

TRAITEMENT DE LA NON-RÉPONSE PARTIELLE PAR IMPUTATION MULTIPLE DANS UNE ENQUÊTE EN POPULATION GÉNÉRALE

Noémie SOULLIER*

(*) DREES, Bureau Handicap-Dépendance

Introduction

La non-réponse partielle est un problème récurrent dans les analyses statistiques. Il faut alors choisir comment traiter la non-réponse pour obtenir des résultats fiables. Dans les enquêtes en population générale, la non-réponse peut être gérée par imputation simple par la moyenne ou par imputation aléatoire par strate. Face à un taux élevé de non-réponse, l'imputation multiple a été préférée, afin d'assurer une meilleure estimation de l'incertitude. Son utilisation dans le cadre d'une enquête en population générale est ici présentée.

1. Les enquêtes Handicap-Santé

L'enquête Handicap-Santé est une enquête nationale sur la santé et le handicap, réalisée en 2008-2009 par la Direction de la recherche, des études, de l'évaluation et des statistiques (*Drees*) et l'Institut national de la statistique et des études économiques (*Insee*). Cette enquête permet de mieux appréhender les difficultés rencontrées dans leur vie quotidienne par les personnes malades, en situation de handicap ou de perte d'autonomie, de recenser les aides dont elles bénéficient, ainsi que celles dont elles auraient besoin. Dans le but de mieux identifier et évaluer les difficultés spécifiques à ces personnes, cette enquête est menée sur l'ensemble de la population, quel que soit l'état de santé ou de handicap. Les données ainsi collectées et analysées servent à mieux cibler les politiques publiques. Cette opération de collecte renouvelle l'enquête nationale Handicaps-Incapacités-Dépendance (HID) réalisée en 1998-2001.

L'enquête Handicap-Santé est composée de trois volets :

- une enquête en ménages ordinaires,
- une enquête en institution,
- une enquête auprès des aidants de l'entourage.

1.1. L'enquête Handicap-Santé en ménages ordinaires (HSM)

L'enquête Handicap-Santé en ménages ordinaires (Insee, 2008-2009) a permis de collecter des informations auprès de 30 000 personnes vivant à domicile en 2008. Un des principaux objectifs de l'enquête est d'estimer le nombre de personnes rencontrant des difficultés pour réaliser les activités de la vie quotidienne en raison de problèmes de santé ou d'un handicap, et ceci en prenant en compte l'environnement dans lequel elles évoluent. Elle s'attache également à mesurer les désavantages sociaux vécus par ces personnes, et à relever la nature, la quantité et l'origine des aides existantes, ainsi que les besoins demeurant non satisfaits. Outre la santé, l'enquête aborde les thèmes de l'emploi, de la scolarité, des loisirs, de l'accessibilité du logement. Elle contient également un module spécifique sur l'aide reçue de la part de professionnels ou de non-professionnels.

L'enquête Handicap-Santé couvre toute la population, y compris les personnes en bonne santé, afin de repérer les difficultés spécifiques rencontrées par les personnes handicapées ou les personnes âgées dépendantes. Pour que les personnes en situation de dépendance ou de handicap soient interrogées en nombre suffisant, une enquête filtre a été réalisée au préalable afin de constituer l'échantillon de Handicap-Santé.

1.2. L'enquête Handicap-Santé en institutions (HSI)

L'enquête Handicap-Santé en institutions (Insee, 2009) a permis de collecter des informations auprès de 10 000 personnes hébergées dans 1 500 structures sanitaires, médico-sociales et sociales. Elle complète l'enquête HSM, de façon à fournir une image complète de l'état de santé de la population sur le territoire français. Cette enquête a pour objectif de connaître l'état de santé des personnes vivant en institution et les caractéristiques de leurs situations de handicap et de dépendance, ainsi que les difficultés rencontrées par ces personnes dans leur vie quotidienne, les aides dont elles bénéficient et les aides dont elles auraient besoin.

Pour cette enquête, des structures sanitaires (services de psychiatrie ou de soins de longue durée), médico-sociales et sociales ont été tirées au hasard.

1.3. L'enquête Handicap-Santé auprès des aidants de l'entourage (HSA)

L'enquête Handicap-Santé auprès des aidants de l'entourage (Drees, 2008) fait suite à l'enquête HSM. Elle interroge près de 5 000 aidants de 16 ans ou plus faisant partie de l'entourage (famille, amis, voisins) et prodiguant régulièrement pour des raisons de santé ou de handicap une aide à la vie quotidienne, ou une aide matérielle ou financière, ou un soutien moral à une personne interrogée dans le cadre de l'enquête HSM. L'objectif général de l'enquête HSA est de décrire les relations d'aide entre les personnes aidées et leurs aidants de l'entourage, ainsi que les conséquences de cette aide sur les conditions de vie des aidants. L'enquête permet notamment de mieux connaître ce que font les aidants, leur nombre, leurs difficultés et leurs besoins.

Pour en savoir plus sur les enquêtes HSM et HSA : <http://www.sante.gouv.fr/handicap-sante.html>.

1.4. Contexte de l'aide aux personnes âgées à domicile

L'enquête HSM permet de décrire l'aide reçue par les personnes âgées à domicile. En 2008, 3,6 millions de personnes âgées de 60 ans ou plus et vivant à domicile sont aidées régulièrement en raison d'un problème de santé ou d'un handicap [1]. Cette aide peut être une aide dans les tâches de la vie quotidienne, et/ou une aide matérielle ou financière, et/ou un soutien moral.

