représentent à elles seules plus de la moitié de l'échantillon (67,7 %).

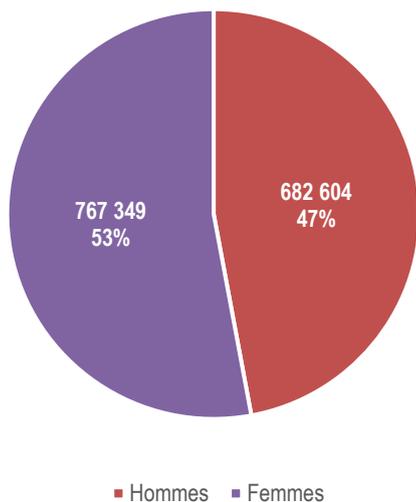


Figure 3 • Répartition de l'échantillon RNIPP en fonction du sexe

Source : RNIPP 2020

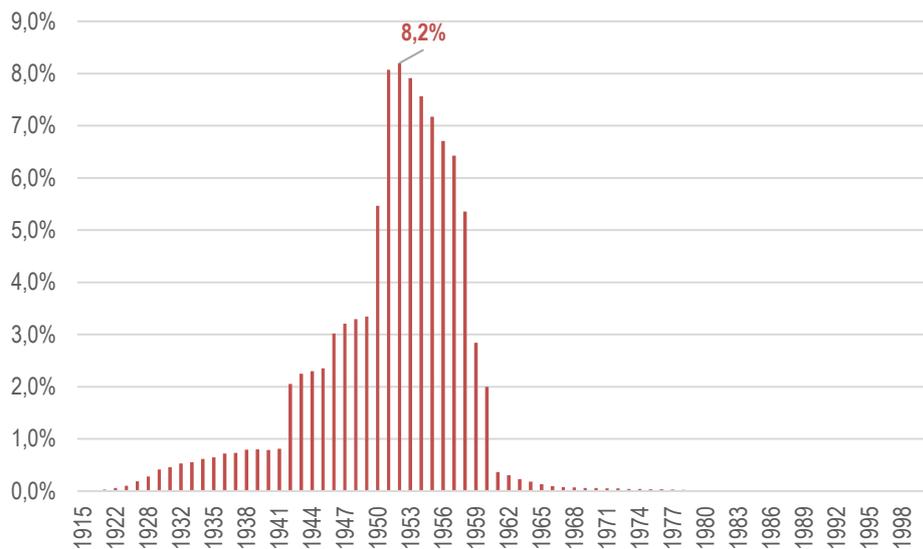


Figure 4 • Répartition de l'échantillon RNIPP en fonction de la génération

Échantillons EIR 2020 vs. EIR 2016 : comparaison des tailles d'échantillon

L'échantillon EIR 2020 est plus de 2 fois plus grand que celui de la vague 2016, avec respectivement 1 449 953 et 644 572 individus, et il est plus de 4 fois plus grand que celui de la vague 2012 (cf. [Figure 5](#)).

Cela s'explique par l'augmentation du taux de sondage pour les générations proches de l'âge de la retraite, qui étaient déjà surreprésentées pour la vague 2016 mais le sont plus encore pour la vague 2020, ainsi que par l'ajout de nouvelles générations.

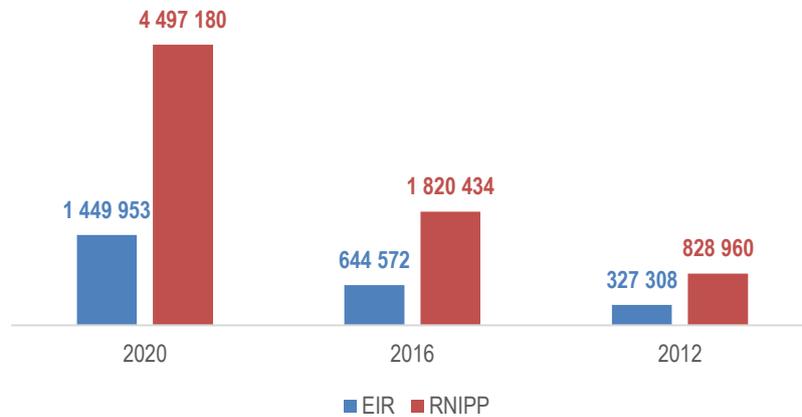


Figure 5 • Taille de l'échantillon du RNIPP vs de l'EIR depuis 2012

Source : EIR 2020, 2016 et 2012 et RNIPP 2020, 2016 et 2012.

■ COLLECTE, CONTRÔLES ET REDRESSEMENTS

Collecte des données

Organismes de retraite participants

Comme mentionné précédemment, chaque caisse de retraite constitue et envoie sa propre base de données à la Drees. Les fichiers qui constituent l'EIR 2020 sont réalisés en rassemblant l'ensemble des informations constituées par les caisses. La quasi-totalité des organismes de retraite obligatoire à savoir, les régimes de base et complémentaires et régimes intégrés sont représentées dans l'EIR, soit 72 au total pour la vague 2020.

En effet, couvrir le plus de régimes de retraite possible est l'un des intérêts principaux de l'EIR, étant donnée la fragmentation importante du système de retraite français et la distribution du nombre de régimes par pensionné (la moitié des assurés perçoivent des droits, directs ou dérivés, d'au moins 3 régimes de retraite). Les retraites issues de régimes supplémentaires non obligatoires sont en revanche exclues du champ de l'opération, ainsi que les revenus provenant de l'épargne individuelle volontaire.

La liste des différentes caisses et régimes participants figure en annexe I. La Cavimac (Caisse d'assurance vieillesse, invalidité et maladie des cultes) appartient au champ de l'EIR mais n'a pas pu participer à la vague 2020 pour des raisons techniques. Les régimes qui sont exclus du champ du panel⁶ sont les suivants :

- la Caisse de retraite du personnel de la Comédie-Française (403 retraités ; 352 cotisants) ;
- la Caisse de retraites des personnels de l'Opéra national de Paris (1 811 retraités ; 1 894 cotisants) ;
- le Port autonome de Strasbourg (203 retraités ; 156 cotisants) ;
- la Caisse de retraite des députés de l'Assemblée nationale ;
- la Caisse de retraite des Membres du Conseil social, économique et environnemental ;
- la Caisse autonome de retraite des anciens sénateurs ;
- la Caisse des Retraites du Personnel du Sénat (769 retraités) ;

Variables demandées

Pour chacun des individus de l'EIR, les caisses de retraite renseignent dans une base de données :

- **Les caractéristiques individuelles du pensionné** : sexe, date de naissance, situation matrimoniale, lieu de résidence, nombre d'enfants etc.
- **Des informations sur les prestations versées** :
 - Sur les droits directs : montant en rente ou en Versement Forfaitaire Unique (VFU), dates d'entrée en jouissance / de liquidation, type de prestation versée, taux de liquidation, taux plein (et pour quel raison), nombre de trimestres de surcote / décote, etc.
 - Sur les droits dérivés : montant en rente ou VFU, dates d'entrée en jouissance / de liquidations, type de prestation perçue par l'ayant droit, etc.
 - Sur les différents montants de majorations conjoint à charge, tierce personne, pour nombre d'enfants, etc.
 - Sur les durées : les durées liquidables validées dans le régime et les durées d'assurance tous régimes
 - Sur les conditions de liquidation des droits à la retraite : âge de liquidation, taux de liquidation, durée de carrière validée, décote, surcote, pension portée à un minimum ou écartée, etc.
 - Sur les points attribués au titre du droit direct (pour les régimes à points)
- **Des caractéristiques de la carrière** : première et dernière année de cotisation / validation d'un droit direct, salaire annuel moyen

⁶ Les informations sur les régimes sont reprises du site du GIP Union Retraite (<https://www.info-retraite.fr/portail-info/sites/PortailInformationnel/home/qui-sommes-nous/les-membres-de-lunion-retraite.html>)

Au total, plus de 170 variables sont renseignées pour chaque individu et par régime (cf. les dessins de fichiers en annexe).

Parmi celles-ci, on en compte 18 nouvelles :

- La commune de résidence (CR_i et CR_p)⁷ ;
- Le taux de prélèvement à la source (TXPAS et NATURE_PAS) ;
- L'Allocation supplémentaire d'invalidité (M8_ASI) et allocation de solidarité aux personnes âgées (M8_ASPA) ;
- Le montant mensuel brut de la retraite du combattant (M1_RC)
- La première perception du minimum vieillesse (ENTMVAA et ENTMVMM) ;
- Les variables liées à la liquidation unique dans les régimes alignés (LURA_RG, LURA_SSI, LURA_MSA, CC_FUSION) ;
- Les coefficients temporaires de solidarité et majorant à la liquidation à l'Agirc-Arrco (INDICCOEFLIQA, DU-REECOEFLIQA) ;
- L'indicatrice cadre/non-cadre avant la fusion Agirc-Arrco (CADRE)
- Le nombre de points cotisés sous le plafond de la sécurité sociale à l'Agirc-Arrco (NPOINDD1_PSS)
- Les variables liées à l'incapacité permanente et l'invalidité (TXINVALINCAP et NATURE_RENTE)

Contrôles et redressements par caisse

Cette phase de contrôle puis de redressement i.e. corrections correspond à une longue étape de la production de l'EIR. En effet, les fichiers bruts, envoyés par les caisses présentent parfois des erreurs, incohérences et des valeurs manquantes. Les données transmises par les régimes de retraite sont donc parfois amenées à être modifiées, le plus souvent dans le but d'harmoniser les données, mais également pour corriger ces erreurs / incohérences ou valeurs manquantes.

Cette phase est réalisée pour chacune des bases de données des caisses de retraite une à une au moment de leur réception.

La Drees commence par faire un premier « nettoyage des données » et s'assure que dans chaque base de données par caisse il y ait, une seule ligne par individu (pas de doublon) et que chaque individu ait au moins un droit direct ou dérivé versé en rente ou en VFU.

Ensuite, les contrôles des variables sont de 5 types :

- 1. Contrôle du remplissage des variables :**
 - pour les variables qualitatives, il s'agit de s'assurer que la variable est renseignée avec les bonnes modalités,
 - pour les variables quantitatives, il s'agit de s'assurer qu'il n'y ait pas de données aberrantes et, que les maxima et minima soient bien respectées s'il y en a.
- 2. Contrôle de la cohérence entre les informations individuelles fournies pour les caisses et celles fournies par l'Insee :** il s'agit, par exemple, de vérifier que l'année de naissance renseignée par la caisse corresponde à celle renseignée par l'Insee
- 3. Contrôle de la présence dans l'EIR 2020 des individus de l'EIR 2016 s'ils ne sont pas renseignés comme étant décédés entre les deux vagues**
- 4. Contrôle la cohérence entre les variables :** par exemple pour les variables DR et PR respectivement, le département et pays de résidence, il arrive qu'un individu soit renseigné avec, à la fois, un département de résidence français et un pays de résidence qui n'est pas la France. Dans ce cas, la Drees harmonise ces deux variables entre elles, pour qu'elles soient en parfaite concordance.

⁷ La correspondance code postal / code Insee : <https://www.data.gouv.fr/fr/datasets/base-officielle-des-codes-postaux/>

5. Contrôle de l'évolution des montants versés (M1, M2 et MONT) : il s'agit de contrôler les montants renseignés en 2020 vs. en 2016 pour les individus présents dans les 2 vagues.

Après un ou plusieurs échanges avec les caisses celles-ci et / ou la Drees effectuent des corrections, qu'on appelle les « redressements par caisse ».

Redressements globaux

Une fois les contrôles et les redressements par caisses réalisés, la Drees fusionne les bases des données des différents régimes. D'autres contrôles et redressements sont effectués à partir de cette base fusionnée, notamment pour vérifier les cohérences des informations renseignées par les différentes caisses. La Drees vérifie par exemple que la date de naissance d'un individu est la même dans toutes les caisses. C'est cette base de données finale qui, après divers traitements (apurement, codifications, etc.) sert à constituer les fichiers de l'EIR.

■ PONDÉRATIONS

L'EIR est un échantillon de la population des retraités. Pour calculer des statistiques représentatives sur l'ensemble des retraités, les individus de l'EIR sont pondérés. La construction des variables de pondération est détaillée dans cette partie.

L'EIR comporte deux jeux de pondération. Le premier jeu rend l'échantillon représentatif au 31 décembre 2020 (variable POND20_CAL). Pour certaines analyses, il peut être utile de corriger de la mortalité différentielle. En effet, les assurés qui meurent de façon anticipée peuvent avoir un profil spécifique ce qui peut induire un biais dans la comparaison entre générations sur la base des seuls pensionnés vivants fin 2020. Pour pallier cela, la pondération est ajustée (variable pond20_genliq).

