

# **ENC EHPAD 2016**

## **Note de fin de campagne**

**Mars 2019**

## *Introduction*

La Direction générale de la cohésion sociale (DGCS) et la Caisse nationale de solidarité pour l'autonomie (CNSA) ont souhaité approfondir et actualiser la connaissance des coûts supportés par les établissements d'hébergement pour personnes âgées dépendantes (EHPAD).

La réalisation de deux enquêtes de coûts, sur les données comptables 2012 et 2013, avait été confiée à l'Agence technique de l'information sur l'hospitalisation (ATIH), en raison de son expertise dans le recueil et l'analyse de données médico-économiques.

Dans la continuité de ces enquêtes, deux études nationales de coûts (ENC), financées par la CNSA, ont été lancées sur les données comptables 2015 puis 2016. Ces études sont plus détaillées que les enquêtes de coûts puisque certaines données sont recueillies au niveau de chaque résident par les établissements participants.

L'ensemble de la méthodologie de l'ENC a été défini lors de comités techniques constitués, entre autres, des fédérations et associations représentatives du secteur.

Les éléments concernant l'ENC EHPAD 2016 sont disponibles sur le site Internet de l'ATIH<sup>1</sup>. Les établissements participants ont été sélectionnés au second semestre 2015 ; ce sont des établissements volontaires, qui ont bénéficié à chaque étape de l'aide d'un superviseur. Ils ont transmis leurs données en deux phases : leurs données d'activité (observées sur une semaine) pour chaque trimestre en 2016 puis leurs données comptables en 2017.

Enfin, cette ENC a la particularité d'intégrer les charges de soins de ville grâce aux informations transmises par la Caisse nationale de l'assurance maladie (CNAM). Ces charges sont issues du croisement de l'outil RESID-EHPAD et des soins remboursés par l'Assurance maladie.

A l'issue de la collecte, pour chaque résident, est calculé le coût de sa prise en charge par trimestre. Par ailleurs, des groupes ont été constitués par des experts (médecins des établissements, des fédérations, de la CNSA et de l'ATIH) qui se sont réunis en 2016. Leurs travaux ont permis de définir 83 groupes homogènes en termes d'état de santé et de perte d'autonomie<sup>2</sup>. Les coûts issus du recueil des données 2016 sont calculés par groupe de résidents.

Ce document décrit le déroulement de la campagne de l'ENC 2016, l'échantillon ainsi que la méthodologie de calcul des coûts<sup>3</sup> qui a été mise en œuvre.

---

<sup>1</sup> <http://www.atih.sante.fr/> dans la rubrique : Domaines d'activités / Information sur les coûts / ENC EHPAD / Documentation / ENC EHPAD Données 2016 (<https://www.atih.sante.fr/enc-ehpad-donnees-2016>)

<sup>2</sup> L'annexe 1 est consacrée à la constitution de ces groupes de résidents. La liste des groupes est disponible dans le fichier Excel des coûts de référence publié en juillet 2018 sur le site de statistiques de l'ATIH : <https://www.scansante.fr/couts-de-referance-enc-ehpad-2016>. Une note de lecture est jointe à ce fichier pour en faciliter la prise en main

<sup>3</sup> Les coûts redressés sont disponibles dans le fichier Excel des coûts de référence publié en juillet 2018 sur le site de statistiques de l'ATIH : <https://www.scansante.fr/couts-de-referance-enc-ehpad-2016>. Une note de lecture est jointe à ce fichier pour en faciliter la prise en main

## Sommaire

<b>1. Déroulement de la campagne pour l'ENC 2016</b>	<b>4</b>
1.1 LES EHPAD PARTICIPANT A L'ETUDE	4
1.2 DONNEES A RECUEILLIR PAR LES EHPAD	4
<b>2. Nouveautés sur la méthodologie de l'ENC 2016</b>	<b>4</b>
<b>3. Description de l'échantillon ENC 2016 (après contrôles qualité)</b>	<b>5</b>
3.1 L'ECHANTILLON SELON LES OPTIONS TARIFAIRES	5
3.2 L'ECHANTILLON SELON LES STATUTS JURIDIQUES	7
3.3 L'ECHANTILLON SELON LA TAILLE DES ETABLISSEMENTS	8
3.4 L'ECHANTILLON SELON LES GROUPES DE RESIDENTS	9
<b>4. Méthodologie de calcul des coûts</b>	<b>10</b>
4.1 PERIMETRE DES COUTS	10
4.2 CONTROLES QUALITE	11
4.3 REDRESSEMENT DES COUTS	11
4.4 LA QUALITE DE L'ESTIMATION	13
<b>Annexes</b>	<b>15</b>
ANNEXE 1 : METHODOLOGIE DE CONSTRUCTION DES GROUPES DE RESIDENTS	16
ANNEXE 2 : CALENDRIER DE LA CAMPAGNE ENC 2016	19
ANNEXE 3 : LISTE DES 12 PROFILS DE SOINS, DES 50 ETATS PATHOLOGIQUES ET DES DOMAINES D'ETATS PATHOLOGIQUES	20
ANNEXE 4 : VARIABLES UTILISEES POUR LE REDRESSEMENT DE L'ECHANTILLON	23
ANNEXE 5 : QU'EST-CE QUE L'ERE ?	25

## 1. Déroulement de la campagne pour l'ENC 2016

### 1.1 Les EHPAD participant à l'étude

L'étude nationale de coûts repose sur un échantillon d'EHPAD sélectionnés par l'ATIH, volontaires et représentatifs de la diversité du secteur. Les établissements, formés par l'ATIH, sont accompagnés par un superviseur dédié.

En contrepartie des moyens matériels et humains mobilisés, les EHPAD sélectionnés bénéficient d'une compensation financière. Son versement est conditionné par la signature d'une convention engageant l'EHPAD, notamment en termes de qualité des données recueillies.

Pour la campagne 2016, 82 EHPAD se sont portés volontaires. 9 établissements se sont désistés avant le début du recueil des données, et 4 établissements ont abandonné en cours de campagne. Ainsi, 69 EHPAD ont transmis l'intégralité des données à recueillir.

### 1.2 Données à recueillir par les EHPAD

L'EHPAD recueille des données d'activité (PATHOS, AGGIR, variables complémentaires et minutes). L'EHPAD doit réaliser quatre recueils de données d'activité (observées sur une semaine) pour chaque trimestre de l'année afin de tenir compte de l'évolution de l'état de santé de ses résidents. Le premier recueil est réalisé au cours du premier trimestre et la coupe PATHOS est validée par le médecin de l'ARS. Ensuite, l'établissement actualise ses données d'activité à chaque nouveau trimestre. L'année suivante, l'établissement retraite ses données comptables. Les recueils de données sont transmis à l'ATIH à partir des outils informatiques mis à disposition des EHPAD. Le calendrier de la campagne ENC est disponible en annexe 2.