Le recours à l'aide augmente graduellement avec l'âge, le vieillissement s'accompagnant d'une augmentation de la perte d'autonomie. Ainsi, 15 % des personnes de 60 à 74 ans sont aidées, alors que 50 % de celles de 75 ans ou plus le sont. Le recours à l'aide est également lié au degré de la perte d'autonomie. Ainsi, la quasi-totalité des personnes âgées dépendantes (au sens de la grille AGGIR¹) reçoivent une aide régulière.

Parmi les personnes âgées aidées, 48% reçoivent uniquement une aide de leur entourage, 20% uniquement l'aide de professionnels et 32% une aide mixte provenant à la fois de leur entourage et de professionnels. Ainsi, huit personnes âgées aidées sur dix sont aidées par leur entourage. La dépendance sévère s'accompagne d'une prépondérance de l'aide mixte. Dans l'ensemble, un tiers des personnes âgées aidées régulièrement pour des raisons de santé ou de handicap ressent le besoin d'une aide humaine supplémentaire pour la vie quotidienne.

Le volume d'heures d'aide est un point central lorsque l'on étudie l'implication de l'entourage et des professionnels auprès des personnes âgées à domicile. Il permet d'évaluer la quantité de l'aide apportée par les aidants qu'ils soient professionnels ou de l'entourage.

¹ Voir annexe 1

2. Les volumes d'heures d'aide à domicile

2.1. Champ de l'étude

L'étude de la non-réponse au volume d'aide s'inscrit dans le cadre de l'étude de l'aide aux personnes âgées de 60 ans ou plus et vivant à domicile. Elle s'appuie principalement sur l'enquête HSM et le champ d'étude du volume d'aide est défini comme l'ensemble des personnes âgées de 60 ans ou plus déclarant être aidée pour raison de handicap ou de santé :

- par un professionnel pour l'étude du volume d'aide professionnelle,
- par un proche pour l'étude du volume d'aide informelle.

L'aide professionnelle pour raison de santé ou de handicap correspond à l'aide d'un professionnel pour :

- les activités de la vie courante (ADL),
- les activités instrumentales de la vie courante (IADL)
 - lorsque l'incapacité est due à l'état de santé ou à un handicap,
 - lorsqu'une situation de dépendance est repérée par l'aide de l'entourage (ADL, IADL, autres tâches, aide financière) pour raison de santé, par une forte limitation dans les activités habituelles depuis au moins 6 mois, par une forte restriction dans les activités de la vie quotidienne (ADL, IADL), par la présence d'une allocation d'adultes handicapés ou une pension d'invalidité,
- d'autres tâches de la vie quotidienne pour raison de santé ou handicap.

L'aide informelle pour raison de santé ou de handicap correspond à l'aide de l'entourage pour :

- les activités de la vie courante (ADL),
- les activités instrumentales de la vie courante (IADL) lorsque l'incapacité est due à l'état de santé ou à un handicap,
- d'autres tâches de la vie quotidienne pour raison de santé ou handicap,
- une aide financière, matérielle ou un soutien moral pour raison de santé ou handicap.

Ainsi, 3,6 millions de personnes âgées de 60 ans ou plus sont aidées pour raison de santé ou de handicap : 1,9 million reçoivent au moins l'aide d'un professionnel et 2,9 millions reçoivent au moins l'aide de leur entourage (1,2 million reçoivent les deux types d'aide).

2.2. Questionnaire

Dans l'enquête HSM, la personne aidée est questionnée sur le volume horaire de l'aide apportée par ses aidants professionnels et de l'entourage, lorsqu'elle est aidée par ceux-ci pour les tâches de la vie quotidienne. Ainsi, 1,9 million de personnes âgées recevant une aide à la vie quotidienne apportée par un ou plusieurs aidants professionnels est questionné sur le volume horaire de l'aide apportée par ces aidants ; et 2,2 millions recevant cette aide de la part d'un ou plusieurs proches sont questionnés sur le temps passé par leur(s) proche(s) pour les aider dans les tâches de la vie quotidienne. Les 900 000 personnes qui reçoivent une aide à la vie quotidienne de la part des deux types d'aidants répondent donc aux deux questionnements.

Pour les aidants professionnels, le questionnement est le suivant :

1. À quelle fréquence vient-il/elle chez vous pour vous aider ? → nombre de fois par jour/semaine/mois
2. Globalement, combien d'heures vous aide-t-il/elle par jour/semaine/mois ? (le pas de temps est le même que celui utilisé par la personne interrogée à la question précédente)

Les mêmes questions sont posées pour les aidants de l'entourage qui ne résident pas avec la personne aidée.

En revanche, pour les aidants cohabitants, la question de la fréquence des visites n'est pas pertinente et seule la question du nombre d'heures est posée :

- Globalement, combien d'heures vous aide-t-il/elle par semaine ?

2.3. Correction des incohérences

2.3.1. Aide professionnelle

Certaines fréquences renseignées semblent importantes par rapport au pas de temps choisi (supérieures à 7 fois par jour, 7 fois par semaine et 31 fois par mois). La notion de fréquence des visites peut avoir été mal comprise et il se peut que soit alors renseigné non pas la fréquence de l'aide mais le nombre d'heures d'aide. Certains choix de pas de temps posent également question : 4 à 7 fois par jour ne serait-il pas plutôt par semaine ?