Tableau 2 • Variables de pondération

Nom de la variable	Dans quel cas les utiliser
POND20_CAL	Réaliser des statistiques par âge quinquennaux ou agrégés (représentativité au 31 décembre 2020)
POND20_GENLIQ	Réaliser des statistiques par génération (représentativité de l'ensemble des personnes de la génération ayant perçu une pension de retraite).

Les pondérations ont été calculées de la façon suivante :

- D'abord, des **pondérations initiales** sont déterminées suivant deux modalités. Pour les générations les plus anciennes (entre 1915 et 1928), qui représentent plusieurs générations, on utilise les effectifs par sexe, origine et âge donnés par le RNIPP (répertoire national d'identification des personnes physiques). Des corrections sont apportées pour corriger :
 - La sous-mortalité des individus nés à l'étranger
 - La sous-mortalité des individus plus âgés (principalement à partir de 90 ans)
 - Le biais d'échantillonnage
- Pour les générations 1930 à 2000 qui ne représentent qu'elles-mêmes, les pondérations initiales sont déterminées comme étant le ratio entre le nombre de jours de tirage et le nombre de jours dans l'année (pour les générations 1930-1941 pour lesquelles on ne tire que les 10 premiers jours d'octobre), dans le trimestre (pour les générations 1942-1950 et 1961-2000 pour lesquelles on ne tire que des premiers jours des premiers mois des différents trimestres) ou dans le mois (pour les générations 1951-1960 pour lesquelles on a tiré des jours dans l'ensemble des mois de l'année).
- Ensuite, les pondérations initiales sont calées (par la technique statistique du calage sur marges) sur des variables exogènes (effectifs, masses de pension, etc.) de l'enquête annuelle auprès des caisses de retraites (EACR) à un niveau fin (POND20_CAL).
- Enfin, des pondérations corrigeant la mortalité différentielle sont également calculées (POND20_GENLIQ). Les probabilités de survie entre la liquidation d'un premier droit à retraite et l'année de collecte de l'EIR sont estimées pour chaque retraité. Ces pondérations permettent d'être représentatifs de l'ensemble des retraités ayant perçu une pension de retraite.

Ces 3 étapes sont explicitées dans les parties ci-après.

Les pondérations initiales

La construction des pondérations initiales (variable POND20_INI) vise à rendre l'échantillon représentatif de l'ensemble des individus qui ont pu acquérir des droits à pension dans un régime de retraite français. Pour cela, on procède selon deux modalités et en plusieurs étapes.

Pour les générations 1915 à 1928, qui représentent aussi les générations adjacentes

- On commence par calculer des pondérations qui rendent l'échantillon représentatif du RNIPP ;
- Ensuite, on corrige à trois niveaux :
 - Ayant constaté dans le RNIPP une sous-mortalité aux âges élevés des individus nés à l'étranger (par rapport aux personnes nées en France), on la corrige, en abaissant les pondérations des assurés nés à l'étranger, et ce d'autant plus que leur âge est élevé.
 - Une seconde correction est appliquée à partir de 90 ans, pour se caler sur la mortalité du Bilan démographique de l'Insee ;
 - Enfin, une troisième correction est appliquée, pour corriger du biais d'échantillonnage dans les mois de l'année.

Calage sur le RNIPP

Dans un premier temps, l'échantillon Insee est calé sur les effectifs par sexe, par lieu de naissance et par âge du RNIPP, en ajustant les pondérations. Le calage est nécessaire car l'échantillon Insee est vraisemblablement biaisé, car il ne contient pas toutes les générations et uniquement des individus nés début octobre pour ceux nés avant 1949. Par ailleurs, pour certains individus nés à l'étranger, le critère d'échantillonnage ne peut pas être le mois de naissance.

Le poids initial P d'un individu i de sexe s , dont la génération g appartient au groupement de générations R , inscrit dans la section sec du RNIPP et présent dans l'échantillon de l'EIR s'écrit :

$$P_{i, g \in R, s, sec} = \frac{\sum_{g \in R} N_{i, g \in R, s, sec}}{\sum_{g \in R} n_{i, g \in R, s, sec}}$$

où N est le nombre total d'individus de sexe s , appartenant à la génération g inscrit dans la section sec du RNIPP vivants au 31/12/N, et n le nombre d'individus de ce sexe et de cette génération retrouvés dans l'échantillon une fois appliquées les règles de tirage spécifiée pour l'EIR.

Correction de la sous-mortalité dans individus nés à l'étranger

La première correction vise à pallier le fait que le RNIPP contient une proportion importante de « faux vivants », car les décès sont parfois enregistrés avec un délai ou ne sont pas enregistrés dans le répertoire. Cette sous-estimation des décès concerne majoritairement les générations âgées et les individus nés à l'étranger. L'échantillon Insee étant tiré au sein du RNIPP, il est également concerné par ce problème. Une génération de l'échantillon Insee peut être représentative de plusieurs générations du RNIPP, or la sous-estimation des décès n'affecte pas de la même manière les différentes générations. Les pondérations calculées peuvent donc être biaisées. Il est important de corriger l'échantillon Insee et le RNIPP avant de calculer les pondérations.

Les taux de survie en 2020 calculés à partir du RNIPP 2019 et 2020 sont plus élevés pour les individus nés à l'étranger (section hors métropole SHM) que pour ceux nés en France (section métropole – SM). Il est vraisemblable que ce différentiel soit dû uniquement au fait que certains décès ne sont pas renseignés dans le RNIPP-SHM, les taux de survie devant sinon être similaire à âge et sexe donné. On « corrige » donc les effectifs des individus nés à l'étranger de manière à ce que leur taux de survie soit le même que celui des individus nés en France. Par exemple, pour les personnes qui ont 44 ans en 2020 (génération 1976), on applique la correction suivante :

$$\text{Effectif corrigé}_{44}^{SHM} = \text{Effectif RNIPP}_{44}^{SHM} * \frac{\text{TxSurvie}_{44}^{SM}}{\text{TxSurvie}_{44}^{SHM}} * \frac{\text{TxSurvie}_{43}^{SM}}{\text{TxSurvie}_{43}^{SHM}} * \dots * \frac{\text{TxSurvie}_{20}^{SM}}{\text{TxSurvie}_{20}^{SHM}}$$

Pour calculer le taux de survie à 20 ans de la génération 1968, il faudrait se baser sur le RNIPP 1988 et le RNIPP 1987. Ne disposant pas de ces bases de données, nous nous basons uniquement sur le RNIPP 2020 et 2019 pour déterminer l'ensemble des taux de survie. La correction apportée permet de rapprocher les taux de survie des individus nés à l'étranger et des individus nés en France et tend à réduire les effectifs d'individus nés à l'étranger de façon importante.

Correction de la sous-mortalité dans individus plus âgés

La deuxième correction vise à corriger la sous-mortalité aux âges élevés. En effet, en comparant le RNIPP après application de la correction apportée aux individus nés à l'étranger et le bilan démographique de l'Insee, il apparaît que le RNIPP sous-estime les décès principalement à partir de 90 ans. Après cet âge, on redresse les effectifs du RNIPP afin de les faire décroître au même rythme que le Bilan démographique. Les mêmes corrections sont apportées à l'échantillon Insee. Une fois ces biais corrigés sur le RNIPP et l'échantillon Insee, des pondérations sont ensuite calculées.

Pour les générations 1930 à 2000

Pour ces générations, qui ne représentent qu'elles-mêmes, on fait l'hypothèse que les jours de naissance des individus sont une variable totalement aléatoire (c'est une hypothèse assez faible bien que le 1^{er} janvier et le 1^{er} juillet sont plus fréquents pour les dates de naissance). Ainsi, on détermine la pondération initiale d'un individu selon son mois de naissance en faisant un ratio du nombre de jours de tirage du mois sur le nombre de jour dans l'année, le trimestre ou le mois selon les générations.

Pour les générations 1930 à 1941, on ne tire que les 10 premiers jours d'octobre. Ainsi, pour les individus de ces générations, la pondération initiale vaudra 365/10 (on fait l'approximation des années bissextiles).

Pour les générations 1942-1949 et 1961-2000, on a tiré les 10 premiers jours d'octobre, les 4 premiers jours d'avril et de juillet ainsi que les 2-5 janvier. Ainsi, pour les individus de ces générations, le poids initial vaut :

- 90/4 si l'individu est né en janvier
- 91/4 si l'individu est né en avril
- 92/4 si l'individu est né en juillet
- 92/10 si l'individu est né en octobre

Suivant la même logique, la pondération initiale de la génération 1950 vaut :

- 90/4 si l'individu est né en janvier
- 91/4 si l'individu est né en avril
- 92/4 si l'individu est né en juillet
- 92/24 si l'individu est né en octobre

Enfin, pour les générations 1951-1960, les pondérations initiales valent :

- 31/4 si l'individu est né en janvier, mars, mai, juillet, août ou décembre
- 30/4 si l'individu est né en avril, juin, septembre ou novembre
- 28/4 si l'individu est né en février
- 31/10 si l'individu est né en octobre

Résultats de la pondération initiale

Au final, les poids initiaux varient entre 3,1 et 147,85, la moyenne s'élevant à 12,08. D'après le [Figure 6](#), on remarque également que :

- 11 % des individus qui perçoivent un droit direct ont le poids minimum,
- 28 % des individus qui perçoivent un droit direct, soit 38 7161 individus, ont un poids de 7,75,
- La médiane s'élève à 7,75
- Seulement 8 % des individus qui perçoivent un droit direct ont un poids supérieur à 25.

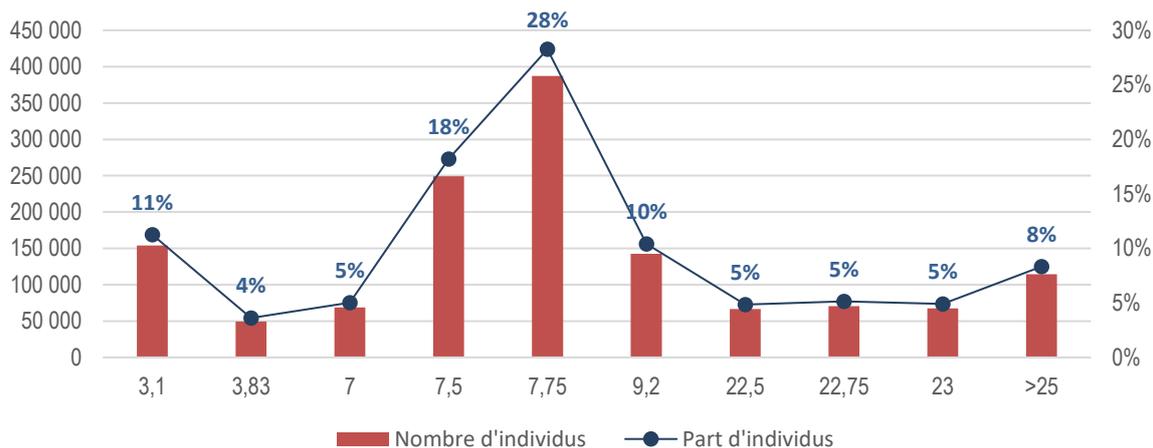


Figure 6 • Distribution des poids initiaux des assurés bénéficiaires d'un droit direct

Source : EIR 2020

Note de lecture : 154 195 individus ayant un droit direct ont un poids de 3,1, ce qui représente 11% de l'échantillon.

Le calage sur marge

Dans un deuxième temps, les pondérations initiales sont ajustées pour assurer un calage sur des marges exogènes connues. Le calage sur marge permet de caler finement les effectifs de retraités et les montants moyens par génération pour les principales caisses de retraites (variable POND20_CAL). Ce calage est réalisé au moyen de la macro CALMAR de l'Insee.

Les marges utilisées sont issues des données de l'Enquête Annuelle auprès des Caisses de Retraite (EACR). L'EACR contient des données agrégées fournies par les régimes de retraite, sur les effectifs de retraités (stock et flux) et les montants moyens de pension pour chaque régime, par sexe (variable SEXI) et –pour les principaux régimes– année et lieu de naissance (variables AI et NAISS). Les marges sont définies à ce niveau de détail, sauf lorsque les effectifs correspondants sont trop faibles (la procédure de calage n'est alors pas possible).

Les marges comportant moins de 100 observations ou entraînant une modification des poids de plus de 25 % sont supprimées en amont du calage.

La méthode de calage est la méthode « logit », en encadrant la déformation maximale à 0,2 et 4. Le calage déforme peu les poids, puisque 97 % des individus ont une déformation comprise entre 0,8 et 1,2 (cf. [Figure 7](#)). Par ailleurs, pour 9 509 individus, le poids n'a pas changé à l'issue de l'étape de calage, soit 0,66% de l'échantillon.