A l'issue de la collecte, pour chaque résident, est calculé le coût de sa prise en charge par trimestre. Par ailleurs, des groupes ont été constitués par des experts (médecins des établissements, des fédérations, de la CNSA et de l'ATIH) qui se sont réunis en 2016. Leurs travaux ont permis de définir 83 groupes homogènes en termes d'état de santé et de perte d'autonomie. A partir de ces éléments, le coût moyen d'une journée d'accompagnement en EHPAD, en fonction du profil des résidents accueillis, est calculé.

Parmi les 69 participants, 2 n'ont pas transmis des données conformes à celles attendues dans le guide méthodologique<sup>4</sup> de l'ENC. C'est pourquoi ils ne sont pas intégrés dans les calculs des coûts de référence par groupes de résidents. Finalement il y a 67 EHPAD intégrés dans le calcul des coûts de référence 2016. Par ailleurs, parmi les 68 EHPAD intégrés lors de l'ENC 2015, 30 sont intégrés aux coûts de référence 2016.

## 2. Nouveautés sur la méthodologie de l'ENC 2016

Il y a une stabilité des méthodes entre l'ENC 2015 et l'ENC 2016.

---

<sup>4</sup> Le guide est disponible sous <https://www.atih.sante.fr/enc-ehpad-donnees-2016>.

### 3. Description de l'échantillon ENC 2016 (après contrôles qualité)

Au sein des 67 établissements intégrés, 19 998 observations ont pu être utilisées<sup>5</sup> dans le calcul des coûts moyens redressés. Ces 19 998 observations concernent 5 946 résidents hébergés dans un des 67 EHPAD. Les 4 coupes réalisées par les établissements (une par trimestre) permettent d'avoir jusqu'à 4 fois les informations pour chaque résident de l'établissement. 66% des résidents de l'ENC 2016 sont présents lors des 4 coupes.

L'observation est le résident-trimestre.

L'unité de restitution des coûts moyens redressés est le groupe de résidents.

L'ENC 2016 porte sur les EHPAD **tarifés au GMPS et ayant une coupe Pathos validée par l'ARS en 2013 ou 2014 ou 2015 ou 2016**, soit 3 696 établissements distincts. Ces EHPAD constituent la population de référence. Ils hébergent 286 184 résidents.

La comparaison des répartitions de l'échantillon et de la population de référence permet d'analyser les similarités de ces deux populations. Un échantillon idéal aurait les mêmes proportions que celles observées pour la population de référence pour toutes les variables importantes. Cependant, il n'est pas nécessaire que l'échantillon ait strictement les mêmes caractéristiques que la population de référence puisqu'un processus de redressement<sup>6</sup> des coûts est réalisé.

De plus, l'étude d'un échantillon intègre l'analyse des taux de sondage. Ce taux est défini comme la part que représente l'échantillon dans la population de référence. Par exemple, l'échantillon des 67 établissements participants à l'étude nationale de coûts EHPAD 2016, représente 1,8% des établissements de la population de référence. Ce taux est calculé en faisant le rapport de 67 sur 3 696 puis en multipliant par 100 pour obtenir la part en %. Différents taux de sondage peuvent donc être calculés : en nombre d'établissements, en nombre de résidents, sur l'échantillon entier ou sur une partie de l'échantillon (comme les EHPAD sous tarif global avec PUI ou encore les EHPAD privés commerciaux, etc.).

#### 3.1 L'échantillon selon les options tarifaires

L'option tarifaire d'un EHPAD définit le mode de financement de la prise en charge des soins. Ainsi, les EHPAD avec pharmacie à usage intérieur (PUI) prennent en charge les médicaments consommés par le résident. Les EHPAD financés au tarif global prennent en charge les soins de médecins généralistes, auxiliaires médicaux libéraux et actes courants de biologie et d'imagerie.

---

<sup>5</sup> Certaines observations sont écartées du calcul du coût redressé. Se reporter à la partie 4.2 *Contrôles qualité*

<sup>6</sup> Le redressement est expliqué dans la partie 4.3

**T 1 I Répartition des EHPAD de l'échantillon et de la population de référence selon les options tarifaires**

<b>Option tarifaire</b>	<b>Echantillon</b>		<b>Population de référence</b>	
	<b>Nb EHPAD</b>	<b>%</b>	<b>Nb EHPAD</b>	<b>%</b>
Global avec PUI	14	20,9	559	15,1
Global sans PUI	9	13,4	408	11,1
Partiel avec PUI	3	4,5	70	1,9
Partiel sans PUI	41	61,2	2 659	71,9
<b>Total</b>	<b>67</b>	<b>100,0</b>	<b>3 696</b>	<b>100,0</b>

La part des EHPAD sous tarification partielle et n'ayant pas de pharmacie à usage intérieur (PUI) est plus petite dans l'échantillon que dans la population de référence. Les autres établissements sont, en proportion, plus nombreux dans l'échantillon que dans la population de référence. L'échantillon présente donc pour toutes les options tarifaires une distorsion plus ou moins importante par rapport à la population de référence.

**T 2 I Taux de sondage en nombre d'établissements ou en nombre de résidents selon les options tarifaires**

<b>Option tarifaire</b>	<b>ENC 2016</b>		<b>ENC 2015</b>	
	<b>En nombre d'établissements</b>	<b>En nombre de résidents</b>	<b>En nombre d'établissements</b>	<b>En nombre de résidents</b>
Global avec PUI	2,5%	3,1%	3,4%	4,0%
Global sans PUI	2,2%	2,7%	1,9%	2,6%
Partiel avec PUI	4,2%	5,2%	1,4%	1,5%
Partiel sans PUI	1,5%	2,2%	2,0%	3,0%
<b>Total</b>	<b>1,8%</b>	<b>2,5%</b>	<b>2,2%</b>	<b>3,1%</b>

Au global, l'échantillon 2016 contient 1,8% des établissements de la population de référence et 2,5%<sup>7</sup> des résidents de la population de référence.

Pour toutes les options tarifaires, les taux de sondage sont légèrement plus élevés lorsqu'ils sont calculés en nombre de résidents.

Pour les EHPAD au tarif partiel avec PUI, les taux de sondage sont plus élevés en 2016, avec plus de 5% des résidents présents dans l'échantillon.

En revanche, pour les établissements au tarif partiel sans PUI, les taux de sondage sont plus bas en 2016, avec 2,2% des résidents présents dans l'échantillon.

<sup>7</sup> Cela peut sembler faible mais grâce au recueil effectué à chaque trimestre, l'information disponible pour chaque résident peut être multipliée jusqu'à 4 fois.