Peu d'impossibilités sont relevées sur le nombre d'heures. Seule l'aide journalière supérieure à 24 heures est incohérente (25 à 61 heures). Les nombres d'heures codés 90 ou 99 laissent à penser que l'enquêteur souhaitait indiquer « ne sait pas ». Les nombres d'heures nuls sont conservés car selon les consignes de collecte, ils peuvent correspondre à un volume horaire compris entre 1 et 29 minutes. Les autres volumes horaires sont plausibles. Un contrôle en parallèle de la fréquence permet de recoder de manière plus cohérente.

La cohérence entre la fréquence et le nombre d'heures est questionnée lorsque la durée de la visite (= durée totale / fréquence) est très faible (moins de 30 minutes) ou très élevée. Une durée de visite plus élevée est tolérée lorsque la fréquence est faible (par semaine ou par mois) : il est possible que le professionnel intervienne très ponctuellement lorsqu'il vient souvent, alors qu'il peut prendre plus de temps à chaque fois lorsque les visites sont plus espacées.

Cas où des données renseignées sont recodées manquantes (n=206) :

- Fréquence = 0
- Fréquence > 7 fois par jour
- Fréquence > 31 fois par mois
- Nombre d'heures = 90, 99 ou 999
- Nombre d'heures > 40 h par jour
- Fréquence et nombre d'heures = 0
- Fréquence comprise entre 12 et 40 fois par semaine et nombre d'heures = fréquence

Ces données font donc partie des non-réponses.

Cas où la fréquence est modifiée (n=82) :

- Fréquence comprise entre 5 et 7 fois par jour et nombre d'heures < 8h par jour → fréquence par semaine
- Fréquence = 1 fois par semaine et nombre d'heures = 14h par semaine → fréquence = 14 fois par semaine
- Fréquence = 2 ou 3 fois par semaine et nombre d'heures > 24h par semaine → fréquence = 2 ou 3 fois par jour

Cas où le nombre d'heures est modifié (n=174) :

- Fréquence non nulle et nombre d'heures nul → nombre d'heures = 0.5 (n=107)
- Fréquence < 5 fois par jour et nombre d'heures compris entre 25 et 40h par jour → nombre d'heures par semaine
- Fréquence = 4 ou 5 fois par semaine et nombre d'heures < 2h par semaine → nombre d'heures par jour
- Fréquence = 6 ou 7 fois par semaine et nombre d'heures < 3h par semaine → nombre d'heures par jour
- Fréquence = 14 fois par semaine et nombre d'heures < 7h par semaine → nombre d'heures par jour
- Fréquence comprise entre 5 et 31 fois par mois et nombre d'heures = 1h par mois → nombre d'heures par jour

Cas où la fréquence et le nombre d'heures sont modifiés (n=2) :

- Fréquence ≥ 5 fois par jour et nombre d'heures compris entre 25 et 40h par jour → fréquence et nombre d'heures par semaine

2.3.2. Aide de l'entourage

Les mêmes corrections ont été apportées pour les fréquences et les nombres d'heures des aidants de l'entourage. La cohérence entre la fréquence et le nombre d'heures ne peut être étudiée que pour les aidants de l'entourage qui ne résident pas avec la personne aidée, puisque la fréquence des visites n'est pas pertinente pour les cohabitants. Pour ces aidants, le recodage tolère des durées de visite plus faibles (10 minutes) et des fréquences plus importantes que celles tolérées pour les professionnels.

Cas où des données renseignées sont recodées manquantes (n=1422) :

- Fréquence = 0
- Fréquence > 7 fois par jour
- Fréquence = (30, 34) par semaine
- Fréquence > 31 fois par mois
- Nombre d'heures = 90, 99 ou 999 (n=1385)
- Nombre d'heures > 40 h par jour
- Fréquence et nombre d'heures = 0
- Fréquence comprise entre 12 et 40 fois par semaine et nombre d'heures = fréquence

Ces données font donc partie des non-réponses.

Cas où la fréquence est modifiée (n=38) :

- Fréquence comprise entre 5 et 7 fois par jour et nombre d'heures < 2h par jour → fréquence par semaine
- Fréquence = 52 par semaine → fréquence = 1 par semaine

Cas où le nombre d'heures est modifié (n=23) :

- Fréquence non nulle et nombre d'heures nul → nombre d'heures = 0.5
- Fréquence < 5 fois par jour et nombre d'heures compris entre 25 et 40h par jour → nombre d'heures par semaine
- Fréquence comprise entre 5 et 31 fois par mois et nombre d'heures = 2h par mois → nombre d'heures par jour

Cas où la fréquence et le nombre d'heures sont modifiés (n=2) :

- Fréquence ≥ 5 fois par jour et nombre d'heures compris entre 25 et 40h par jour → fréquence et nombre d'heures par semaine

2.4. Non-réponse

Chaque réponse correspond à une relation d'aide aidé-aidant.

A l'issue du recodage, le taux de non-réponse concernant les questions sur la fréquence est faible : 1% des réponses sont manquantes pour la fréquence des visites des aidants professionnels et 2% pour les aidants de l'entourage non cohabitant.