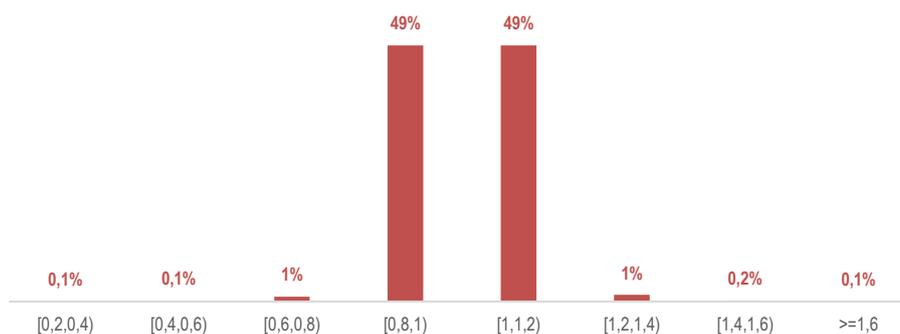


Figure 7 • Distribution des déformations de poids après l'étape de calage en fonction de la génération et du sexe

Source : EIR 2020

Note : La déformation a été calculée en faisant le rapport entre des poids initiaux et les poids calés.

Le calage déconcentre la distribution des poids (cf. Figure 8), cependant l'étendue est plus large puisque les poids varient maintenant de 0,62 à 213,23. Pour autant, seulement 9 % des individus ayant un droit direct ont un poids supérieur à 26. La médiane est inchangée, à 12,08, tandis que la moyenne a très légèrement augmentée passant de 7,75 à 7,78. 45 % des individus ayant un droit direct ont un poids post-calage entre 6 et 8 non inclus.

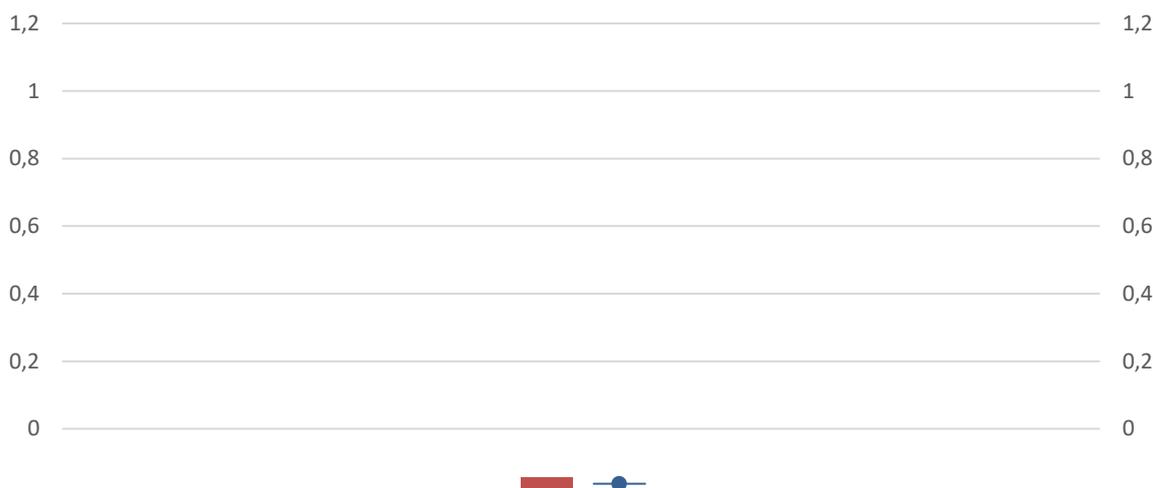


Figure 8 • Distribution des poids calés des assurés bénéficiaires d'un droit direct

Source : EIR 2020

Note de lecture : 200 303 individus ayant un droit direct ont un poids entre 2 et 4, ce qui représente 15% de l'échantillon.

Le Tableau 3, présente le poids final moyen par génération et sexe, ainsi que les écarts de poids entre homme et femme. Les écarts varient de -16 à 7. On remarque également que pour les générations les plus anciennes (1928 et avant, c'est-à-dire pour les individus ayant 92 ans ou plus au 31 décembre 2020) et les plus récentes (1983 et après, c'est-à-dire pour les individus ayant 37 ans ou moins au 31 décembre 2020), on retrouve des écarts de poids moyens supérieurs à 5 en valeur absolue. Pour les autres générations, les poids sont équilibrés entre hommes et femmes, les écarts sont inférieurs à 5 en valeur absolues.

Pour les générations sur-représentées de 1951 à 1960, le poids moyen s'élève à 6,8, c'est-à-dire qu'un individu de l'échantillon représente 6,8 retraités.

Tableau 3 • Poids final moyen par génération, sexe et lieu de naissance

Année de naissance	Ensemble		Ecart Homme vs Femme
	Femme	Homme	
1915	36,08	-	ND
1918	39,29	23,73	-16
1920	38,04	25,96	-12
1922	66,98	65,63	-1
1924	119,98	120,16	0
1926	79,82	81,78	2
1928	80,45	87,04	7
1930	35,71	36,23	1
1931	37,05	36,76	0
1932	36,38	36,56	0
1933	37,19	37,25	0
1934	38,12	37,67	0
1935	37,29	38,24	1
1936	36,95	37,59	1
1937	37,38	37,42	0
1938	36,76	36,53	0
1939	36,75	38,19	1
1940	38,21	38,79	1
1941	35,98	36,15	0
1942	16,25	16,42	0
1943	16,33	16,12	0
1944	16,81	16,20	-1
1945	16,48	16,26	0
1946	16,85	16,69	0
1947	16,87	16,69	0
1948	16,79	16,84	0
1949	16,68	16,47	0
1950	10,26	10,34	0
1951	6,69	6,78	0
1952	6,73	6,77	0
1953	6,82	6,79	0
1954	6,78	6,76	0
1955	6,76	6,79	0
1956	6,86	6,81	0
1957	6,73	6,71	0
1958	6,84	6,78	0
1959	6,84	6,86	0
1960	6,80	6,91	0
1961	16,78	17,22	0
1962	17,05	16,61	0
1963	17,28	16,14	-1
1964	16,70	17,33	1
1965	16,84	16,83	0
1966	15,90	16,70	1
1967	16,35	16,27	0
1968	15,85	16,12	0
1969	15,46	17,02	2
1970	16,93	17,21	0
1971	17,16	16,79	0
1972	15,92	17,39	1
1973	17,70	18,97	1
1974	16,49	17,21	1
1975	14,71	15,67	1
1976	15,22	14,83	0
1977	15,02	15,19	0
1978	15,47	15,76	0
1979	16,41	14,98	-1
1980	16,87	17,18	0
1981	15,22	18,17	3

1982	15,99	15,74	0
1983	10,15	16,24	6
1984	20,11	18,45	-2
1985	10,15	20,11	10
1991	7,50		-7

Source : EIR 2020

Note : les cases en rouge sont les écarts de poids entre homme et femme supérieurs à 5 en valeur absolue.

Note de lecture : une femme de l'échantillon EIR née en 1991 représente en moyenne 7,5 femmes retraitées nées en 1991.

La correction de la mortalité différentielle

Le second jeu de pondération vise à corriger la mortalité différentielle. Cette correction permet de rendre comparables les caractéristiques des diverses générations bien qu'elles ne soient pas observées au même âge dans l'EIR. La mortalité, et donc l'espérance de vie, dépend en effet des caractéristiques des retraités : par exemple, les retraités dont les pensions sont les plus élevées vivent en moyenne plus longtemps. De même, l'âge moyen de décès des hommes est plus précoce que celui des femmes.

Sans cette correction, les retraités des générations les plus anciennes, encore vivants au 31 décembre 2020, ne seraient pas représentatifs de l'ensemble des retraités de ces générations.

En pratique, cela consiste à repondérer les retraités pour rendre l'EIR 2020 représentatif de l'ensemble des personnes ayant perçu une pension de retraite. Concrètement, cette correction augmente les pondérations des assurés dont la mortalité est supérieure à la moyenne (les hommes par exemple). Le but étant que chaque cohorte soit représentative de l'ensemble de ses membres ayant liquidé un droit direct à retraite. Cette pondération n'est calculée que pour les retraités de droit direct.

Comme à l'étape de calage, la Drees a calculé la déformation à l'issue de cette étape de correction de la mortalité. L'étendue de la déformation est large. En effet, celle-ci varie de 0,17 à 1 011,77. Pour autant, comme le calage, la correction déforme peu les poids, puisque 75 % des individus ont une déformation comprise entre 0,17 et 1,2 (cf. **Figure 9**). Par ailleurs, pour 1 435 individus ayant une pension de droit direct en rente, le poids n'a pas changé à l'étape de la correction, soit pour 0,11 % de l'échantillon.

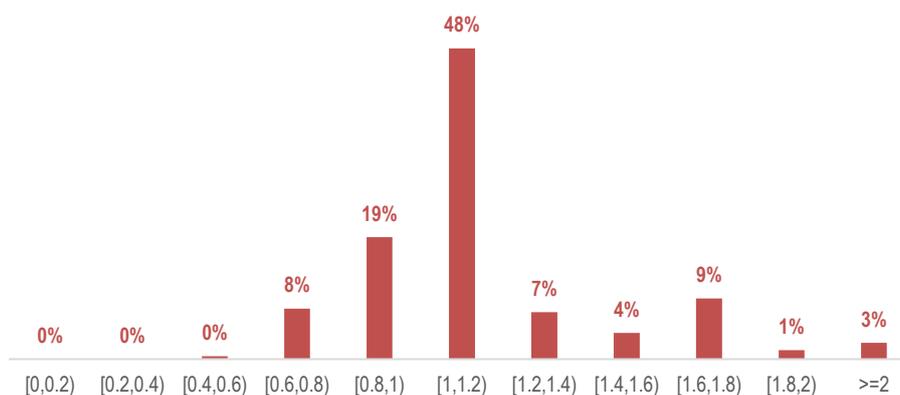


Figure 9 • Poids final moyen par génération, sexe et lieu de naissance

Source : EIR 2020

Note : La déformation a été calculé en faisant le rapport entre des poids calés et les poids corrigés.

Le **Figure 10** montre que, pour les générations les plus récentes, pour lesquelles peu de gens sont décédés, le poids corrigés moyens sont proches des poids calés moyens. En revanche, pour les générations les plus anciennes, la correction de la mortalité différentielle conduit à rehausser les pondérations.

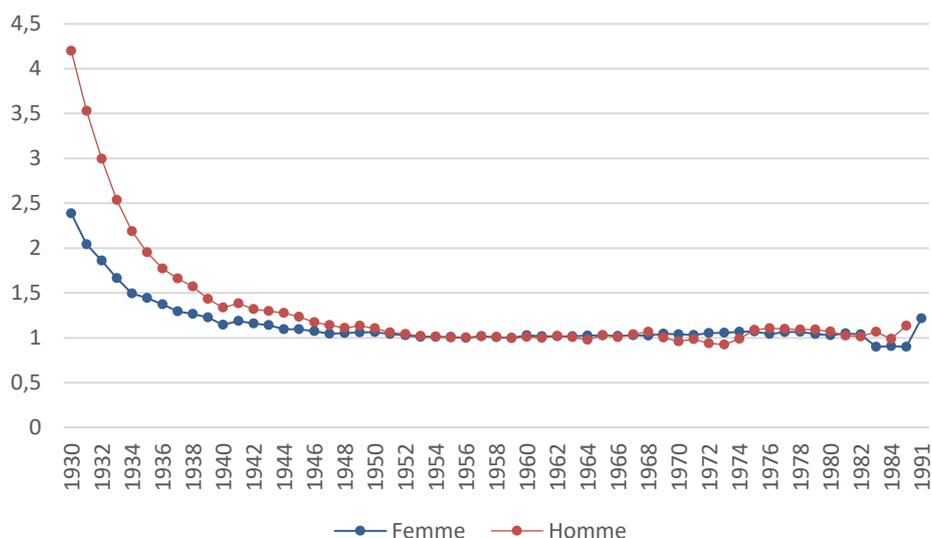


Figure 10 • Déformation des poids moyen à l'issue de l'étape de correction en fonction du sexe

Source : EIR 2020

Note : La déformation a été calculé en faisant le rapport entre des poids calés moyens et les poids corrigés moyens.

À l'aide du panel de données constituées par les différents millésimes de l'échantillon interrégimes de retraités, la Drees estime des probabilités de survie et des espérances de vie individuelles à différents âges à partir des caractéristiques de chaque personne à l'aide d'un modèle de Gompertz.

Les tables de mortalités de l'Insee ne commençant qu'en 1976, pour les anciennes générations les quotients calculés par l'Ined sont utilisés (Vallin Meslé).

Les poids corrigés varient de 3,9 à 22 410,7. 76 % de l'échantillon a un poids inférieur ou égale à 11,4, dont près de 38 % ayant un poids entre 7,6 et 7,8 non compris (cf. **Figure 11**).