Les taux de sondage ont significativement évolué entre 2015 et 2016, surtout pour les établissements au tarif partiel. Plusieurs phénomènes combinés peuvent expliquer ces évolutions. D'une part, la population de référence a augmenté entre 2015 et 2016 (il y a une année de coupe Pathos supplémentaire pour 2016<sup>8</sup>) tandis que le nombre d'établissements intégrés a légèrement diminué (passant de 68 en 2015 à 67 en 2016). D'autre part, les échantillons 2015 et 2016 sont très différents en termes d'établissements. En effet, il y a seulement 30 EHPAD qui ont participé à l'ENC 2016 et qui avaient déjà participé à l'ENC 2015.

### 3.2 L'échantillon selon les statuts juridiques

#### T 3 I Répartition des EHPAD de l'échantillon et de la population de référence selon les statuts juridiques

<i>Statut juridique</i>	Echantillon		Population de référence	
	Nb EHPAD	%	Nb EHPAD	%
Privé commercial	4	6,0	901	24,4
Privé non lucratif	25	37,3	1 224	33,1
Public non rattaché à un établissement sanitaire	28	41,8	1 055	28,5
Public rattaché à un établissement sanitaire	10	14,9	516	14,0
<b>Total</b>	<b>67</b>	<b>100,0</b>	<b>3 696</b>	<b>100,0</b>

Les établissements privés commerciaux sont très sous-représentés dans l'échantillon avec une différence de 18 points par rapport à la structure nationale. Les établissements publics non rattachés à un établissement sanitaire sont, au contraire, surreprésentés dans l'échantillon avec une différence de 13 points.

La part des établissements privés non lucratifs est similaire dans l'échantillon et dans la population de référence. Il en est de même pour les établissements publics rattachés à un établissement sanitaire.

<sup>8</sup> Pour l'ENC 2015 la population de référence correspond aux EHPAD tarifés au GMPS et ayant une coupe Pathos validée par l'ARS en 2013, 2014 ou 2015 (soit 3 084 EHPAD en 2015 contre 3 696 en 2016).

#### T 4 I Taux de sondage en nombre d'établissements ou en nombre de résidents selon les statuts juridiques

<b>Statut juridique</b>	<b>ENC 2016</b>		<b>ENC 2015</b>	
	<b>En nombre d'EHPAD</b>	<b>En nombre de résidents</b>	<b>En nombre d'EHPAD</b>	<b>En nombre de résidents</b>
Privé commercial	0,4%	1,1%	1,7%	3,0%
Privé non lucratif	2,0%	2,8%	3,4%	4,6%
Public non rattaché à un établissement sanitaire	2,7%	3,6%	1,3%	1,8%
Public rattaché à un établissement sanitaire	1,9%	2,3%	2,2%	2,6%
<b>Total</b>	<b>1,8%</b>	<b>2,5%</b>	<b>2,2%</b>	<b>3,1%</b>

Pour tous les statuts juridiques, les taux de sondage évalués sur le nombre de résidents sont plus élevés que ceux évalués en nombre d'établissements.

Les taux de sondage des EHPAD publics non rattachés à un établissement sanitaire sont plus importants en 2016 avec 3,6% des résidents présents dans l'échantillon.

L'absence de stabilité des taux entre 2015 et 2016 s'explique par la constitution de l'échantillon qui est différente entre 2015 et 2016 (avec seulement 30 EHPAD communs les 2 années et un établissement de moins en 2016) mais aussi par l'augmentation de la population de référence avec l'année 2016 ajoutée (par rapport à 2015).

### 3.3 L'échantillon selon la taille des établissements

#### T 5 I Répartition des EHPAD de l'échantillon et de la population de référence selon la taille<sup>9</sup>

<b>Taille</b>	<b>Echantillon</b>		<b>Population de référence</b>	
	<b>Nb EHPAD</b>	<b>%</b>	<b>Nb EHPAD</b>	<b>%</b>
< 80 places	34	50,8	1 997	54,0
Entre 80 et 100 places	22	32,8	1 027	27,8
> 100 places	11	16,4	672	18,2
<b>Total</b>	<b>67</b>	<b>100,0</b>	<b>3 700</b>	<b>100,0</b>

La part des établissements de petite taille (moins de 80 places) et de grande taille (plus de 100 places) dans l'échantillon sont inférieures à celles mesurées sur la population de référence.

<sup>9</sup> La taille des EHPAD est construite en fonction de leur capacité autorisée en nombre de places issue de la base FINESS (Fichier National des Etablissements Sanitaires et Sociaux).



Les établissements de taille moyenne (entre 80 et 100 places) ont en revanche une part plus élevée dans l'échantillon que dans la population de référence (33% dans l'échantillon contre 28% dans la population de référence).

#### **T 6 I Taux de sondage en nombre d'établissements ou en nombre de résidents selon la taille**

<b>Taille</b>	<b>ENC 2016</b>		<b>ENC 2015</b>	
	<b>En nombre d'établissements</b>	<b>En nombre de résidents</b>	<b>En nombre d'établissements</b>	<b>En nombre de résidents</b>
< 80 places	1,7%	2,5%	2,4%	3,7%
Entre 80 et 100 places	2,1%	2,9%	1,8%	2,7%
> 100 places	1,6%	2,2%	2,3%	3,0%
<b>Total</b>	<b>1,8%</b>	<b>2,5%</b>	<b>2,2%</b>	<b>3,1%</b>

Quelle que soit la taille de l'établissement, les taux de sondage mesurés en nombre de résidents sont plus élevés que ceux calculés en nombre d'établissements.

Pour 2016, les taux de sondage en nombre de résidents se situent entre 2% et 3% quelle que soit la taille de l'établissement.

Les établissements de taille moyenne ont le taux de sondage le plus élevé avec presque 3% des résidents de la population de référence présents dans l'échantillon 2016.

De nouveau, la variabilité des taux de sondage entre 2015 et 2016, s'explique par la différence des échantillons entre 2015 et 2016 (avec seulement 30 EHPAD commun les 2 années et un établissement de moins en 2016) mais aussi par l'augmentation de la population de référence avec l'année 2016 ajoutée (par rapport à 2015).

### **3.4 L'échantillon selon les groupes de résidents**

Les 15 groupes ayant les plus gros effectifs dans la population de référence rassemblent 49% des résidents. De plus, 14 de ces 15 groupes ont aussi les plus gros effectifs dans l'ENC 2016.

## T 7 I Répartition des groupes de résidents selon la classe de taux de sondage

<b>Classe de taux de sondage en nombre de résidents</b>	<b>Nb de groupes</b>	<b>Part dans l'ensemble des groupes</b>	<b>Part (en %) en nb de résidents dans la population de référence</b>
0%-1%	2	2,4%	0,3%
1%-2%	31	37,4%	40,6%
2%-2,5%	31	37,4%	42,0%
2,5%-3%	18	21,6%	17,0%
Au moins 3%	1	1,2%	0,1%
<b>Total</b>	<b>83</b>	<b>100,0</b>	<b>100,0</b>

Presque 75% des groupes (62 sur 83) ont un taux de sondage compris entre 1% et 2,5% et ils rassemblent 83% des résidents de la population de référence.