Pour les aidants professionnels, le volume horaire de l'aide est bien rempli : 5% des réponses sont manquantes pour cette question. L'aidant professionnel intervient chez la personne pour un nombre d'heures et de tâches définis à l'avance. Un contrat détermine a priori la durée et le prix de l'intervention, notamment lorsque l'aide professionnelle intervient dans le cadre d'une prise en charge par l'Allocation personnalisée d'autonomie (APA). Il apparaît donc assez aisé pour la personne aidée de savoir combien d'heures un professionnel intervient chez elle.

A contrario, le taux de non-réponse concernant le nombre d'heures d'aide apporté par l'entourage est élevé : 29% des valeurs sont manquantes. La non-réponse touche particulièrement les aidants cohabitants : 22% des valeurs sont manquantes pour le volume horaire d'aide des aidants non-cohabitants et 37% le sont pour les cohabitants.

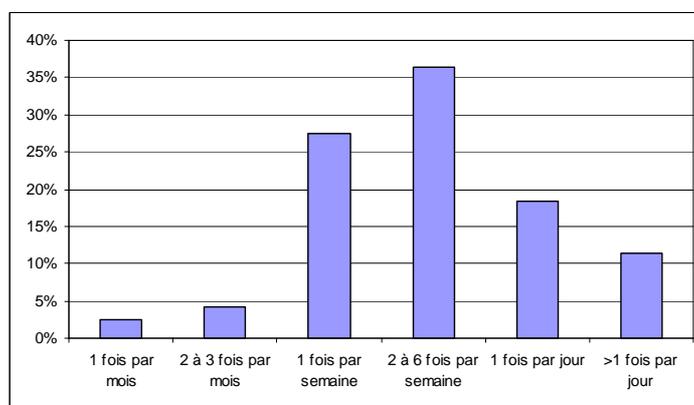
L'intervention d'aidants de l'entourage n'est en effet pas toujours programmée ni chronométrée. Il est alors parfois difficile pour les personnes aidées de dénombrer les heures que leur proche passe à les aider du fait de leur dépendance ou de leur handicap. Cela s'avère encore plus compliqué lorsque l'aidant cohabite avec la personne aidée, et notamment lorsque l'aidant est le conjoint. Il leur faut distinguer l'aide qui relève de leur problème de santé du partage des tâches qui assure le bon fonctionnement du ménage et cela n'est pas aisé. Tout cela peut expliquer le taux élevé de non-réponse sur le volume horaire de l'aide des aidants de l'entourage et en particulier des aidants cohabitants.

Ce fort taux de non-réponse oblige à traiter les réponses manquantes pour le volume horaire de l'aide apportée par l'entourage. Les autres variables sont analysées sur les répondants ; leurs distributions sont données ci-après.

2.5. Distributions des fréquences et du volume horaire de l'aide professionnelle

Pour présenter les résultats, les réponses sont agrégées au niveau de la personne aidée, en séparant les aidants professionnels des aidants de l'entourage. Pour la fréquence des visites, la valeur maximale est sélectionnée parmi les fréquences des aidants de la personne ; les résultats se lisent alors « la personne aidée reçoit l'aide d'un ou plusieurs professionnel(s)/proche(s) au moins ... fois par ... ». Pour le nombre d'heures d'aide, les volumes horaires de tous les aidants de la personne aidée sont sommés. La pondération de l'enquête peut alors être utilisée pour obtenir des résultats à l'échelle de la France entière.

Graphique 1. Fréquence de l'aide professionnelle reçue par les personnes âgées à domicile

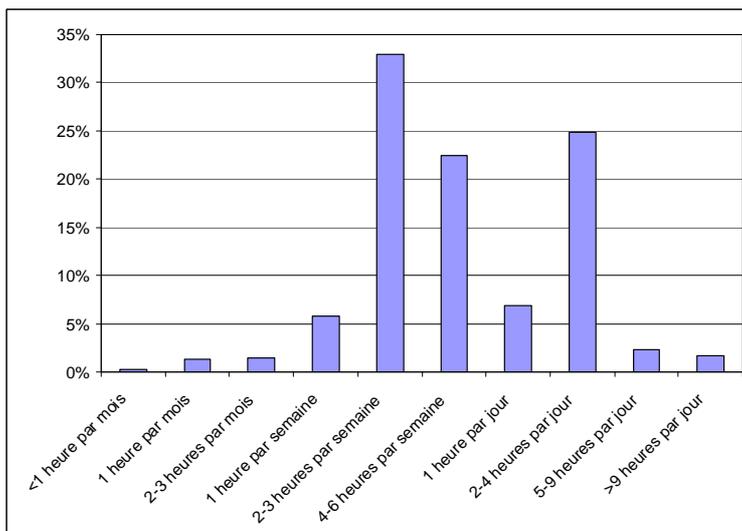


Champ : Personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs professionnel(s).

Source : Enquête Handicap-Santé 2008, volet ménages, Insee.

Note de lecture : 11% des personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs professionnel(s) le sont plus d'une fois par jour.

Graphique 2. Nombre d'heures d'aide professionnelle reçues par les personnes âgées à domicile

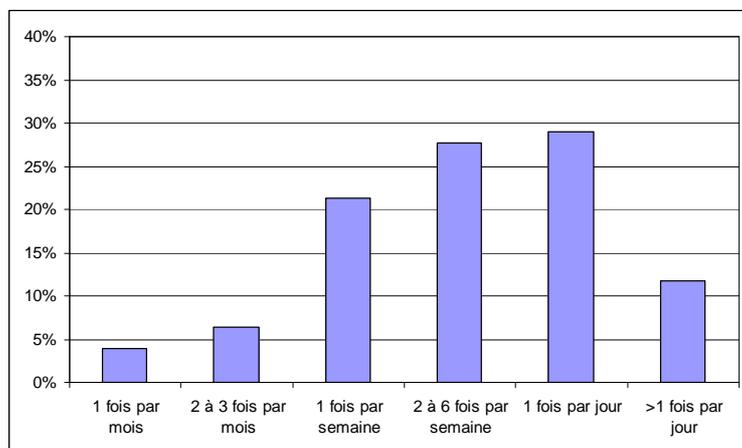


Champ : Personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs professionnel(s).