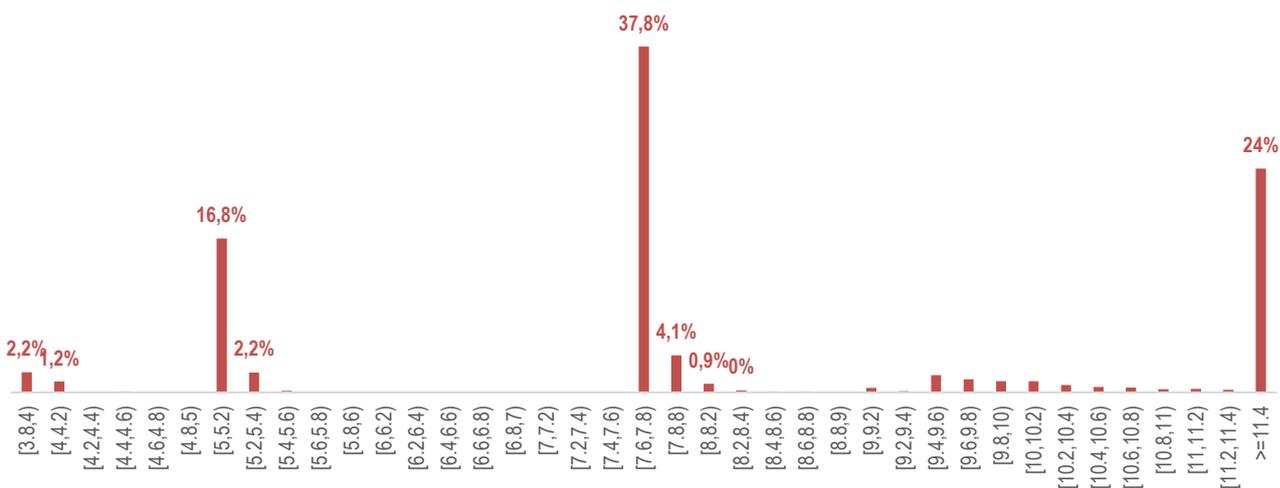


Figure 11 • Distribution des poids corrigés de la mortalité

Source : EIR 2020

■ POST-TRAITEMENTS

À l'issu des phases de contrôles, redressements et pondération, une phase de post-traitement est réalisée. Quatre post-traitements sont appliqués lors desquelles des variables complémentaires sont créées.

Tableau 4 • Variables issus des post-traitements

Nom de la variable	Description
RetrEQCC	Montant mensuel (ou mensualisé) brut de l'avantage principal de droit direct en équivalent carrière complète retrEQCC = M1 / coefficient de proratisation
M1_hors_taux	Montant mensuel (ou mensualisé) brut de l'avantage principal de droit direct, hors taux de décote surcote versé dans le régime. M1_hors_taux= M1 / Taux de décote ou surcote
RetrPleine	Montant mensuel (ou mensualisé) brut de l'avantage principal de droit direct versé à un affilié ayant validé la durée de carrière requise et partant à l'âge minimal où il bénéficie du taux plein (sans anticipation ni prolongation d'activité). retrpleine = M1_hors_taux / coefficient de proratisation
M1_hors_minimum	Montant mensuel (ou mensualisé) brut de l'avantage principal de droit direct, hors surplus lié au minimum de pension (contributif/garanti, majoré ou non au titre des périodes cotisées) M1_hors_minimum=M1 - MAJOMINDD
RESID	Zone de résidence
NBENF	Nombre d'enfant par femme.
PENSIONnet	Pension de retraite de droit direct net des cotisations et augmenté du montant de la majoration pour enfant le cas échéant
tauxrempdav25plein	Taux de remplacement net par rapport aux 5 avant-dernières années à temps plein avant liquidation tauxrempdav25plein = (PENSIONnet / moyennecomplet5m) * 100
txremp50	Taux de remplacement net par rapport au salaire perçu à 50 ans txremp50 = (PENSIONnet / salaire50m) * 100
tauxrempdav25	Taux de remplacement net par rapport à la dernière année avant l'année de liquidation tauxrempdav25 = (PENSIONnet / salairedav25m) *100
moyennecomplet5m	Moyenne des 5 avant-derniers salaires net fiscal mensuel à temps plein avant liquidation
salairedav25m	Moyenne des 5 avant-derniers salaires net fiscal mensuel avant liquidation
secteurdav2liq	Secteur d'activité de la dernière année avant liquidation
salaire50m	Salaire net mensuel à 50 ans
salaire50	Secteur d'activité à 50 ans
secteur2plein	Secteur d'activité pendant la dernière année à temps plein avant l'année de liquidation

Equivalent carrière complète (EQCC)

Principe général

Il s'agit de calculer un montant de pension en EQCC, défini pour chaque assuré comme étant le rapport entre son montant de pension de droit direct et son coefficient de proratisation. Ce dernier se calcule en rapportant la durée de service à une durée de référence définie par la législation, sans pouvoir être supérieur à 100 % au sein de chaque régime.

Cette définition permet d'estimer le montant de la pension dans le cas où la personne aurait une carrière complète (au sens d'une proratisation à 100 % dans le régime). Pour les retraités polypensionnés, le coefficient de proratisation tous régimes correspond à la somme des coefficients de proratisation dans chacun des régimes de base. La pension en EQCC moyenne pour une génération est calculée comme la pension moyenne de la génération divisée par le coefficient de proratisation moyen de la génération.

Les variables suivantes sont calculées pour chaque individu :

- La retraite en équivalent carrière complète (Retreqcc);
- La « retraite pleine » (retrpleine) ;
- La pension hors décote-surcote (M1_hors_taux) ;
La pension hors minima (pensionhmin).

Calcul de RetrPleine : Le montant de « retraite pleine »

Dans la plupart des régimes en annuité, le montant de la pension se calcule comme le produit de trois termes :

$$\text{Pension} = \text{T}x \text{ de décote} - \text{surcote} \times \text{T}x \text{ de proratisation} \times \text{Retrpleine}$$

RetrPleine correspond à la pension versée à un affilié ayant validé la durée de carrière requise et partant à l'âge minimal où il bénéficie du taux plein (sans anticipation ni prolongation d'activité). Elle est estimée à partir des éléments de l'EIR 2020 comme suit :

$$\text{Retrpleine} = \frac{\text{M1_hors_taux}}{\text{coefficient de proratisation}}$$

Pour les régimes complémentaires, la durée de carrière a été estimée à partir des nombres de trimestres validés, hors majorations de durée, dans les régimes de base correspondant. Par exemple, le coefficient de proratisation pour le régime complémentaire obligatoire des non-salariés agricoles est égal au nombre de trimestres validés (hors majorations) divisé par la durée de proratisation dans les régimes de base. L'estimation est plus complexe pour les régimes complémentaires des salariés du privé, dans la mesure où il y a d'une part plusieurs régimes de base correspondant (régime général et MSA salariés, notamment) et où d'autre part différentes configurations d'affiliation sont possibles lorsque les personnes cotisent dans ces régimes de base (ARRCO seule, ARRCO et AGIRC, IRCANTEC, voire aucun régime complémentaire). Pour cette raison, les trois régimes complémentaires de salariés du privé sont étudiés conjointement, en agrégeant les pensions versées par l'ARRCO, l'AGIRC et l'IRCANTEC. La durée correspondante validée en tant que « salarié du privé » sera mesurée à partir des trimestres validés à la CNAV et à la MSA salariés. Si un retraité a validé dans un seul de ces deux régimes, le nombre de trimestres validés hors majoration sera directement observé dans les données de ce régime.

Le montant de retraite pleine pour l'ensemble base / complémentaire est calculé comme la somme des retraites pleines dans chaque régime :

$$\text{RetrPleine_salarié_privé} = \frac{\text{Pension cnavMSAs}/\text{TxDécSur cnavMSAs}}{\text{T}x\text{Prorat cnavMSAs}} + \frac{\sum c \text{ ARRCO AGIRC IRCAN Pension } c/\text{TxDécSur } c}{\text{T}x\text{Prorathors majo compl}}$$

Le montant de retraite pleine tous régimes est calculé, sur le même principe, comme la somme de la retraite pleine de base et de la retraite pleine complémentaire ; la retraite pleine de base étant calculé comme la somme des retraites hors décote/surcote des régimes de base divisée par la somme des taux de proratisation de ces régimes (idem pour les complémentaires) :

$$\text{RetrPleine tous régimes} = \frac{\sum b \in \text{base Pension } b/\text{TxDécSur } b}{\sum \text{T}x\text{Prorat } b} + \frac{\sum c \in \text{Acompl. Pension } c/\text{TxDécSur } c}{\sum \text{T}x\text{Prorathors majo } c}$$

Imputation du lieu de résidence

Les variables géographiques de l'EIR sont :

- RESID indique si l'individu réside en France ou à l'étranger⁸ ;
- DR est le département de résidence (99 pour les résidents à l'étranger) ;
- PR est le pays de résidence.

19,4 % des droits n'avaient pas de département de résidence renseigné (DR) (principalement à l'Agirc Arrco). Cette variable est redressée en priorité par le DR renseigné par la CNAV, puis par un autre régime lorsque cela est possible, si la CNAV ne connaît pas le DR (seulement si STATUT=1, car le critère de résidence pour un VFU ou un suspendu peut être jugé peu fiable). Cela permet de redresser le DR.

Après cette correction, il reste 0,11 % des droits (1 884 lignes d'AVANT, sur 1 670 767) sans résidence attirée : DR= « 00 » ; PR= « » et RESID= « ». Ces droits concernent 0,8 % des personnes de l'échantillon (5 144 personnes).

Pour les régimes de base, plus de 56 % des observations ont été redressées. La grande majorité de ces redressements proviennent de la Cnav qui n'avait pas renseigné cette variable. La variable TRIMDEC a été imputée à partir des variables TAUX et AN.

Contrairement aux vagues précédentes, les variables géographiques dans la base AVANT ont été redressées pour être harmonisées avec les bases INDIV et DDIR. À chaque individu correspond donc un unique département de résidence, duquel dépendent les variables PR et RESID.

Imputation du nombre d'enfant des femmes

Le nombre d'enfant (variable NRE), n'est pas toujours bien renseigné par les caisses (NRE) car cette information n'est pas systématiquement connue dans les systèmes d'informations des caisses. En effet, plus 50% des femmes étaient renseignées comme ayant 3 enfants ou plus (cf. **Erreur ! Source du renvoi introuvable.**) alors que l'on s'attendrait à en avoir autour de 30%.

De fait, la Drees a créé la variable NBENF, issue d'une imputation. Cette imputation a été réalisée uniquement pour les femmes, la variable NBENF est donc vide pour les hommes. Pour ces derniers, on peut seulement savoir s'ils perçoivent une majoration pour les parents d'au moins 3 enfants.

Principe de l'imputation

L'imputation du nombre d'enfants dans l'EIR 2020 est réalisée principalement à partir du nombre de trimestres de majoration de la durée d'assurance pour chaque enfant pris en compte dans la durée validée dans le régime (variable TDD3A).

Cette variable est d'abord redressée afin de la compléter lorsqu'elle est manquante. Ce redressement est effectué, en fonction de plusieurs facteurs :

- la présence d'une bonification pour 3 enfants ou plus,
- la part que représente cette majoration (M5) par rapport au montant de la pension de droit direct et/ou de droit dérivé,
- la présence de cotisations AVPF notamment, information issue de l'EIC.

Une fois la variable TDD3A complétée, on peut en déduire le nombre d'enfants lorsqu'il est manquant de la manière suivante :

- Aucun enfant pour les femmes qui n'ont aucun trimestre de majoration
- Pour la fonction publique et les régimes spéciaux :
 - 1 enfant pour les femmes qui ont 4 trimestres de majoration ou moins
 - 2 enfants pour les femmes qui en ont entre 5 et 8

⁸ Comme dans l'EIR 2012, RESID= « FRANCE » pour la métropole et les DOM, et RESID= « ETRG » pour les TOM/COM et les pays étrangers.

- 3 enfants pour les femmes qui en ont entre 9 et 12
- 4 enfants pour les femmes qui en ont entre 13 et 16
- 5 enfants pour les femmes qui en ont entre 17 et 20
- 6 enfants pour les femmes qui en ont entre 21 et 24
- 7 enfants pour les femmes qui en ont entre 25 et 28
- 8 enfants pour les femmes qui en ont entre 29 et 32
- 9 enfants pour les femmes qui en ont entre 33 et 37
- 10 enfants pour les femmes qui en ont entre 37 ou plus
- Pour le secteur privé et la CNAV :
 - Aucun enfant pour les femmes qui ont 4 trimestres de majoration ou moins
 - 1 enfant pour les femmes qui en ont entre 5 et 12
 - 2 enfants pour les femmes qui en ont entre 13 et 20
 - 3 enfants pour les femmes qui en ont entre 21 et 28
 - 4 enfants pour les femmes qui en ont entre 29 et 36
 - 5 enfants pour les femmes qui en ont 37 ou plus

Le nombre d'enfants est ensuite plafonné à 5.