A noter que 4 groupes ayant des effectifs trop faibles ne pourront pas avoir de coûts moyens redressés<sup>10</sup>. Ce sont les groupes 01, 13, 20 et 71<sup>11</sup>. Cependant ces groupes rassemblent moins de 0,5% des résidents de la population de référence.

## 4. Méthodologie de calcul des coûts

### 4.1 Périmètre des coûts

Les 67 établissements ont transmis la totalité des données avec une qualité jugée satisfaisante. Les coûts moyens redressés sont donc issus de l'ensemble des informations transmises.

Les coûts étudiés sont des coûts d'hébergement en € par journée de prise en charge (présence et absence) et par groupe de résidents.

La méthodologie de l'étude nationale de coûts 2016 consiste à réaliser une répartition de l'ensemble des charges selon les 9 activités suivantes :

- hôtellerie générale,
- restauration,
- blanchisserie,
- soins dispensés,
- accompagnement dans les actes de la vie quotidienne,
- organisation du lien social,

<sup>10</sup> Pour plus de détails, se référer à la partie 4.3 *Redressement des coûts*

<sup>11</sup> Groupe 01. M1 et CH, groupe 13. T1DG ou T2DG, groupe 20. T1R2CH ou T2R2CH et groupe 71. CH et/ou R1 et/ou DG

- gestion générale – management,
- immobilier,
- charges financières.

Notons que les charges financières et les charges de structure immobilière<sup>12</sup> ont également été collectées.

Cependant, redresser l'échantillon selon des variables explicatives de ces charges, telles que, par exemple, le fait d'être propriétaire ou locataire est techniquement impossible. En effet, ces informations sont connues sur l'échantillon mais ne sont pas disponibles sur la population de référence.

Ce type de charges ne peut donc pas être redressé.

L'estimation du coût présenté n'intègre donc pas les charges financières et de structure immobilière. Par conséquent, le coût étudié ne contient qu'une partie de l'activité immobilier, à savoir les énergies et fluides.

Dans les restitutions, l'activité « immobilier » s'entendra hors structure immobilière (c'est-à-dire qu'elle ne contient finalement que les énergies et fluides).

Le coût total est un coût correspondant aux charges issues des données comptables et aux charges de soins de ville récupérées et transmises par la CNAM *via* le croisement de l'outil RESID-EHPAD<sup>13</sup> avec les soins remboursés par l'Assurance maladie.

Les soins de ville inclus dans le périmètre du coût sont : les médicaments, les dépenses de médecin généraliste, masseurs-kinésithérapeutes, pédicures-podologues, orthophonistes, orthoptistes, les dispositifs médicaux, les actes de biologie, les actes d'imagerie n'utilisant pas des équipements matériels lourds.

## 4.2 Contrôles qualité

Des contrôles médicaux et statistiques sur la qualité des données, ont conduit à écarter 3 316 observations (soit 14%) qui ont été jugées non utilisables dans le calcul des coûts de référence par groupe de résident. Finalement les coûts moyens de prise en charge pour chaque groupe de résident sont issus de 19 998 observations.

## 4.3 Redressement des coûts

L'analyse menée dans la partie 3 a montré que les répartitions de l'échantillon (selon l'option tarifaire, le statut juridique ou encore la taille d'établissement) ne correspondent pas parfaitement aux répartitions

---

<sup>12</sup> Activité « Immobilier » : charges d'assurance, d'entretien et réparations sur biens immobiliers, des immobilisations corporelles et incorporelles.

<sup>13</sup> Cet outil permet de connaître l'ensemble des résidents en EHPAD ainsi que leurs mouvements (entrée, sortie provisoire, retour dans l'établissement, sortie définitive). Il s'agit d'un outil déclaratif renseigné par chaque EHPAD, chaque mois. Croisé avec les données de remboursement de l'assurance maladie, il permet de connaître les soins remboursés aux résidents pendant leur période d'hébergement en EHPAD.

observées dans la population de référence<sup>14</sup>. Pour cette raison, il ne serait pas correct d'utiliser, de manière brute, les coûts recueillis par les 67 EHPAD participants.

Pour qu'ils deviennent représentatifs de la population de référence, ces coûts doivent être redressés. Le traitement statistique de ce redressement est appelé calage sur marges. La méthodologie est détaillée dans l'encadré ci-dessous.

## Le calage sur marges : pour quoi faire ?

Le principe du redressement ou calage consiste à tirer profit d'une information auxiliaire (c'est-à-dire en pratique d'un ensemble de variables corrélées au coût complet) disponible dans une base de données nationale afin de modifier les poids de sondage initiaux.

**Pour comprendre la nécessité d'un redressement du coût issu d'un échantillon, un exemple est proposé :**

**Il s'agit de connaître le revenu moyen mensuel d'un français.**

Pour cela un échantillon de 100 individus français, composé de 20 femmes et 80 hommes est constitué. Dans cet échantillon, l'âge moyen des femmes est de 32,8 ans et celui des hommes est de 55,2 ans. Le revenu moyen mensuel pour cet échantillon est de 3 078 €.

Par ailleurs, au niveau national (source INSEE), il y a 48,0% de femmes et 52,0% d'hommes, l'âge moyen des femmes est de 42,6 ans et celui des hommes est de 39,8 ans.

L'échantillon contient donc plus d'hommes plus âgés que la population française nationale.

Or on sait que le lien entre le revenu moyen, le sexe et l'âge est très fort.

Ainsi le revenu moyen d'un français ne sera certainement pas égal à 3 078 € puisque l'échantillon ne respecte pas les proportions de la population nationale concernant le sexe et l'âge. D'ailleurs le revenu moyen mensuel d'un français est estimé à 1 712 € par l'INSEE (donc bien différent de celui observé dans l'échantillon).

Il convient donc de redresser l'échantillon, et la méthode choisie est appelée calage sur marges.

**Parallèle avec l'étude nationale de coûts EHPAD :**

Tout comme dans l'exemple ci-dessus, les répartitions des caractéristiques de l'échantillon des 67 EHPAD participants à l'enquête ne sont pas égales à celles de la population de référence. Il n'est donc pas convenable d'utiliser les données comptables telles quelles pour estimer un coût de prise en charge moyen pour la population de référence. Les données de l'échantillon doivent être redressées.

**Le calage sur marges : comment ça marche ?**

Ce redressement ne modifie en rien les valeurs des coûts recueillis mais **il modifie la contribution de chaque donnée dans le calcul des coûts moyens.**

Le calage est réalisé en deux grandes étapes : la recherche des variables de calage puis le calcul des pondérations.