Source : Enquête Handicap-Santé 2008, volet ménages, Insee.

Note de lecture : 2% des personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs professionnel(s) le sont plus de neuf heures par jour.

Graphique 3. Fréquence de l'aide de l'entourage non cohabitant reçue par les personnes âgées à domicile



Champ : Personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs proche(s).

Source : Enquête Handicap-Santé 2008, volet ménages, Insee.

Note de lecture : 12% des personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs proche(s) le sont plus d'une fois par jour.

3. L'étude de la non-réponse du volume horaire de l'aide de l'entourage

3.1. Premières pistes pour réduire la non-réponse

3.1.1. Limiter le champ

La non-réponse révèle la possible difficulté des personnes âgées aidées à répondre à la question du volume horaire de l'aide apportée par leur(s) proche(s). Il est envisageable que cette difficulté soit plus ou moins grande selon que l'aide apportée est plus ou moins importante / concrète. Ainsi, une réduction du champ pourrait atténuer le taux de non-réponse. Plusieurs solutions ont été testées :

- réduire le champ aux premiers ou aux deux premiers aidants cités par la personne aidée. Ces aidants sont probablement les plus impliqués dans l'aide, puisque ce sont les premières personnes auxquelles la personne aidée a pensé. Déterminer un nombre d'heures d'aide pourrait alors être plus facile pour ces aidants.
- réduire le champ aux aidants qui apportent exclusivement une aide à la vie quotidienne. La question du volume horaire n'est posée que pour les aidants qui apportent une aide à la vie quotidienne, mais ils peuvent également apporter en supplément une aide financière ou matérielle et/ou un soutien moral. La combinaison de plusieurs types d'aide pourrait rendre plus difficile l'estimation du volume horaire par la personne aidée.
- réduire le champ aux aidants qui apportent exclusivement une aide à la vie quotidienne, à l'exception d'une aide qui consiste uniquement en « assurer une présence, une compagnie » et/ou « vérifier ce que la personne fait »². Ces deux postes d'activité correspondent à des tâches moins concrètes, pour lesquelles il peut être plus difficile de déterminer le temps passé.

La réduction du champ ne s'avère pas une bonne solution puisque le taux de données manquantes reste incompressible quelque soit la réduction du champ apportée (voir Tableau 1). De plus, cette réduction aurait pour effet de diminuer grandement la taille de l'échantillon d'étude, ce qui n'est pas souhaitable. Cela revient alors à étudier la question du volume horaire de l'aide sur une sous-population ; toute l'information souhaitée ne pourrait être recueillie sur ce champ.

Tableau 1. Taux de données manquantes sur le volume horaire de l'aide apportée par les aidants de l'entourage aux personnes âgées vivant à domicile, selon le champ choisi

	Tous	Premiers aidants	Deux premiers aidants	Aide à la vie quotidienne exclusive	Aide à la vie quotidienne exclusive hors présence ou vérification exclusives
Aidants non cohabitants	22	18	20	21	21
Aidants cohabitants	37	37	37	33	33
<i>Effectifs non cohabitants</i>	<i>2506</i>	<i>1244</i>	<i>1974</i>	<i>774</i>	<i>756</i>
<i>Effectifs cohabitants</i>	<i>2328</i>	<i>2089</i>	<i>2270</i>	<i>506</i>	<i>503</i>

Champ : Personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs proche(s).

Source : Enquête Handicap-Santé 2008, volet ménages, Insee.

Note de lecture : 22% des valeurs sont manquantes pour le nombre d'heures des aidants non cohabitants.

² Dans l'enquête Handicap-Santé en ménages ordinaires (HSM), l'aide à la vie quotidienne a été déclinée en huit postes d'activité : les soins personnels (toilette, habillage, repas), les tâches ménagères, les démarches administratives, assurer une présence, vérifier ce que la personne fait, s'occuper des problèmes de santé, faire les courses, aider pour d'autres activités.

3.1.2. Compléter en utilisant l'enquête HSA

L'enquête HSA est le pendant, auprès des aidants de l'entourage, de l'enquête HSM. Lorsque la personne aidée n'a pas répondu dans HSM, il est possible que l'aidant dont il est question ait été enquêté par HSA et qu'il ait lui répondu à la question du volume horaire de l'aide qu'il apporte.

Ainsi, si l'on essaie de compléter les valeurs manquantes dans HSM par les réponses apportées par les aidants dans l'enquête HSA, le taux de non-réponse diminue à 16% pour le nombre d'heures des aidants non-cohabitants et à 24% pour le nombre d'heures des aidants cohabitants. Cependant, la concordance entre la déclaration de l'aidé et celle de l'aidant ne va pas de soi a priori. La pertinence de cette complétion est donc à remettre en cause.

Une étude approfondie de la concordance confirme que l'aidé et l'aidant se réfèrent aux mêmes types d'aide dispensées, mais que leur perception du volume horaire diffère.