Résultats de l'imputation

Après imputation, on obtient la distribution suivante :

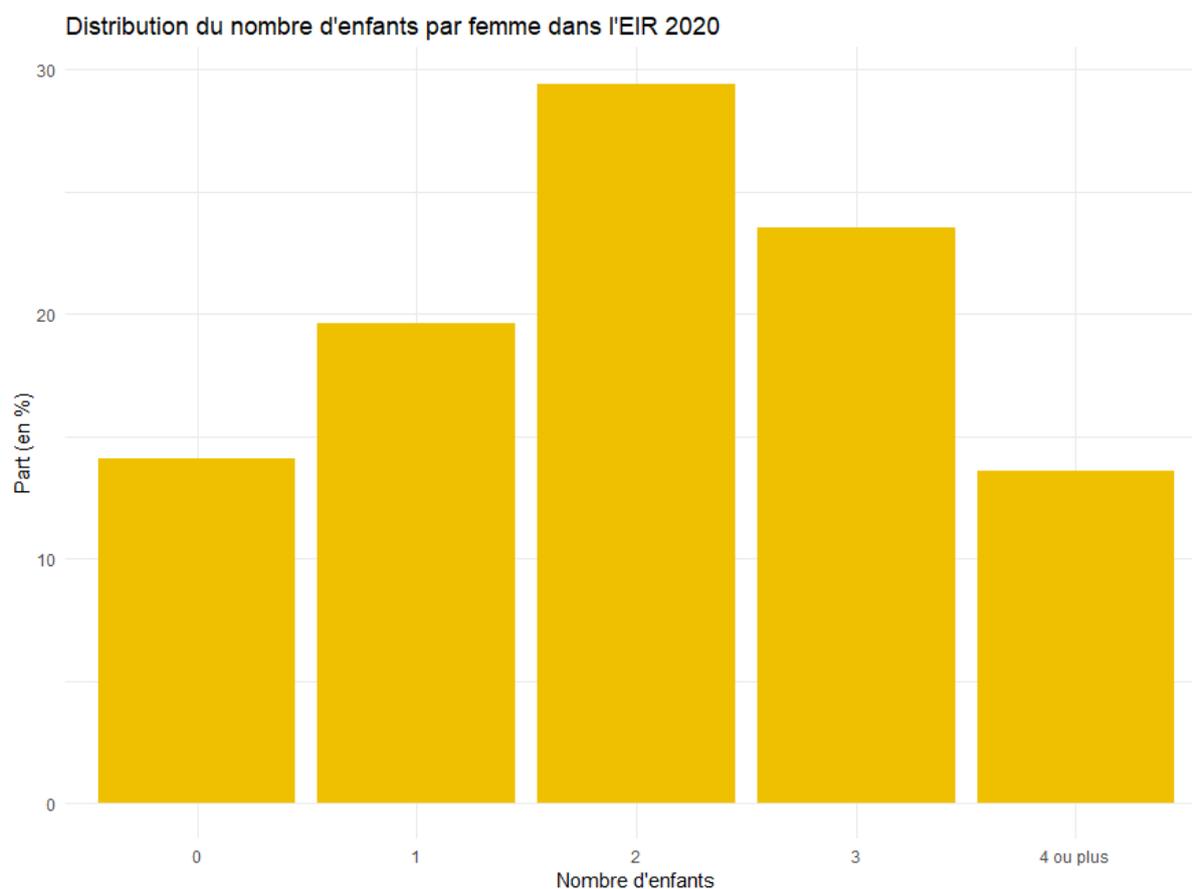


Figure 13 • Distribution du nombre d'enfants par femme dans l'EIR 2020

On peut également comparer la distribution à ce qui était observé dans l'enquête famille et logement (EFL) de 2011 de l'Insee. Le champ des deux sources n'est pas totalement comparable. En effet, l'EFL porte sur les ménages de France métropolitaine, résidant en logement ordinaire. L'EIR inclut lui l'ensemble des personnes bénéficiant d'une pension de retraite, résidant en France ou à l'étranger.

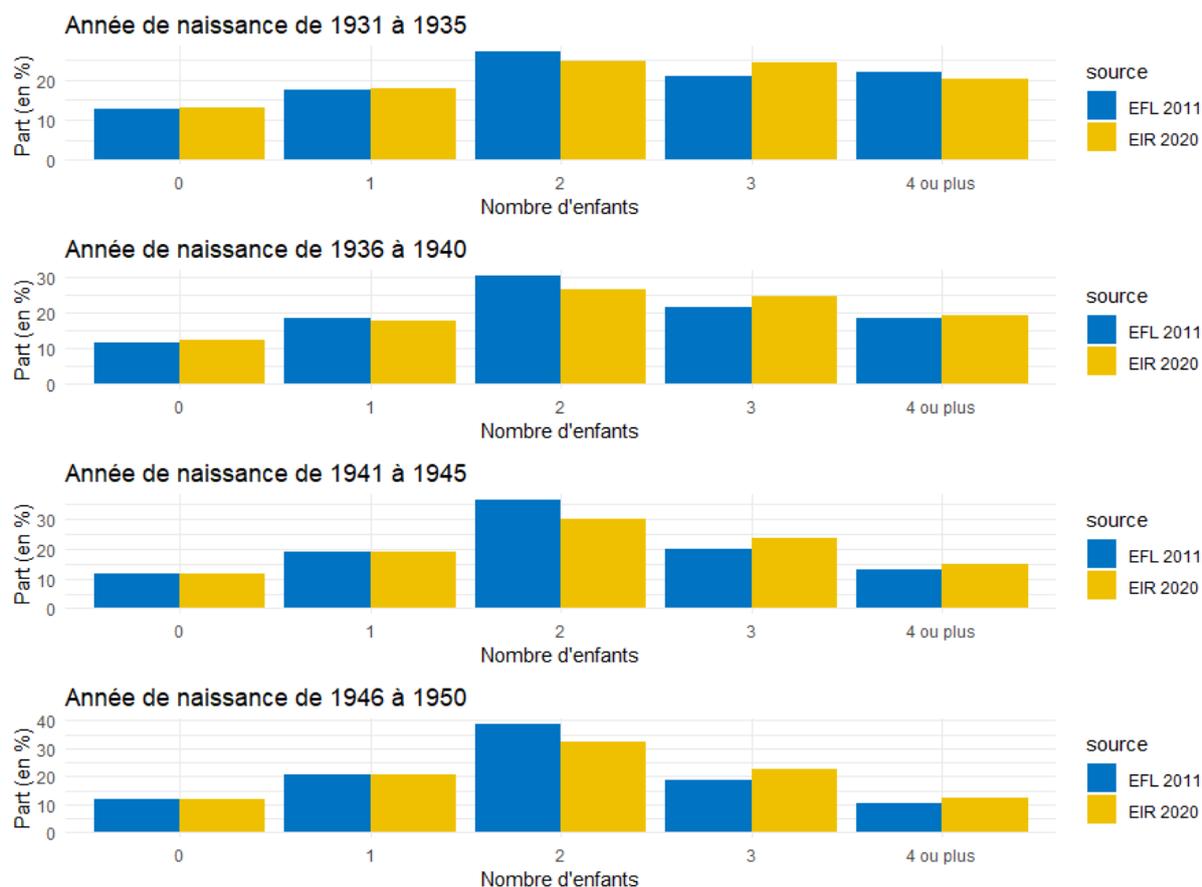


Figure 13 • Distribution du nombre d'enfants par femme dans l'EIR 2020 et dans l'enquête famille et logement de 2011

■ CRÉATION DE NOUVELLES VARIABLES ET DES BASES FINALES AVANT, INDIV ET DDIR

Pour compléter les informations renseignées par les caisses de retraite, la Drees crée des variables complémentaires et les ajoute aux bases finales. Elles sont listées dans les parties « Les variables calculées » du dictionnaire des variables. Cette partie consiste à présenter 3 bases finales et à expliciter ces variables « calculées ».

Description des tables

À l'issue des différentes étapes décrites précédemment, 3 bases de données sont créées, dont l'ensemble constitue l'EIR 2020. Les dessins de fichier de ces bases sont présentés dans un classeur Excel en annexe.

Les 3 tables sont :

- **AVANT20** qui décrit tous les avantages perçus en rente, de droit direct ou dérivé, par les retraités en fonction de la caisse émettrice de la pension. Il y a donc une ligne par individu x régime. Ne figure pas dans cette base :
 - les prestations d'invalidité versées aux individus n'ayant pas l'âge d'être à la retraite (qui figurent dans la table INVALIDITE20)
 - les VFU
 - les droits suspendus

La table AVANT20 compte 3 543 053 observations, qui concernent 1 449 953 individus, et 273 variables. Parmi ces individus, 5 412 personnes qui n'ont ni droit direct ni droit dérivé mais perçoivent un minimum social du SASPA.

- **INDIV20**, recense pour chaque individu l'ensemble de ses droits à pension. Pour constituer cette base, nous avons sommé l'ensemble des droits perçus en rente pour un individu donné à partir de la base AVANT20. Dans INDIV20, une observation correspond donc à un individu.

Au total elle contient 1 449 953 observations/individus et 94 variables.

- **DDIR20**, recense, pour chaque individu percevant au moins une pension de droit direct dans un régime de base, l'ensemble de ses droits à pension (total de ses avantages de droit direct, droit dérivé, etc.). Cette table est construite à partir d'INDIV20 et comprend donc également une seule observation par individu.

Au total, elle contient 1 370 006 observations/individus et 162 variables.

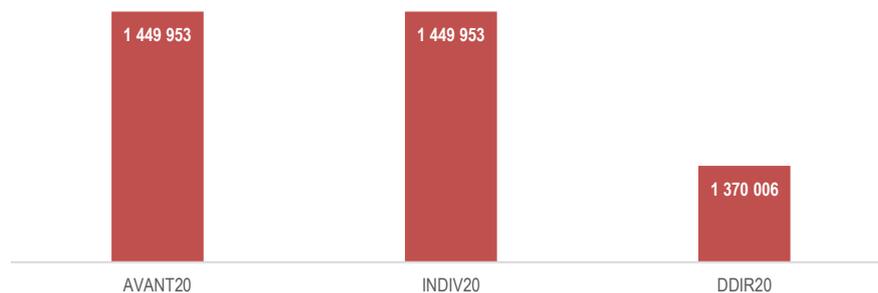


Figure 12 • Nombre d'individus dans les bases AVANT20, INDIV20 et DDIR20

Calculs des âges et des tranches d'âges quinquennaux : les mois de naissance inconnus (variable AGEQUIN_AN)

La variable AGEQUIN_AN est la classe d'âge quinquennal calculée via les variables AN et MN. Elle n'est pas calculée via les variables de l'Insee, AI et MNI car il y a davantage de valeurs manquantes notamment pour la variable MNI.

Détermination du régime principal d'affiliation (variable CC1)

Dans les bases finales INDIV et DDRI, quand un individu dispose de plusieurs droits directs issus de plusieurs régimes différents, le régime principal (variable **CC1**) est défini comme étant le régime de base dans lequel l'individu a validé le plus de trimestres pour un droit direct (TDD1/TDD1E). En cas d'égalité, la taille du régime prévaut (CNAV, MSA salariés, FPE civil, CNRACL...)

Les informations individuelles sur le sexe (variable SEXE), le mois de naissance (variable MN), le lieu de résidence (variables RESID et DR) et le statut matrimonial (variable SM) sont celles du régime principal et sont harmonisés pour les autres régimes.

Variable synthétique d'affiliation (variable POPREV)

La variable **POPREV** (des bases INDIV et DDIR) classe les individus percevant un droit direct dans un régime de base selon leur(s) régime(s) d'affiliation. Pour les polypensionnés, on choisit le régime dans lequel la durée d'assurance (TDD1) dépasse la moitié de la durée d'assurance totale (somme des TDD1 des régimes de base)⁹.

Pour les cas où l'individu aurait la même durée d'assurance dans deux régimes d'affiliation, l'ordre de priorité donné est le suivant : '0010-CNAV', '0012-SRE civil', '0013-SRE militaire', '0032-CNRACL', '0021-MSA salariés', '0022-MSA non-salariés', '0040-RSI commerçants', '0050-RSI artisans', 'Professions libérales', 'régimes spéciaux', 'autres régimes (CANSSM et Cavimac)'.

Un retraité touchant un droit direct dans au moins trois régimes de base est codé POPREV='AUTRES' si aucun de ses régimes d'affiliation ne représente plus de la moitié de la durée d'assurance totale.

Pour les unipensionnés, la variable POPREV est construite de la même façon que pour l'EIR 2016, (100 signifiant 100 % du total des trimestres validés dans les régimes – variable TOTTRIM).