**PREMIERE ETAPE :** Recherche des variables de calage

Le calage peut être effectué grâce à la connaissance de certaines caractéristiques de la population de référence. Toutes les variables connues sur les 3 696 EHPAD peuvent donc être testées. Cependant, plusieurs contraintes doivent être respectées dans le choix des variables retenues :

---

<sup>14</sup> La population de référence pour 2016 est constituée des EHPAD tarifés au GMPS et ayant une coupe Pathos validée par l'ARS en 2013, 2014, 2015 ou 2016, soit 3 696 établissements distincts.

- ◆ elles doivent être explicatives du coût que l'on cherche à estimer,
- ◆ elles ne doivent pas être corrélées entre elles,
- ◆ elles ne doivent pas être en trop grand nombre (sous peine que le calage échoue).

C'est la régression linéaire qui sert d'outil pour choisir le modèle le plus satisfaisant (celui qui est conservé *in fine*). Il doit satisfaire à quelques exigences :

- ◆ les variables explicatives doivent être significatives, c'est-à-dire que leur effet sur le coût est non nul,
- ◆ la part de la variance du coût expliquée par ces variables doit être la plus élevée possible,
- ◆ les résidus de la régression doivent suivre une loi normale.

Dans l'exemple sur l'estimation du revenu des français, les variables de calage sont le sexe et l'âge. Pour les EHPAD, la recherche des variables de calage s'est faite sur chaque groupe de résidents<sup>15</sup> ayant au moins 30 observations (soit 79 groupes sur 83). La liste des variables retenues est donc différente d'un groupe à un autre. Selon les groupes, de 1 à 11 variables de calage ont été retenues<sup>16</sup>.

### SECONDE ETAPE : Calcul des pondérations

Le but de cette étape est d'obtenir un poids relatif à chaque observation (donc une pondération pour chacune des observations de chaque groupe de résidents). Les poids seront tels que, pour chaque groupe de résidents, les caractéristiques de l'échantillon pondéré seront exactement égales aux caractéristiques de la population de référence. Ce système de poids est trouvé grâce à une macro SAS élaborée par l'INSEE, nommée CALMAR<sup>17</sup> (pour CALage sur MARGes).

Autrement dit, après calage, pour chaque groupe de résidents, l'échantillon redressé a les mêmes caractéristiques que celles observées sur la population de référence, et cela pour l'ensemble des variables de calage conservées dans le modèle final.

Après le calage, l'échantillon redressé a des répartitions (par exemple par variable discriminante AGGIR<sup>18</sup> ou par option tarifaire ou par zone géographique –rural ou urbain–, ou par taille de l'établissement ou selon la présence d'une unité PASA<sup>19</sup> ou encore selon l'âge des résidents...) identiques à celles de la population de référence.

## 4.4 La qualité de l'estimation

Une fois l'échantillon redressé, une estimation du coût moyen pour chaque groupe de résidents est disponible. Cette estimation peut être de plus ou moins bonne qualité selon les groupes.

---

<sup>15</sup> Pour plus de détails sur la constitution des groupes de résidents, se reporter à l'annexe 1

<sup>16</sup> La liste des variables retenues dans au moins un groupe est disponible en annexe 4

<sup>17</sup> Cette macro est disponible sur le site de l'INSEE dans la rubrique « Définitions, méthodes et qualité – Modèles méthodes et outils statistiques – Les outils statistiques – La macro SAS CALMAR »

<sup>18</sup> Les 10 variables discriminantes concernent la partie « activité corporelle et mentale » de la grille AGGIR. Ce sont la cohérence, l'orientation, la toilette, l'habillement, l'alimentation, l'élimination, les transferts, les déplacements à l'intérieur, les déplacements à l'extérieur et la communication pour alerter.

<sup>19</sup> Pôle d'Activités et de Soins Adaptés

La qualité d'une estimation se juge avec l'ERE<sup>20</sup> (Erreur Relative d'Echantillonnage). Si l'ERE est inférieure à 10% alors l'estimation sera jugée comme étant de **bonne** qualité, si l'ERE est comprise entre 10% et 30% alors l'estimation sera **correcte**. En revanche si l'ERE est supérieure à 30% alors la qualité de l'estimation est **mauvaise**.

Il conviendra donc, avant de prendre connaissance du coût moyen estimé pour un groupe de résidents, de regarder cette ERE afin d'apprécier la qualité de l'estimation correspondante.

Un intervalle de confiance est associé au coût moyen du groupe ainsi qu'à son ERE. Plus l'ERE est élevé, plus l'intervalle de confiance est large. Pour une estimation de bonne qualité avec une ERE très faible, l'intervalle sera assez resserré.

Par exemple, pour le groupe 22<sup>21</sup>, le coût moyen 2016 par journée de prise en charge a été estimé à 113€. L'ERE est de 5,5% et l'intervalle de confiance associé est [101€ ; 125€]. Cela signifie que le vrai coût<sup>22</sup> a 95% de chances d'être compris entre 101€ et 125€.

---

<sup>20</sup> L'annexe 5 décrit l'ERE en détail

<sup>21</sup> Groupe 22. P2 – rang 1 : personnes âgées avec troubles psychiatriques, rang 1

<sup>22</sup> Le vrai coût correspond à celui qui aurait été obtenu si tous les EHPAD de la population de référence avaient pu participer à l'ENC

# Annexes

## ANNEXE 1 : Méthodologie de construction des groupes de résidents

L'étude nationale de coûts en EHPAD a pour objectif de fournir un coût moyen journalier par groupes de résidents hébergés. Le groupe de chaque résident est défini en fonction de son état de santé et de son degré d'autonomie.

Ainsi, pour les besoins de la restitution des résultats de l'ENC EHPAD 2015 et en l'absence de typologie déjà existante, une typologie des résidents hébergés en EHPAD a été élaborée « à dire d'experts ». Un groupe de travail composé de médecins coordonnateurs désignés par les fédérations d'EHPAD et animé par les médecins de la CNSA et de l'ATIH a été mis en place en 2016. L'objectif de ce groupe de travail était de constituer *a priori* des groupes de résidents homogènes d'un point de vue médical, c'est-à-dire en termes de perte d'autonomie et d'état de santé.

La constitution des groupes devait tenir compte de plusieurs contraintes :

- Les variables mobilisées devaient être issues des outils de recueil utilisés en routine dans les EHPAD, à savoir la grille AGGIR pour le niveau d'autonomie et l'outil PATHOS pour l'état de santé.
- La constitution des groupes devait être qualitative. Chaque groupe devait être élaboré « à dire d'expert » pour être, *a priori*, « homogène en termes d'autonomie et d'état de santé ». Par conséquent, les ressources consommées par les résidents, ainsi que les classifications médico-économiques déjà existantes (par exemple : groupes iso-ressources ou postes de ressources et points PATHOS) ne pouvaient pas être utilisées pour l'élaboration des groupes.
- L'effectif des groupes ne devait être ni trop faible (si possible  $\geq 50$  résidents) pour rendre possible le calcul d'un coût redressé, ni trop élevé pour favoriser « l'homogénéité médicale » des résidents au sein d'un groupe.  
Cependant, si un groupe de « petit effectif » ne pouvait pas être regroupé avec un autre groupe sous peine d'hétérogénéité médicale au sein du groupe final, il a été choisi de conserver le groupe de « petit effectif » même si aucun coût redressé n'a pu être calculé pour ce dernier.