Compte tenu de ces résultats, il ne semble pas souhaitable d'imputer les valeurs de HSA dans HSM. Cela pourrait créer un nombre important d'erreurs sans qu'il soit impossible de les détecter. Il est donc préconisé de séparer les analyses « du point de vue des aidés » (HSM) et « du point de vue des aidants » (HSA).

Les premières pistes s'avérant sans issue, le traitement de la non-réponse est fait par imputation.

3.2. Caractéristiques des non-répondants

Cette analyse est réalisée sur les relations d'aide où :

- l'aidé est âgé de 60 ans ou plus,
- la valeur du nombre d'heures non-recodée est ., 99 ou 999 (non-réponse).

Une fois la distinction cohabitants / non-cohabitants prise en compte, il n'y a pas de différence entre les répondants et les non-répondants (voir Tableau 2). Un modèle logistique univarié expliquant la non-réponse par chacune des caractéristiques est testé pour les non-cohabitants et pour les cohabitants séparément. Pour les non-cohabitants, la fréquence (classes), le degré de dépendance (discret), l'ordre de citation de l'aidant (discret) apparaissent significatifs, mais aucune tendance ne peut être dégagée. L'âge de l'aidant (continu) est également significatif, les aidants étant un peu plus âgés chez les non-répondants. Pour les cohabitants, l'âge de l'aidé (continu) et l'âge de l'aidant (continu) sont significatifs, les aidés et les aidants étant un peu plus âgés chez les non-répondants.

Tableau 2. Caractéristiques des répondants / non-répondants HSM

	Non-cohabitants		Cohabitants	
	Répondants	Non-répondants	Répondants	Non-répondants
Nombre d'aidants	1980	526	1468	860
Proportion de				
Aide plus d'une fois par semaine	65	65	NA	NA
Age de l'aidé ≥ 75 ans	70	73	54	60
Femmes aidées	75	71	59	59
Dépendants (GIR < 4)	25	20	29	31
Aidés par un professionnel	55	55	36	40
Aidés par famille proche (conjoint/enfant/fratrie/parent)	58	63	93	93
Aidants ≥ 60 ans	26	29	67	74
Femmes aidantes	69	65	53	55
Aidant cité en premier ou deuxième	80	73	98	97

Champ : Personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs proche(s).

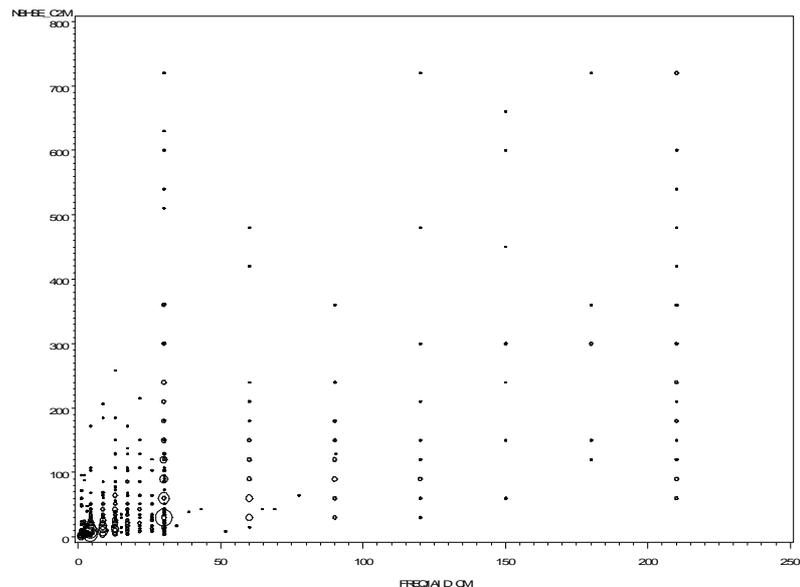
Source : Enquête Handicap-Santé 2008, volet ménages, Insee.

Note de lecture : 65% des personnes âgées aidées par un proche non-cohabitant et répondant à la question du nombre d'heures d'aide sont aidées plus d'une fois par semaine.

La non-réponse chez les cohabitants peut s'expliquer par le fait qu'il soit difficile de chiffrer le nombre d'heures d'aide lorsque l'aidant (en majorité des conjoints) cohabite. Une première solution serait alors d'étudier uniquement le volume d'aide chez les non-cohabitants, en essayant de prédire le nombre d'heures par la fréquence (et d'autres facteurs).

Cependant, les nombres d'heures sont très dispersés pour chaque fréquence (voir Graphique 4). Une régression linéaire expliquant le nombre d'heures mensuelles par la fréquence mensuelle et l'âge de l'aidant donne un R^2 de 0,37 seulement.

Graphique 4. Nombre d'heures d'aide mensuel selon la fréquence mensuelle chez les non-cohabitants



Champ : Personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs proche(s) non cohabitant(s).

Source : Enquête Handicap-Santé 2008, volet ménages, Insee.

NB : La taille des points correspond à la fréquence de la valeur.

Contrairement à ce que l'on pouvait attendre, la fréquence des visites est peu prédictive du nombre d'heures d'aide. L'imputation du nombre d'heures ne peut pas se baser uniquement sur la fréquence, et un nombre important de variables entrent potentiellement en jeu dans la détermination du volume horaire de l'aide. L'imputation multiple permet de prendre en compte facilement l'influence de ces nombreuses variables et de mieux estimer l'incertitude, ce qui paraît important face au taux élevé de non-réponse.