■ Un seul régime de base :

- 100_10 : Salariés du secteur privé (Régime Général)
- 100_12 : Fonctionnaires civils d'État
- 100_13 : Fonctionnaires militaires d'État
- 100_21 : Salariés agricoles (MSA)
- 100_22 : Exploitants agricoles (MSA)
- 100_32 : Fonctionnaires CNRACL
- 100_42 : Sécurité Sociale des Indépendants (ex-Régime Social des Indépendants) (SSI)
- 100_PL : Professions libérales
- 100_SP : régimes spéciaux (SNCF, RATP, ENIM, CRPCEN, Banque de France, IEG, FSPOEIE, Altadis ex Seita, RAVGDT, RETREP équivalent base)
- 100_AU : autres régimes (CANSSM, CAVIMAC)

⁹ Dans les cas où TDD1 n'est renseigné pour aucun régime de base, on prend le régime où le montant de pension de droit direct (M1) est le plus élevé.

■ Plusieurs régimes de base :

- 50_10 : Salariés du secteur privé (Régime Général)
- 50_12 : Fonctionnaires civils d'État
- 50_13 : Fonctionnaires militaires d'État
- 50_21 : Salariés agricoles (MSA)
- 50_22 : Exploitants agricoles (MSA)
- 50_32 : Fonctionnaires CNRACL
- 50_42 : Sécurité Sociale des Indépendants (ex-Régime Social des Indépendants) (SSI)
- 50_PL : Professions libérales
- 50_SP : régimes spéciaux
- 50_AU : autres régimes (CANSSM, CAVIMAC)
- AUTRE : autres (3 régimes de base au moins, dont aucun ne représente plus de la moitié de la durée d'assurance totale).

Variables indicatrices sur la « carrière complète » (variables CAR-COMP_XX)

Dans les bases INDIV et DDIR, 2 variables correspondent à la notion de « carrière complète » :

- CARCOMP_TAUX : carrière complète par rapport à la durée d'assurance requise tous régimes pour accéder au taux plein
- CARCOMP_PRORAT : carrières complète par rapport à la durée de service tous régimes requise par une proratisation à 100 %.

Les carrières complètes au sens du taux plein ou du coefficient ne correspondent pas exactement, pour trois raisons :

- 1) les durées de référence pour le taux de la pension et pour le coefficient de proratisation ne sont pas identiques pour toutes les générations ;
- 2) certaines durées sont comptabilisées uniquement pour la durée tous régimes (qui sert à définir la carrière complète selon la durée requise pour le taux plein) mais pas dans la durée au sein du régime (qui sert à définir la carrière complète selon le coefficient de proratisation). Par exemple, les trimestres de majoration de durée d'assurance (MDA) dans la fonction publique sont utilisés pour le calcul de la durée tous régimes (et donc du taux de la pension), mais pas pour le calcul du coefficient de proratisation.
- 3) la durée utilisée pour le calcul du taux de la pension (durée tous régimes) plafonne la durée validée chaque année dans l'ensemble des régimes à 4 trimestres, alors que ce n'est pas le cas pour le calcul du taux de proratisation.

Variables de durée et d'exhaustivité des carrières

L'EIR comprend de nombreuses variables sur la durée, puisqu'il s'agit d'un paramètre pour le calcul des droits. La durée validée tous régimes détermine en effet le taux de la pension et l'éligibilité au minimum contributif, tandis que la durée validée dans chaque régime détermine le coefficient de proratisation.

La variable DUREEVALIDEE

Dans les bases INDIV et DDIR, la variable DUREEVALIDEE propose une durée d'assurance tous régimes validée par le retraité, permettant de trancher entre l'ensemble des variables de durée renseignées par les caisses. Cette variable est construite « arbitrairement » et ne peut être considérée comme une durée réelle ou effective, mais elle donne une indication. Elle est construite selon les raisonnements suivants :

Dans la plupart des cas, DUREEVALIDEE est égale à DASmax (pour privilégier la durée validée au sens tous régimes) sauf :

- Si NBDDIR>1 (polypensionné) et ENTDDAAmin >= 1983 et (DUREEREGmax – DUREEVALIDEE) > =2 alors DUREEVALIDEE = DUREEREGmax (le DASmax est inférieur de 2 trimestres à DUREEREGmax, signifiant que le DASmax est forcément sous-estimé, on le contraint à DUREEREGmax minimum) ;
- Si aucune DAS n'a été renseignée par les régimes de base (DASMAX =.), alors DUREEVALIDEE = DUREEREG (le maximum entre TDD1 et DASREGIME du régime) ;
- Si le premier droit direct a été liquidé avant 1983 (ENTDDAAmin < 1983) et que DASmax < TDD1max alors DUREEVALIDEE = DUREEREG car avant 1983, la notion de DAS n'avait pas lieu d'être, donc DAS a pu être mal remplie (sous-estimée) pour ces anciennes liquidations ;
- Si NBDDIR=1 (unipensionné) et DASmax<DUREEREGmax alors DUREEVALIDEE=DUREEREG ;
- Enfin, on limite DUREEVALIDEE à 280 trimestres maximum (70 annuités).

La variable DUREECOT

Dans les bases INDIV et DDIR, la durée cotisée correspond au nombre de trimestres validés au seul titre de l'emploi. La durée validée inclut en outre les trimestres assimilés de retraite (chômage, maladie, etc.), les trimestres d'assurance vieillesse des parents au foyer (AVPF) et les majorations de durée d'assurance. La variable DUREECOT est calculée de la façon suivante :

DUREECOT = Durée hors majorations – périodes assimilées soit :

$$\text{DUREECOT} = \text{DUREEVALIDEE} - \text{TDD3A} - \text{TDD2B} - \text{TDD2C}$$

Ce calcul conduit à une durée cotisée négative pour 226 individus, principalement à la Cnav.

À la Cnav, les variables TDD3A, TDD2B et TDD2C proviennent du système d'information « carrière » et non du système d'information « retraites » (contrairement à la DUREEVALIDEE). Autrement dit, les variables TDD3A, TDD2B et TDD2C peuvent continuer à évoluer entre l'année de liquidation et l'année de l'enquête (cumul-emploi retraite etc.).

La variable ECARTDUREE

Dans la base finale DDIR, la variable ECARTDUREE mesure l'écart, en trimestres, entre la durée d'assurance tous régimes (DASmax) et la somme des durées d'assurance intra-régime de tous les régimes de base de l'individu (DUREEREG).

La plupart du temps, cet écart est nul ou très restreint. Il peut toutefois y avoir un écart significatif dans les cas de trimestres effectués à l'étranger ou dans des régimes non présents dans l'EIR 2020 (trimestres non captés dans les TDD1 mais potentiellement repérés par les variables DAS). Ou encore en cas de périodes de pluricotisation : la personne valide des trimestres au sein de régimes différents sur une même période donnée (ex : 4 trimestres validés au RG et 2 trimestres validés à la CNAVPL une même année donne 6 trimestres de TDD1 mais 4 trimestres seulement de DAS).

$$\text{ECARTDUREE} = \text{DASMAX} - \text{DUREEREG}$$

DASETRGmax est la valeur maximale de DASETRG renseignée par un régime de base (durée tous régimes). DASETRG n'a été renseignée que par la CNAV et la MSA. Le problème est que cette variable n'a pas toujours pu être écartée par ces régimes à 4 trimestres par année, il peut donc y avoir des périodes de chevauchement de durées validées à l'étranger.

Par ailleurs, on constate que certains individus ont un nombre élevé de trimestres validés à l'étranger mais qui ne compteraient pas pour le calcul de la DAS. Prenons l'exemple d'un individu unipensionné de la CNAV :

$$\text{DASETRGmax} = 118 ; \text{DASmax} = 143 ; \text{DUREEREG} = 143$$

Pour cet individu, les trimestres à l'étranger n'entrent pas en compte dans le calcul de la DAS. Dans d'autres cas, les trimestres à l'étranger comblent parfaitement l'écart entre la durée validée intra-régime et la durée tous régimes.

$$\text{DASETRGmax} = 72 ; \text{DASmax} = 109 ; \text{DUREEREG} = 37$$

Nous avons créé une seconde variable, DASETRGcret, de façon à ce que les trimestres validés à l'étranger ne puissent être supérieurs au nombre de trimestres d'écart entre la durée validée tous régimes (DASmax) et la somme des durées validées intra-régime de tous les régimes de base de l'individu (DUREEREG). Ainsi, les trimestres validés à l'étranger peuvent expliquer une partie de l'écart constaté, voire la totalité de cet écart, mais pas d'avantage. La variable DASETRGcret pourrait s'interpréter comme la part des trimestres validés à l'étranger comptant pour la durée d'assurance tous régimes.

$$\text{DASETRGcret} = \min(\text{DASETRGmax}, \max(\text{ECARTDUREE}, 0))$$

La variable ECARTDUREE2 reprend ECARTDUREE mais ajoute les trimestres validés à l'étranger écartés.

$$\text{ECARTDUREE2} = \text{ECARTDUREE} - \text{DASETRGcret}$$

DUREEVALIDEEFR est la durée validée (au sens tous régimes) au sein des régimes français :

$$\text{DUREEVALIDEEFR} = \text{DUREEVALIDEE} - \text{DASETRGcret}$$

Il reste néanmoins 941 individus pour lesquels la variable DUREEVALIDEE demeure non renseignée. Il s'agit exclusivement d'unipensionnés pour lesquels les variables TDD1/TDD1E, DAS et DASREGIME sont non remplies. 65,8% d'entre eux sont des unipensionnés du RETREP.

La variable EXHAU

Dans la base finale DDIR, la variable EXHAU indique l'exhaustivité de la carrière d'un retraité de droit direct, en comparant sa durée validée tous régimes et la somme de ses durées validées intra-régime de ses régimes de base.

La variable **EXHAU** présente les différents cas possibles (même méthode que pour l'EIR 2016) :

- EXHAU = 0 si ECARTDUREE = 0 et NMISSTDD1 = 0 : la somme des durées intra régime (DUREEREG) est parfaitement cohérente avec la durée tous régimes (DASmax) ;
- Sinon EXHAU = 1 si ECARTDUREE = 0 mais NMISSTDD1 > 0 : cohérence parfaite alors qu'il manque au moins un TDD1 dans un régime de base ;
- Sinon EXHAU = 2 si $1 < \text{ECARTDUREE} < 1$: cohérence parfaite en arrondissant les trimestres ;
- Sinon EXHAU = 3 si aucune DAS n'est renseignée ou aucun DUREEREG : pas assez d'élément pour calculer un écart ;
- Sinon EXHAU = 4 si NMISSTDD1 > 0 et ECARTDUREE2 > 0 : l'écart est sans doute le fait qu'un TDD1 manque dans au moins un régime de base renseigné ;
- Sinon EXHAU = 5 si DASETRGmax > 0 et ECARTDUREE2 = 0 : l'écart s'explique exactement par les trimestres validés à l'étranger ;
- Sinon EXHAU = 6 si DASETRGmax > 0 et ECARTDUREE > 0 : l'écart peut s'expliquer par des trimestres à l'étranger mais ils sont mal comptabilisés ;
- Sinon EXHAU = 8 si POLYMONOmax = 1 et NBDDIR = 1 et ECARTDUREE2 > 0 et NBVFUDD <= 0 : l'individu est déclaré polycotisant par un régime, mais il est unipensionné au sens de l'EIR ; l'écart peut donc provenir du manque d'un régime, soit qu'il n'a pas encore été liquidé, soit qu'il s'agisse d'un trou de collecte ;
- Sinon EXHAU = 9 si POLYMONOmax = 1 et NBDDIR = 1 et ECARTDUREE2 > 0 et NBVFUDD > 0 : l'individu est unipensionné, et l'écart peut s'expliquer par le fait qu'un régime de base verse une pension en VFU à l'individu et donc que l'on n'ait pas la durée validée au sein de ce régime correspondante ;
- Sinon EXHAU = 10 si POLYMONOmax <= 0 et $0 < \text{ECARTDUREE} < 8$: faibles écarts pour les déclarés unicotisants, pouvant provenir par exemple du fait que DASEQUIV n'a pu être remplie par tous les régimes ou de petites différences de comptabilité des durées d'assurance entre régime ;
- Sinon EXHAU = 11 si $0 < \text{ECARTDUREE} < 8$: idem que cas précédent mais pour les déclarés polycotisants ;
- Sinon EXHAU = 12 si ECARTDUREE2 < 0 et NBDDIR > 1 : écart négatif (DUREEREG > DASmax) pour des polypensionnés, possibilité de chevauchement de carrière ou DASmax sous-estimée dans le cas des vieilles liquidations par exemple ;
- Sinon EXHAU = 13. Par exemple : écart négatif pour les unipensionnés (DAS sous-estimée) ; écart positif significatif (individu polypensionné dans l'EIR mais qui n'a pas encore liquidé un troisième régime de base, prise en compte du temps partiel chez les polypensionnés, etc.)