Deux variables ont été choisies pour la constitution des groupes de résidents :

- o Les profils de soins<sup>23</sup> de [l'outil PATHOS](#) qui indiquent le niveau de soins requis pour chaque état pathologique d'un résident.
- o Le rang ou profil homogène de niveau d'autonomie de la [grille AGGIR](#) qui regroupe les résidents ayant des profils d'expression de l'autonomie significativement proches.

Au total, **83 groupes ont été constitués**<sup>24</sup> « à dire d'experts médicaux » à partir du croisement entre la combinaison des profils de soins PATHOS et le rang AGGIR des résidents. Lorsque le nombre de résidents pour une combinaison de profils de soins donnée était trop faible, les rangs ont été soit regroupés, soit non pris en compte.

L'arbre de décision présenté ci-après explique le classement de chaque résident dans les groupes. Cet arbre se lit de haut en bas et de gauche à droite. Chaque groupe est exclusif : un résident ne peut appartenir qu'à un groupe et un seul.

---

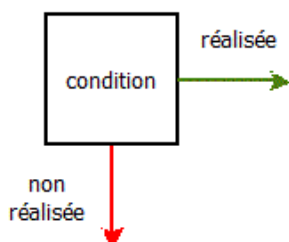
<sup>23</sup> L'annexe 3 détaille les 12 profils de soins, les 50 états pathologiques et les domaines d'état pathologique

<sup>24</sup> La liste des groupes est disponible dans le fichier Excel des coûts de référence publié sur le site de l'ATIH : <https://www.scansante.fr/coouts-de-referance-enc-ehpad-2016>.



## Symboles et abréviations utilisés dans l'arbre de décision

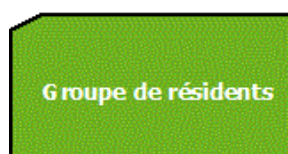
Les tests dichotomiques portant sur les données médicales des résidents, Profils de soins (PS) issus de l'outil PATHOS et rangs (=profil homogène d'autonomie) issus de la grille AGGIR, sont représentés ainsi :



Un cercle signale que la condition porte sur le profil de soins. Pour un résident, la condition est considérée comme remplie lorsque le résident a le ou les profils de soins associé(s) à au moins un état pathologique (codage via l'outil PATHOS).



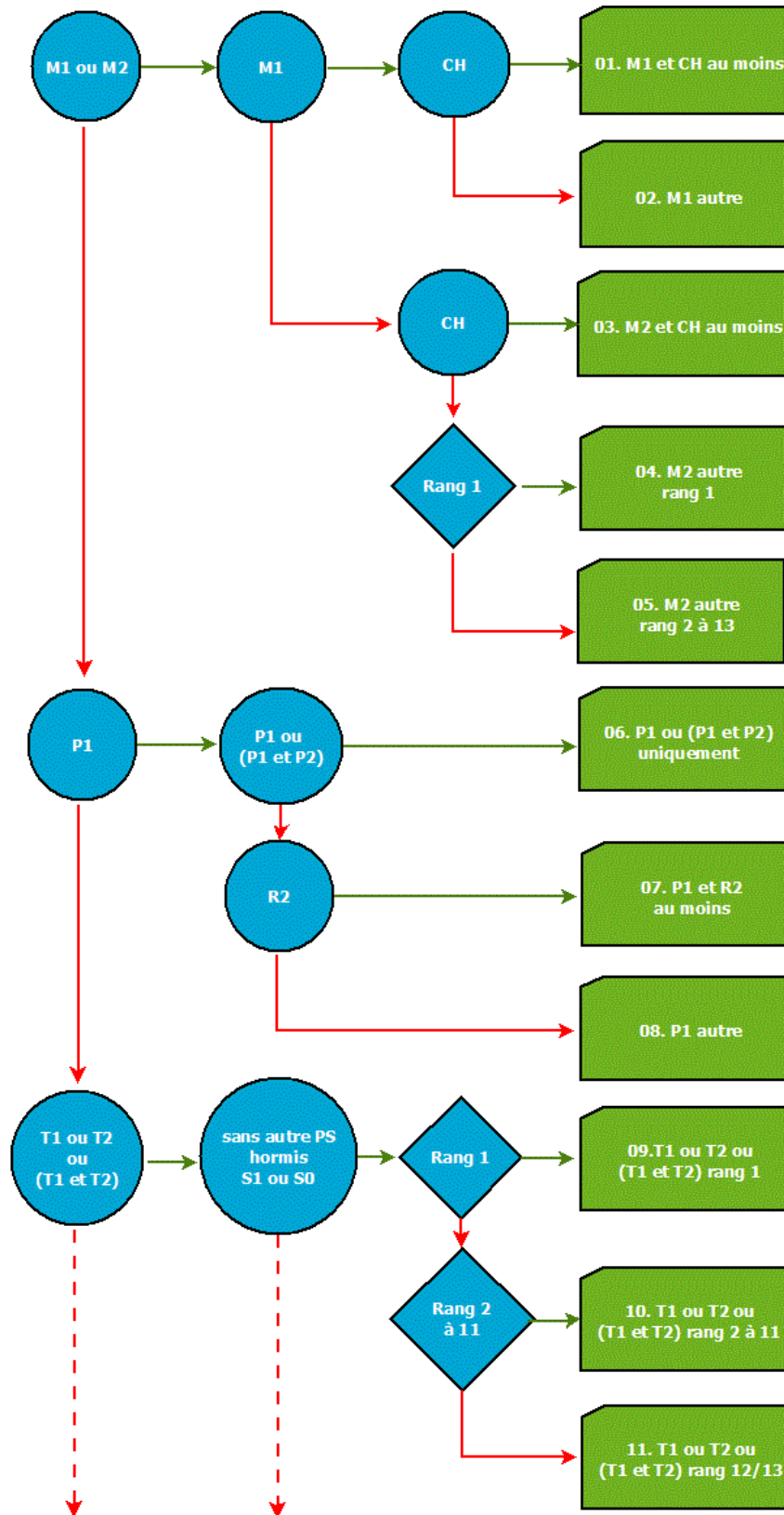
Un losange signale que la condition porte sur le rang AGGIR. La condition est considérée comme remplie lorsque le résident a le niveau d'autonomie correspondant au rang AGGIR indiqué (codage via la grille AGGIR).



Un rectangle signale que le résident a pu être classé dans un groupe de résidents. Le numéro et le nom du groupe sont précisés dans le rectangle vert.