4. L'imputation multiple

4.1. Théorie

4.1.1. Principe général

L'imputation multiple est apparue dans les travaux de Rubin il y a trente ans [2]. Initialement, la méthode a été développée pour répondre au problème des données manquantes dans les enquêtes par sondage [3].

Lors de l'imputation de données manquantes, il est important de tenir compte de l'information contenue dans les données observées, à partir du moment où il y a une association entre les variables explicatives et la variable à imputer. L'imputation par la régression permet notamment cela, et remplace alors chaque valeur manquante par une valeur plausible. Des données complètes sont restituées et l'analyse peut s'effectuer sur ces données complétées. Mais lors de cette analyse, les données imputées sont traitées comme si elles étaient observées. Ceci est peu satisfaisant, dans la mesure où la vraie valeur d'une donnée manquante n'est pas connue et qu'on ne fait que l'estimer. Ces méthodes sur-estiment alors la précision de l'étude. Pour pallier ce problème, l'imputation multiple a été développée. En remplaçant chaque valeur manquante par plusieurs valeurs plausibles, elle permet de prendre en compte l'incertitude autour de la valeur imputée, et se révèle alors plus efficace que l'imputation simple en ce qui concerne l'estimation de la variance [4].

De plus, l'imputation multiple est implémentée dans les logiciels statistiques standards (tels que SAS par exemple) et est donc facile à mettre en œuvre.

Le principe général de l'imputation multiple est de remplacer chaque valeur manquante par plusieurs valeurs plausibles. L'imputation de plusieurs valeurs a pour but de refléter l'incertitude concernant la valeur imputée : les valeurs respectives des valeurs manquantes ne sont pas connues précisément, seule la distribution des valeurs manquantes est connue. Afin d'imputer des valeurs plausibles, l'imputation multiple doit se baser sur un modèle qui relie les valeurs observées aux valeurs manquantes.

L'imputation multiple suppose que les données sont manquantes au hasard. C'est-à-dire que, conditionnellement aux données observées, la probabilité qu'une donnée soit manquante ne dépend pas des données non-observées, qu'elles soient des valeurs manquantes des données ou des variables non-observées.

4.1.2. Estimation

L'imputation multiple ne nécessite pas un grand nombre de répétitions [5]. Cinq imputations suffisent dans la plupart des cas. Dix ou vingt imputations peuvent être effectuées pour s'assurer de l'efficacité de l'imputation, notamment lorsque la proportion d'information manquante est élevée et que le nombre d'observations permet à la procédure de tourner assez rapidement. Il n'y a pas d'intérêt à effectuer plus de vingt imputations. Par la suite, le nombre d'imputations est désigné par M .

L'imputation multiple se décompose en trois phases :

1. Les données manquantes sont remplies M fois pour générer M jeux de données complétés.
2. Les M jeux de données complétés sont analysés séparément pour produire M estimations.
3. Les M estimations sont combinées pour obtenir une estimation finale unique.

4.1.2.1. Imputation

Pour mener à bien la première phase, l'imputation multiple utilise la distribution conditionnelle des observations manquantes sachant les données observées. Les valeurs imputées représentent l'information sur les données manquantes contenue dans les données observées, pour le modèle spécifié.

Les données sont désignées par une matrice Y , constituée d'une partie observée Y^{obs} et d'une partie manquante Y^{miss} . Par ailleurs, Y est distribuée selon la fonction $f(Y|\phi)$, où ϕ désigne l'ensemble des paramètres du modèle d'imputation. Supposant que les données sont manquantes au hasard, la distribution prédictive $f(Y^{miss}|Y^{obs})$ est donnée par :

$$\begin{aligned} f(Y^{miss}|Y^{obs}) &= \int f(Y^{miss}, \phi|Y^{obs}) d\phi \\ &= \int f(Y^{miss}|Y^{obs}, \phi) f(\phi|Y^{obs}) d\phi \end{aligned}$$

Les données manquantes sont imputées en deux étapes. Tout d'abord, une valeur pour les paramètres est simulée d'après leur distribution a posteriori observée sur les données $f(\phi|Y^{obs})$. Les chaînes de Markov sont souvent utilisées pour cela. Cette première étape permet d'obtenir une variabilité dans les paramètres. Ensuite, les données manquantes sont simulées d'après leur distribution conditionnelle a posteriori $f(Y^{miss}|Y^{obs}, \phi)$, en utilisant la valeur de ϕ générée à la première étape. Par exemple, les données manquantes sont générées d'après un modèle de régression de covariables Y^{obs} et de coefficients ϕ [6]. Pour imputer M valeurs, on tire M fois dans la distribution prédictive a posteriori. Les valeurs manquantes de chaque individu sont imputées selon ses propres valeurs observées, en ajoutant un bruit aléatoire afin de préserver une variabilité au sein des données imputées [7].

Par rapport à une imputation par la régression, l'imputation multiple considère que les paramètres de la régression sont des variables et que leur estimation est issue d'une distribution. Premièrement, une droite de régression est tirée dans sa distribution estimée. Deuxièmement, un résidu est tiré dans la distribution normale estimée des résidus et ajouté à la droite de régression tirée pour obtenir la valeur imputée. La première étape conduit à ce que la droite de régression qui sert à l'imputation s'écarte légèrement de la droite de régression moyenne. La deuxième étape se traduit par des valeurs imputées qui s'écartent de leur droite de régression.