La variable NONEXHAU

Dans la base finale DDIR, la variable NONEXHAU permet de cibler les retraités pour lesquels on a une forte incohérence entre la durée validée tous régimes et la somme des durées validées intra-régime. C'est-à-dire de repérer les individus pour lesquels les durées validées intra-régime expliquent moins de 90 % de la durée validée tous régimes. L'écart pouvant alors s'expliquer par différentes raisons :

- une durée validée manquante au sein d'un régime présent dans l'EIR 2020,
- un régime de base non encore liquidé, des trimestres validés à l'étranger, un trou de collecte, etc. (cf. la variable EXHAU).

La variable est renseignée comme suit :

Si $ECARTDUREE > 0,1 \times DUREEVALIDEE$ alors NONEXHAU = 1; 0 sinon

Pour les individus ayant NONEXHAU=0, la somme des durées validées intra-régime expliquent pour au moins 90 % de la durée validée tous régimes. On peut donc penser, pour ces individus, que l'EIR 2020 est quasi exhaustif sur les durées de carrière intra et tous régimes (et donc aussi sur les montants correspondants).

Bien sûr, il ne peut pas être certain que chacun des individus considérés comme ayant une carrière « exhaustive » dans l'EIR aient bien l'intégralité de leur carrière reconstituée dans l'EIR. Néanmoins on peut dire que, pour ces individus, toutes les informations fournies par l'EIR 2020 laissent à penser que ces individus ont les bonnes durées validées intra et tous régimes et que l'on a bien l'ensemble des montants de pension correspondants.

Calcul de la pension nette et du taux de prélèvement sociaux (variables PSOC et TXPSOC)

L'EIR 2020 comprend 4 variables « calculées » relatives aux prélèvements sociaux :

- **La variable PSOC** correspond à un montant de prélèvements sociaux dans le régime (base AVANT)
- **La variable PSOCtot** correspond au montant de prélèvements sociaux tous régimes (bases INDIV et DDIR).
- **La variable TXPSOC** correspond au taux de prélèvement sociaux dans le régime (base AVANT)
- **La variable TXPSOCtot** correspond au taux de prélèvement sociaux tous régimes (bases INDIV et DDIR).

Ces variables prennent en compte la contribution sociale généralisée (CSG), la contribution à la réduction de la dette sociale (CRDS), la cotisation à l'assurance maladie¹⁰ et la contribution additionnelle de solidarité pour l'autonomie (Casa) et sont appliquées à la somme de la pension de droit direct, la pension de droit dérivé et les éventuelles majorations de pension pour 3 enfants ou plus.

¹⁰ Cette cotisation est prévue à l'article L. 131-2 du code de la Sécurité sociale, et son taux fait l'objet du D. 242-8. En 2020, les pensions versées par les régimes de base n'y étaient pas soumises ; dans les régimes complémentaires, le taux de cette cotisation était de 1,0 % pour la plupart des régimes. Pour les résidents à l'étranger, ce taux s'élève à 4,2 %.

■ INFORMATIONS COMPLÉMENTAIRES SUR CERTAINES VARIABLES

Variables relatives au minimum de pension

Dans l'EIR 2020, différentes variables relatives au minimum de pension sont disponibles :

- **La variable M1_hors_minimum** : montant mensuel de la pension de droit direct en rente, hors majoration du minimum contributif / garanti
- **La variable INDD** : indicateur de l'application de la règle du minimum et du maximum pour le droit direct
- **La variable MICO** : indicateur de pension portée au minimum contributif (RGA) ou garanti (FP) ou autre minimum de ce genre existant
- **La variable TAUX_AVANTMICO** : taux de liquidation de la pension avec les trimestres de surcote, avant que la pension soit portée ou non au minimum contributif
- **La variable MINDD** : conditions d'attribution et de service du minimum contributif (à partir de 2016) :
- **La variable MAJOMINDD** : surplus de pension lié au minimum de pension de droit direct
- **La variable INDR** : indicateur de l'application de la règle du minimum et du maximum pour le droit dérivé
- **La variable MAJOMINDR** : le montant de surplus de pension lié au minimum de pension de droit dérivé

Taux de liquidation de la pension (TAUX)

La variable TAUX est défini comme le **taux de liquidation de la pension après application de la décote (ou de l'abattement) ou de la surcote**, lorsque celle-ci est effective.

Cependant, les régimes ont calculé un TAUX parfois avant application de la surcote, parfois après.

A noter, la variable n'est pas renseignée pour :

- l'ENIM
- les pensions de coordination (TYPDD=25) à la SNCF
- la CGPCE
- la MSA complémentaire exploitants.

Régimes général et alignés

- A la CNAV (régime général) et aux régimes alignés, le taux de liquidation (TAUX) est calculé avant proratisation.
- Le taux de liquidation maximum, c'est-à-dire le taux appliqué à un individu n'ayant pas de trimestres de surcote ou de décote est fixé à 50 %. A noter, les individus bénéficiant d'une surcote ont également un taux de liquidation égale à 50 %, mais le montant de la pension (M1) sera majoré par rapport au nombre de trimestres supplémentaires. Il est donc possible d'avoir des individus TAUX=50 et TRIMSUR>0 (nombre total de trimestres tous régimes pris en compte dans le calcul de la surcote).
- Le taux minimum est fixé à 25 %, il est appliqué aux retraités qui compte 20 trimestres de décote, soit le nombre maximal possibles de trimestres (un trimestre de décote étant valorisé à 1,25 % pour les liquidations après 2004).
- La **surcote liée à la durée d'assurance tous régimes** est appliquée à partir de l'année 2004. En ce qui concerne la surcote liée à l'âge, c'est-à-dire celle appliquée aux pensionnés partis à la retraite après l'âge d'annulation de la décote, elle est appliquée après 2004.
- Le taux de majoration de la surcote dépend de la législation en cours.

- Pour la CNAV et le SSI, il est possible que la surcote ne soit pas intégrée dans la variable TAUX, pour les pensions liquidées avant le 1^{er} avril 2009. En effet, pour ces pensions, la surcote s'applique au montant de la pension avant comparaison au minimum contributif (cf. Encadré 1).
 - Si le minimum contributif est supérieur à la pension majorée de la surcote, alors le retraité perçoit le minimum contributif, sans majoration liée à la surcote.
 - Ainsi, un retraité ayant liquidé sa pension avant 2009, peut avoir un TRIMSUR>0 (nombre total de trimestres tous régimes pris en compte dans le calcul de la surcote), TAUX = 50 %, s'il est bénéficiaire du minimum contributif (variable INDD). TAUX étant alors un TAUX effectif du calcul de la pension. Pour plus de clarté ont été créées deux variables supplémentaires :
 - 1) **TAUX_SURCOTE** : majoration de TAUX liée à la surcote instaurée en 2004
 - 2) **TAUX_AVANTMICO** : taux de liquidation de la pension après application de la décote ou de la surcote avant prise en compte du minimum contributif (supérieur à 50 % pour tous ceux qui ont TRIMSUR>0, y compris les bénéficiaires du minimum contributif liquidant avant 2009).
- Pour les pensions liquidées après cette date, la surcote est appliquée après le calcul du minimum contributif.

Encadré 2 : Le minimum contributif au régime général

Le montant de la pension de retraite versée à taux plein par le régime général ne peut être inférieur à un montant minimum, appelé « minimum contributif ». Fin 2020, son montant s'élève à 642,93 euros par mois pour une carrière complète dans le régime (702,54 euros avec la majoration). En cas de carrière incomplète, ce montant est versé au prorata de la durée validée pour la retraite. Depuis le 1^{er} janvier 2012, les conditions d'attribution du minimum contributif ont été modifiées. Ce dernier n'est versé qu'aux assurés ayant liquidé l'ensemble de leurs droits à retraite (condition de subsidiarité) et dont le montant de la pension de droit direct tous régimes confondus est inférieur à un seuil, fixé à 1 191,56 euros par mois à partir de janvier 2020 (encadré 1). En cas de dépassement du seuil, le montant du minimum est partiellement ou totalement écarté. Ces nouvelles règles excluent donc du dispositif un certain nombre de retraités, notamment les polypensionnés dont la pension tous régimes dépasse ce montant.

Source : Les retraités et les retraites – Édition 2023

Régimes de la Fonction publique (SRE, CNRACL, FSPOEIE)

- La variable TAUX renseigné est calculé après proratisation.
- La surcote a été appliquée à partir de 2004 et la décote à partir de 2006 seulement. Concernant les liquidations antérieures à 2004, elles sont calculées selon le modèle :

$$taux = 75 \times \frac{TDD1E}{DUREE_PRORAT}$$

Avec :

DUREE_PRORAT : durée de proratisation

- TAUX peut ne pas intégrer les trimestres de décote (TRIMDEC) si ces derniers n'ont pas effectivement minoré la pension. Avant la réforme de 2010, la pension minorée par la décote pouvait être augmentée jusqu'à atteindre le minimum garanti (cf. Variables relatives au minimum de pension, p.41), et les trimestres de décote n'étaient alors pas "effectifs". Depuis 2011, il est impossible de bénéficier du minimum garanti en cas de trimestres de décote.

Régimes de la CNAVPL

- Les régimes de la CNAVPL fonctionnent par points. Le TAUX indiqué dans l'EIR est donc toujours fixé à 100 % si la pension est servie sans abattement ou sans surcote.
- La surcote au sens durée d'assurance tous régimes est possible depuis 2004.
- La décote au sens de la durée d'assurance tous régimes est instaurée depuis 2004 mais auparavant existait un abattement par l'âge (nombre de trimestres manquants pour atteindre 65 ans).

Régimes spéciaux (SNCF, RATP, Banque de France, CRPCEN)

- Le taux de liquidation renseigné tient compte de la proratisation.
- Le taux est de 75 % pour une carrière complète au sein du régime (ou 80 % selon certains cas).
- La surcote n'est possible que pour les liquidations réalisées à partir de juillet 2008 et la décote depuis 2010.

Régimes complémentaires

- Le taux de liquidation dans la plupart des régimes complémentaires est de 100 % en cas de pension sans abattement/majoration.

Nombre de points pour les droits directs (variables du type NPOINDD)

Dans la plupart des cas, les variables NPOINDD comprennent déjà l'abattement dans les cas où le taux de liquidation est inférieur à 100 % (TAUX < 100). Le montant de l'avantage principal de droit direct à la liquidation (antérieure au 31 décembre 2020) est alors $M1 = NPOINDD1 \times \text{valeur du point}$.

Cette formule s'applique dans la plupart des régimes. Néanmoins, pour certaines caisses, notamment parmi les professions libérales, elle connaît des exceptions, car la valeur du point ne dépend pas uniquement de l'année de liquidation, mais également de l'année d'acquisition.

Au RAFP, la variable NPOINDD1 ne tient pas compte du taux de liquidation qui est autour de 1 %.

Majoration de pension pour les parents de 3 enfants ou plus

Dans certains régimes de retraite, les parents ayant élevé 3 enfants ou plus bénéficient d'une majoration de pension.

Au régime général et dans la fonction publique, cette majoration prend la forme d'une bonification de pension, accordée dès la liquidation de la pension, et qui est intégrée à la pension (et acquise définitivement). Cette majoration est différente de la majoration de pension pour personne à charge, qui est une majoration temporaire de la pension accordée lorsque la personne à la retraite a des personnes à charge. Cette dernière est enregistrée dans l'EIR en M9.

Les montants enregistrés dans les variables M1 / MTOT1 (avantage principal de droit direct) et en M2 / MTOT2 (avantage principal de droit dérivé) ne tiennent pas compte de cette majoration. Elle est enregistrée en M5 / MTOT5, qu'elle soit issue d'un droit direct ou dérivé.

Pour distinguer la part du M5 issue d'un droit direct de la part issue d'un droit dérivé, l'usage consiste alors à répartir le montant M5 entre deux composantes :

- $M5 \times \frac{M1}{(M1+M2)}$ est la majoration du droit direct,
- $M5 \times \frac{M2}{(M1+M2)}$ est la majoration du droit dérivé.

C'est cette convention qui est retenue pour les tableaux de la fiche sur le niveau des pensions du panorama annuel *Les retraités et les retraites*.

Variables relatives au minimum vieillesse

L'EIR 2020 contient 5 variables de montant relatives au minimum vieillesse :

- M6, le montant mensuel de l'ASV, attribuée sous conditions de résidence en France ;
- M7, le montant mensuel de l'allocation de 1^{er} étage (ancienne L. 814-2), attribuée sans condition de résidence et donc plébiscitée par les non-résidents en France ;
- M8, le montant mensuel de l'ASPA ou de l'Allocation Supplémentaire d'Invalidité (ASI)

Variables ajoutées en 2020

- M8_ASI, le montant mensuel de l'ASI ;
- M8_ASPA, le montant mensuel de l'ASPA.

Par construction, M8 est exclusif de M6 et M7. M8 et M6 sont en théorie réservés aux assurés résidant en France. En pratique, dans l'EIR 2020, quelques individus qui perçoivent du M6 ou du M8 sont indiqués comme résidant à l'étranger.