## Arbre de décision

L'arbre de décision présenté explique le classement des résidents dans les 11 premiers groupes de résidents. La même logique est à appliquer pour le classement des résidents dans les 72 groupes suivants.



## ANNEXE 2 : Calendrier de la campagne ENC 2016

La campagne ENC EHPAD 2016 a débuté en janvier 2016 et s'est clôturée en septembre 2017.

ENC EHPAD 2016	2016												2017								
	Trimestre 1			Trimestre 2			Trimestre 3			Trimestre 4			Trimestre 1			Trimestre 2			Trimestre 3		
	Janv	Fév	Mars	Avril	Mai	Juin	Juillet	Août	Sept	Oct	Nov	Déc	Janv	Fév	Mars	Avril	Mai	Juin	Juillet	Août	Sept
EHPAD		Recueil T1		Recueil T2			Recueil T3			Recueil T4								Retraitement données comptables 2016			
SUPERVISEUR		Prise de contact																			Validation données comptables
ARS			Validation coupe PATHOS																		
ATI	Formation et mise à disposition logiciel données activité			Analyse données T1			Analyse données T2			Analyse données T3			Analyse données T4				Formation et mise à disposition logiciel données comptables			Analyse données comptables	

## ANNEXE 3 : Liste des 12 profils de soins, des 50 états pathologiques et des domaines d'états pathologiques

<b>Profil de soins</b>	<b>Description</b>
T1	Pronostic vital en jeu dans l'immédiat
T2	Équilibration thérapeutique et surveillance rapprochée
P1	Prise en charge psychiatrique d'un état de crise
P2	Prise en charge psychothérapique de soutien
R1	Rééducation Réadaptation Fonctionnelle intensive individuelle
R2	RRF d'entretien ou allégée
CH	Soins locaux complexes et longs
DG	Investigations pour une pathologie froide non encore diagnostiquée
M1	Soins terminaux psychothérapiques et /ou techniques lourds
M2	Soins terminaux d'accompagnement
S1	Surveillance épisodique programmée
S0	Aucun soin médical ou technique

Lecture : les états pathologiques 1 à 8 correspondent au domaine d'états pathologiques nommé « affections cardio-vasculaires ».

<b>Etat Pathologique</b>	<b>Description</b>
<b>Affections cardio-vasculaires</b>	
01	Insuffisance cardiaque
02	Coronaropathie
03	Hypertension artérielle
04	Troubles du rythme
05	Phlébites
06	Embolie et thrombose artérielle, amputation
07	Artériopathies chroniques
08	Hypotension orthostatique
<b>Affections neuro-psychiatriques</b>	
09	Malaises, vertiges, perte de connaissance brève, chutes
10	Accidents vasculaires cérébraux
11	Comitialité focale et généralisée
12	Syndrome parkinsonien
13	Syndrome confusionnel aigu

<b>Etat Pathologique</b>	<b>Description</b>
14	Troubles chroniques du comportement
15	Etats dépressifs
16	Etats anxieux
17	Psychose, délires, hallucinations
18	Syndrome démentiel
<b>Affections broncho-pulmonaires</b>	
19	Broncho-pleuro-pneumopathies
20	Insuffisance respiratoire
21	Embolies pulmonaires
<b>Pathologies infectieuses</b>	
22	Syndromes infectieux généraux
23	Syndromes infectieux locaux
24	Infections urinaires basses
<b>Affections dermatologiques</b>	
25	Escarres
26	Autres lésions cutanées
<b>Affections ostéo-articulaires</b>	
27	Pathologie de la hanche
28	Pathologie de l'épaule
29	Pathologie vertébro-discale
30	Autres pathologies osseuses
31	Polyarthrite et pathologies articulaires
<b>Affections gastro-entérologiques</b>	
32	Syndromes digestifs hauts
33	Syndromes abdominaux
34	Affection hépatique, biliaire, pancréatique
35	Dénutrition
<b>Affections endocriniennes</b>	
36	Diabète
37	Dysthyroïdie
38	Troubles de l'hydratation
<b>Affections uro-néphrologiques</b>	
39	Rétention urinaire
40	Insuffisance rénale
41	Incontinence

<b>Etat Pathologique</b>	<b>Description</b>
<b>Autres domaines</b>	
42	Anémies
43	Etats cancéreux
44	Hémopathies malignes
45	Syndrome inflammatoire, fièvre inexpliquée
46	Pathologies oculaires évolutives
47	Etat grabataire
48	Etat terminal
49	Autres pathologies
50	Aucune pathologie pertinente à retenir

## ANNEXE 4 : Variables utilisées pour le redressement de l'échantillon

Le tableau suivant décrit les variables qui ont été utilisées dans au moins un groupe de résidents pour le redressement de l'échantillon. Certaines variables caractérisent l'EHPAD tandis que d'autres caractérisent le résident.

Variable retenue pour le calage	Description de la variable
Tarification de l'EHPAD	La variable prend 2 valeurs différentes selon la tarification de l'établissement <i>global</i> <i>partiel</i>
Taille de l'EHPAD	La variable prend 3 valeurs différentes selon la capacité en nombre de places autorisées <i>moins de 80 places</i> <i>de 80 à 100 places</i> <i>plus de 100 places</i>
Présence d'un PASA dans l'EHPAD	La variable prend 2 valeurs différentes selon que l'EHPAD possède un Pôle d'Activités et de Soins Adaptés ou pas <i>présence</i> <i>absence</i>
Statut juridique de l'EHPAD	La variable prend 2 valeurs différentes selon le statut juridique de l'EHPAD <i>public</i> <i>privé</i>
Variables discriminantes AGGIR <sup>25</sup>	La variable prend 3 valeurs différentes selon le codage de la variable discriminante <i>A</i> <i>B</i> <i>C</i>
Présence d'une UHR dans l'EHPAD	La variable prend 2 valeurs différentes selon que l'EHPAD possède une Unité d'Hébergement Renforcée ou pas <i>présence</i> <i>absence</i>
Zone géographique	Typologie mise en place par l'INSEE définissant une commune. Ici c'est la commune de l'établissement qui est typée. <i>rural</i> <i>urbain</i>
Rattachement à un établissement sanitaire	La variable prend 2 valeurs différentes selon que l'EHPAD fait partie d'un établissement juridique dans lequel il y a un établissement sanitaire ou pas. <i>rattaché</i> <i>non rattaché</i>
Indicatrice d'âge du résident	La variable prend 2 valeurs différentes selon l'âge du résident. <i>moins de 80 ans</i> <i>plus de 80 ans</i>

<sup>25</sup> Il y a 10 variables discriminantes AGGIR : Transferts, Alimentation, Déplacements extérieurs, Alerter, Cohérence, Elimination, Habillage, Déplacements intérieurs, Toilette et Orientation. Les variables AGGIR sont codées A si le résident « fait seul, totalement, habituellement et correctement », B si le résident « fait partiellement, ou non habituellement ou non correctement » et C si le résident « ne fait pas »

<b>Variable retenue pour le calage</b>	<b>Description de la variable</b>
<b>Présence d'une PUI</b>	La variable prend 2 valeurs différentes selon qu'il y a une Pharmacie à Usage Intérieur dans l'EHPAD ou pas <i>présence</i> <i>absence</i>
<b>Domaine d'état pathologique<sup>26</sup></b>	Une variable est créée pour chaque domaine d'état pathologique. La variable prend 2 valeurs différentes : si un profil de soins (autre que S0 et S1) est codé sur un des états pathologiques du domaine alors la variable prend la valeur 1. Sinon la variable vaut 0. <i>0</i> <i>1</i>
<b>Tranches d'âge du résident</b>	La variable prend 6 valeurs différentes selon l'âge du résident. <i>moins de 80 ans</i> <i>80-85 ans</i> <i>85-90 ans</i> <i>90-95 ans</i> <i>95-100 ans</i> <i>plus de 100 ans</i>
<b>Nombre de profils de soins<sup>27</sup> codés sur les 50 états pathologiques</b>	Variable numérique prenant pour valeur le nombre de fois où le profil de soins est codé parmi les 50 états pathologiques
<b>Nombre de points sur le poste ressource imagerie ou psychothérapie (PATHOS)</b>	La variable prend 5 valeurs différentes selon le nombre de points sur le poste ressource (imagerie ou psychothérapie). Les seuils <sup>28</sup> sont différents pour chaque poste ressource. <i>0</i> <i>A</i> <i>B</i> <i>C</i> <i>autre</i>

<sup>26</sup> Les domaines d'états pathologiques sont décrits en annexe 3. Les domaines utilisés pour le calage sont : affections ostéo-articulaires, affections neuropsychiatriques (hors troubles du comportement), troubles chroniques du comportement, affections cardio-vasculaires, état cancéreux, autres (hors état cancéreux), affections broncho-pulmonaires, affections endocriniennes, escarres, affections gastroentérologiques, pathologies infectieuses, affections uro-néphrologiques.

<sup>27</sup> Les profils de soins sont disponibles en annexe 3. Les profils de soins utilisés pour le redressement sont : P2 : Prise en charge psychothérapique de soutien, T2 : Equilibration thérapeutique et surveillance rapprochée.

<sup>28</sup> Seul 2 postes ressources sont pris en compte dans le calage : imagerie et psychothérapie.

La variable créée avec le nombre de points sur le poste ressource (PATHOS) **d'imagerie** est codée « autre » si la valeur est manquante, 0 si le nombre de points est strictement inférieur à 9, A si le nombre de points est compris entre 9 (inclus) et 16 (exclus), B si le nombre de points est compris entre 16 (inclus) et 36 (exclus) et C si le nombre de points est supérieur ou égal à 36.

La variable créée avec le nombre de points sur le poste ressource (PATHOS) de **psychothérapie** est codée « autre » si la valeur est manquante, 0 si le nombre de points est strictement inférieur à 19, A si le nombre de points est compris entre 19 (inclus) et 61 (exclus), B si le nombre de points est compris entre 61 (inclus) et 100 (exclus) et C si le nombre de points est supérieur ou égal à 100.



## ANNEXE 5 : Qu'est-ce que l'ERE ?

L'ERE (Erreur Relative d'Echantillonnage), est un indicateur permettant de mesurer la qualité d'une estimation. Pour les groupes ayant moins de 30 observations (résident-trimestre), l'ERE n'est pas calculable.

Il est important, avant de prendre connaissance du coût moyen estimé pour un groupe de résidents, de regarder son ERE afin d'apprécier la qualité de l'estimation correspondante.

Si l'ERE est inférieure à 10% alors l'estimation sera jugée comme étant de **bonne** qualité, si l'ERE est comprise entre 10% et 30% alors l'estimation sera **correcte**. En revanche si l'ERE est supérieure à 30% alors la qualité de l'estimation est **mauvaise**.

Le calcul de l'ERE prend en compte les variables utilisées dans le redressement (ou calage) au travers des résidus  $U_{ij}$  (avec  $i$  l'établissement et  $j$  l'observation) de la régression multiple des coûts complets  $C_{ij}$  sur l'ensemble des variables utilisées pour le calage du groupe. La régression est conçue groupe par groupe. Pour un groupe donné, la régression est effectuée sur l'ensemble des observations de l'échantillon de l'ENC.

**Pour un groupe  $g$  donné**, soit  $U_i = \sum_{j \in i} U_{ij}$  et  $\bar{U} = \frac{1}{m} \sum_i U_i$

où  $m$  est le nombre d'établissements dans l'échantillon de l'ENC.

La variance estimée  $\hat{V}$  du coût moyen se formule de la façon suivante :

$$\hat{V} = \frac{1}{N^2} \left(1 - \frac{m}{M}\right) \frac{M^2}{m(m-1)} \sum_i (U_i - \bar{U})^2$$

où :  $N$  est le nombre d'observations pour la population de référence,

$m$  est le nombre d'établissements dans l'échantillon de l'ENC,

$M$  est le nombre d'établissements dans la population de référence.

Il est important de noter que les variables de régression mobilisées pour le calcul des résidus  $U_{ij}$  ne sont pas les mêmes pour tous les groupes puisque les variables de calage sont propres à chaque groupe.

Pour chaque groupe, l'ERE est définie comme :

$$ERE = \frac{\sqrt{\hat{V}}}{\hat{C}}$$

où  $\hat{C}$  est le coût de référence estimé pour le groupe. Exprimée en pourcentage, l'ERE permet d'apprécier l'incertitude relative qui entoure l'estimation du coût moyen du groupe. Cette incertitude est mesurée grâce à l'intervalle de confiance qui se calcule ainsi :

$$\left[ \hat{C} (1 - 1,96 ERE) ; \hat{C} (1 + 1,96 ERE) \right]$$

Plus l'ERE est élevée, plus l'intervalle de confiance est large. Pour une estimation de bonne qualité avec une ERE très faible, l'intervalle de confiance sera assez resserré.

Concrètement, le vrai coût<sup>29</sup> a 95% de chances d'être compris dans cet intervalle de confiance.

Par exemple, pour le groupe 22<sup>30</sup>, le coût moyen 2016 par journée de prise en charge a été estimé à 113€. L'ERE est de 5,5% et l'intervalle de confiance associé est [101€ ; 125€]. Cela signifie que le vrai coût pour le groupe 22 a 95% de chances d'être compris entre 101€ et 125€.

---

<sup>29</sup> Le vrai coût correspond à celui qui aurait été obtenu si tous les EHPAD de la population de référence pouvaient participer à l'ENC

<sup>30</sup> Groupe 22. P2 – rang 1 : personnes âgées avec troubles psychiatriques, rang 1