Aucune de ces trois variables n'est utilisée pour le calage sur marges. Néanmoins, les effectifs sont globalement très proches de ceux obtenus dans le cadre de l'enquête annuelle de la Drees sur le minimum vieillesse. Avec l'EIR 2020, on estime que 619 697 personnes de 65 ans ou plus et résidant en France sont allocataires du minimum vieillesse en 2020, tandis que l'enquête sur le minimum vieillesse indique un total de 635 300 personnes.

Retraite progressive (TYPDD=26)

La retraite progressive est un dispositif qui permet de lisser la transition entre emploi et retraite : les bénéficiaires réduisent leur quotité de travail et perçoivent une fraction de leur pension de retraite. Quand ils partent définitivement à la retraite, leurs pensions sont recalculées en tenant compte des droits acquis depuis. Le nombre de bénéficiaires de la retraite progressive était encore faible fin 2016 (11 490 personnes au régime général, 720 à la MSA salariés).

Les organismes participants à l'EIR doivent déclarer les bénéficiaires d'une retraite progressive comme retraités, en les identifiant avec la modalité TYPDD=26.

Dans l'EIR 2020, 6 386 personnes sont indiquées en retraite progressive dans au moins un régime. Plus de la moitié ont entre 60 ans et 62 ans au 31 décembre 2020. On note néanmoins que certains ont 70 ans ou plus.

Annexe 1. Liste des codes caisses utilisés – Variables CC et CCGP

Régimes de base		
CC	Acronyme du régime	Signification
0010	CNAV	Caisse Nationale d'Assurance Vieillesse
0010	CNAM	Caisse Nationale d'Assurance Maladie
0012	SRE civil	Service des Retraites de l'État - civil
0013	SRE militaire	Service des Retraites de l'État - militaire
0021	MSA - Salariés agricoles	Mutualité Sociale Agricole - Salariés agricoles
0022	MSA - Non salariés agricoles	Mutualité Sociale Agricole - Non salariés agricoles
0031	MSA - SASPA	Mutualité Sociale Agricole - Service de l'allocation de solidarité aux personnes âgées
0032	CDC - CNRACL	Caisse des Dépôts et Consignation - Caisse Nationale de Retraite des Agents des Collectivités Locales
0033	CDC - FSPOEIE	Caisse des Dépôts et Consignation - Fonds Spécial des Pensions des Ouvriers des Etablissements Industriels de l'État
0042	SSI (ex-RSI)	Sécurité Sociale des Indépendants (ex-Régime Social des Indépendants)
0060	CPRPF	Caisse de Prévoyance et de Retraite du Personnel Ferroviaire (CPRPF)
0070	ENIM	Établissement National des Invalides de la Marine
0080	CDC - CANSSM	Caisse des Dépôts et Consignation - Caisse Autonome Nationale de la Sécurité Sociale dans les Mines
0090	CAVIMAC	Caisse d'Assurance Vieillesse Invalidité et Maladie des Cultes
0100	CNIEG	Caisse Nationale des Industries Electriques et Gazières
0300	CRP RATP	Caisse de Retraite et Prévoyance RATP
0301	CRP RATP - pensions en coordination - régimes de base	Caisse de Retraite et Prévoyance RATP
0500	CRPCEN	Caisse de Retraite et de Prévoyance des Clercs et Employés de Notaires
0510	CRPCEN Alsace-Moselle	Caisse de Retraite et de Prévoyance des Clercs et Employés de Notaires - Alsace-Moselle
0600	Caisse de réserve des employés de la Banque de France	Caisse de réserve des employés de la Banque de France
0700	APC - ALTADIS (ex SEITA)	Association de Prévoyance Collective - Alliance-tabac-distribution (ex-Service d'exploitation industrielle des tabacs et allumettes)
0800	APC - RAVGDT	Association de Prévoyance Collective - Régime d'Allocations Viagères des Gérants de Débits de Tabac
0901	APC - RETREP - équivalent régime de base	Association de Prévoyance Collective - Régime Temporaire de Retraite de l'Enseignement Privé
2011	CPRN	Caisse de prévoyance et de retraite des notaires
2021	Groupe Berri - CAVOM	Caisse d'Assurance Vieillesse des Officiers Ministériels, officiers publics et des compagnies fiduciaires
2031	CARMF	Caisse Autonome de Retraite des Médecins de France
2041	CARCDSF - chirurgiens dentistes	Caisse Autonome de Retraite des Chirurgiens Dentistes et des Sages-Femmes
2051	CAVP	Caisse d'Assurance Vieillesse des Pharmaciens
2061	CARCDSF - sages femmes	Caisse Autonome de Retraite des Chirurgiens Dentistes et des Sages-Femmes
2071	CARPIMKO	Caisse Autonome de Retraite et de Prévoyance des Infirmiers, Masseurs-Kinésithérapeutes, Pédiçures-podologues, Orthophonistes et Orthoptistes
2081	CARPV	Caisse Autonome de Retraite et de Prévoyance des Vétérinaires
2091	Groupe Berri - IRCEC (*)	Institution de Retraite Complémentaire de l'Enseignement et de la Création
2101	CAVAMAC	Caisse d'Allocation Vieillesse des Agents généraux et des Mandataires non salariés de l'Assurance et de la Capitalisation
2111	Groupe Berri - CIPAV	Caisse Interprofessionnelle de Prévoyance et d'Assurance Vieillesse
2121	Groupe Berri - CAVEC	Caisse d'Assurance Vieillesse des Experts-Comptable
2211	CNBF	Caisse Nationale des Barreaux Français

2221	CNBF - ex CAVOM	Caisse Nationale des Barreaux Français - ex Caisse d'Assurance Vieillesse des Officiers Ministériels
2231	CNBF - ex CIPAV	Caisse Nationale des Barreaux Français - ex Caisse Interprofessionnelle de Prévoyance et d'Assurance Vieillesse
Régimes complémentaires		
0023	MSA - Non salariés agricoles	Mutualité Sociale Agricole - Non salariés agricoles
0043	SSI (ex-RSI) - Commerçants	Sécurité Sociale des Indépendants (ex-Régime Social des Indépendants) - Commerçants
0302	CRP RATP - pensions en coordination - régimes complémentaires	Caisse de Retraite et Prévoyance RATP
0902	APC - RETREP - équivalent régime complémentaire	Association de Prévoyance Collective - Régime Temporaire de Retraite de l'Enseignement Privé
0911	APC - RAEP - fraction régime de base	Association de Prévoyance Collective - Régime Additionnel de l'Enseignement Privé
0912	APC - RAEP - fraction régime complémentaire	Association de Prévoyance Collective - Régime Additionnel de l'Enseignement Privé
1001	CDC - IRCANTEC - Régime général	Caisse des Dépôts et Consignation - Institution de Retraite Complémentaire des Agents Non-Titulaires de l'Etat et des Collectivités publiques
1002	CDC - IRCANTEC - Régime des élus	Caisse des Dépôts et Consignation - Institution de Retraite Complémentaire des Agents Non-Titulaires de l'Etat et des Collectivités publiques
2014	CPRN avantage complémentaire section "variable B"	Caisse de Prévoyance et de Retraite des Notaires
2015	CPRN avantage complémentaire section "Colmar et Metz"	Caisse de Prévoyance et de Retraite des Notaires
2016	CPRN avantage complémentaire section "variable C" (*)	Caisse de Prévoyance et de Retraite des Notaires
2022	Groupe Berri - CAVOM	Caisse d'Assurance Vieillesse des Officiers Ministériels, officiers publics et des compagnies fiduciaires
2032	CARMF	Caisse Autonome de Retraite des Médecins de France
2042	CARCDSF - Chirurgiens dentistes	Caisse Autonome de Retraite des Chirurgiens Dentistes et des Sages-Femmes
2052	CAVP	Caisse d'Assurance Vieillesse des Pharmaciens
2062	CARCDSF - Sages-femmes	Caisse Autonome de Retraite des Chirurgiens Dentistes et des Sages-Femmes
2072	CARPIMKO	Caisse Autonome de Retraite et de Prévoyance des Infirmiers, Masseurs-Kinésithérapeutes, Pédiatres-podologues, Orthophonistes et Orthoptistes
2082	CARPV	Caisse Autonome de Retraite et de Prévoyance des Vétérinaires
2092	Groupe Berri - IRCEC	Institution de Retraite Complémentaire de l'Enseignement et de la Création
2093	Groupe Berri - IRCEC RAEL ou RACD **	Institution de Retraite Complémentaire de l'Enseignement et de la Création
2102	CAVAMAC	Caisse d'Allocation Vieillesse des Agents généraux et des Mandataires non salariés de l'Assurance et de la Capitalisation
2112	Groupe Berri - CIPAV	Caisse Interprofessionnelle de Prévoyance et d'Assurance Vieillesse
2122	Groupe Berri - CAVEC	Caisse d'Assurance Vieillesse des Experts-Comptable
2212	CNBF	Caisse Nationale des Barreaux Français
2232	CNBF - ex CIPAV	Caisse Nationale des Barreaux Français - ex Caisse Interprofessionnelle de Prévoyance et d'Assurance Vieillesse
3000	CDC - ERAFP	Caisse des Dépôts et Consignation - Établissement de Retraite Additionnelle de la Fonction Publique
4000	CRPN - Retraite des navigants	Caisse de Retraite du Personnel Navigant professionnel de l'aéronautique civile - Retraite des navigants
4500	CGPCE	Caisse Générale de Prévoyance des Caisses d'Épargne
5600	AGIRC - ARRCO	Association Générale des Institutions de Retraite des Cadres Association des Régimes de Retraite Complémentaire
Régimes supplémentaires		
2033	CARMF	Caisse Autonome de Retraite des Médecins de France
2043	CARCDSF - régimes des praticiens conventionnés - Chirurgiens dentistes	Caisse Autonome de Retraite des Chirurgiens Dentistes et des Sages-Femmes
2053	CAVP - régimes des praticiens conventionnés	Caisse d'Assurance Vieillesse des Pharmaciens
2063	CARCDSF - régimes des praticiens conventionnés - Sages-femmes	Caisse Autonome de Retraite des Chirurgiens Dentistes et des Sages-Femmes
2073	CARPIMKO - régimes des praticiens conventionnés	Caisse Autonome de Retraite et de Prévoyance des Infirmiers, Masseurs-Kinésithérapeutes, Pédiatres-podologues, Orthophonistes et Orthoptistes

Annexe 2. Les appariements avec l'EIR

L'EIR est apparié avec d'autres sources de données produites par l'Insee. En effet, tous les régimes de retraite ne bénéficient pas nécessairement d'un système d'informations aussi exhaustif. Ces appariements permettent d'enrichir les informations disponibles et transmises par les différents régimes de retraite français.

Appariement avec le Panel tous salariés et l'Échantillon démographique permanent

Cf. **Erreur ! Source du renvoi introuvable.**, p.**Erreur ! Signet non défini.**

Appariement avec les données fiscales

Un appariement de l'EIR avec les données fiscales a été réalisé pour la première fois avec les données de l'EIR 2012 sur quelques régimes de retraite seulement. L'opération a été généralisée pour l'EIR 2016 et s'est faite en parallèle de la collecte de ce dernier. L'appariement a été réalisé par l'Insee, à l'aide de données identifiantes (nom, prénom, adresse) transmises par les régimes de retraite à l'Institut. Cette opération s'est révélée très lourde en gestion, pour les trois types d'organismes intervenant : les régimes de retraite, l'Insee et la Drees.

Une fois appariées avec les données de l'EIR, les données fiscales permettent d'étudier les revenus du foyer au-delà des pensions de retraites présentes dans l'EIR. Les variables concernées portent sur la structure du foyer et les dates de naissance de ses membres ; les informations relatives au revenu des personnes présentes dans l'échantillon interrégimes de retraités ainsi qu'au revenu de leur foyer. Ces informations fournissent une vue d'ensemble des revenus du ménage, au-delà des montants individuels des pensions. Ces données permettent notamment de mieux connaître le niveau de vie des retraités, d'approfondir le lien entre pension de retraite et autres revenus, et de mieux appréhender le non-recours au minimum vieillesse¹¹.

Dans un contexte de moyens contraints, et de développement des solutions d'appariement, une solution alternative, plus économique, a été réalisée pour l'EIR 2020. Les informations fiscales ont été obtenues grâce à un appariement avec l'EDP qui les contient depuis 2011, pour les personnes présentes dans les deux échantillons. Enfin, un appariement avec FIDELI sera réalisé ce qui permettra un meilleur taux d'appariement.

¹¹ [Pauline Meinzel \(DREES\) \(2022, mai\). Le non-recours au minimum vieillesse des personnes seules. Les dossiers de la DREES, 97.](#)

Annexe 3. Dessins de fichiers des tables de l'EIR 2020 et nomenclature de certaines variables

Les tables de l'EIR 2020 ayant un nombre important de variables, les dessins de fichiers ne sont pas présentés dans ce document mais sont joints dans un fichier Excel. Les nomenclatures utilisées sont également détaillées dans ce fichier Excel.