4.1.2.2. Analyse

Une fois les données manquantes imputées, l'analyse prévue pour obtenir le paramètre d'intérêt peut être effectuée normalement : calcul d'une moyenne, de paramètres de régression... Mais elle est effectuée M fois, une fois sur chaque jeu de données complétées.

La deuxième phase souligne un avantage majeur de l'imputation multiple : une fois les données imputées, les analyses statistiques standards peuvent être utilisées. Chaque jeu de données complété peut être analysé avec la méthode prévue pour des données complètes. Cette phase évoque également le domaine d'application originel de l'imputation multiple. En effet, l'intérêt de Rubin se portait sur de grandes enquêtes destinées à créer des bases de données publiques, ensuite analysées par de nombreux et différents utilisateurs. La séparation entre la première phase d'imputation et la deuxième phase d'analyse permet d'envisager que tous les utilisateurs travaillent sur les mêmes jeux de données complétés, imputés préalablement par une seule et même personne.

4.1.2.3. Combinaison des estimations

La troisième phase s'effectue selon les règles suivantes (règles de Rubin³). Soit Q le paramètre à estimer (moyenne, coefficient de régression...) et U sa variance. L'imputation de M valeurs conduit à M estimations de Q et U , que l'on note \hat{Q}^m et \hat{U}^m pour $m = 1, \dots, M$.

L'estimateur combiné \hat{Q}^* est la moyenne des M estimations :

$$\hat{Q}^* = \frac{1}{M} \sum_{m=1}^M \hat{Q}^m$$

La variance combinée \hat{U}^* est composée de deux parties. La variance intra-imputation \bar{U} rend compte des variances de chacun des \hat{Q}^m , elle est estimée par la moyenne des M variances :

$$\bar{U} = \frac{1}{M} \sum_{m=1}^M \hat{U}^m$$

La variance inter-imputation B rend compte de la variance de chacun des \hat{Q}^m par rapport à l'estimateur combiné. Elle correspond à la variance des moyennes a posteriori des \hat{Q}^m :

$$B = \frac{1}{M-1} \sum_{m=1}^M (\hat{Q}^m - \hat{Q}^*)^2$$

La variance combinée \hat{U}^* est la somme pondérée de la variance intra-imputation et de la variance inter-imputation :

$$\hat{U}^* = \bar{U} + \left(1 + \frac{1}{M}\right) B$$

Une approximation de Student est utilisée pour les tests et les intervalles de confiance : $(\hat{Q}^* - Q) / \sqrt{\hat{U}^*} \sim t_v$, où le degré de liberté vaut :

$$v = (M-1) \left[1 + \frac{\bar{U}}{(1+1/M)B} \right]^2$$

4.1.3. Sélection des variables du modèle d'imputation

La validité de l'imputation multiple dépend de la manière dont les valeurs imputées sont estimées. La sélection des variables à inclure dans le modèle d'imputation est une étape importante.

Le modèle d'imputation doit rendre compte des relations entre les variables mesurées et la variable à imputer, mais également chercher à expliquer les données manquantes. Il doit donc inclure toutes les variables potentiellement liées à la variable à imputer, ainsi que toutes les variables potentiellement liées au fait que la donnée soit manquante [8].

De manière générale, la perte de précision liée à une augmentation du nombre de variables dans le modèle est négligeable devant l'assurance d'estimations correctes lors de la phase d'analyse. Cependant, il faut bien entendu veiller à ce que le modèle reste identifiable, notamment en limitant le nombre de variables par rapport au nombre d'observations. Si en théorie il ne semble pas nécessaire d'inclure plus de 25 variables dans le modèle d'imputation [9], en pratique les jeux de données permettent généralement d'en sélectionner une dizaine [10].

³ Les règles de Rubin [3] pour la combinaison des estimations sont présentées ici dans le cas d'un scalaire (un seul paramètre d'intérêt). La combinaison des estimations pour un vecteur de paramètres a été développée par Schafer [8](pages 112-118).

En conclusion, il est recommandé d'inclure toute l'information disponible afin de s'assurer que les imputations multiples ont un biais minimal et une certitude maximale. En incluant le plus de covariables possibles, on augmente également les chances de se placer dans le cadre de données manquantes au hasard [11]. En effet, plus les données observées contiennent d'information pour prédire les données manquantes, moins la distribution conditionnelle du processus de données manquantes dépend des données manquantes, et donc plus on s'approche de données manquantes au hasard [8].

4.2. Recherche des variables explicatives

4.2.1. Facteurs expliquant la non-réponse

Dans un premier temps, des régressions logistiques univariées expliquant la non-réponse sont ajustées pour un grand nombre de variables HSM portant sur l'aide reçue par la personne âgée, sa santé, ses limitations fonctionnelles, ses restrictions d'activité, son réseau familial, ses revenus et les allocations qu'elle perçoit.

Parmi l'ensemble des variables significatives, les variables qui semblent pertinentes ont été conservées :

- Age de l'aidé (6 classes)
- Besoin d'aide humaine supplémentaire pour les ADL ou les IADL
- Ancienneté de l'aide de l'entourage
- Limité à cause d'un problème de santé
- Difficultés psychologiques
- État de santé général (très bon ou bon/assez bon/mauvais/très mauvais)
- Difficultés à se rappeler des activités
- Diplôme le plus élevé (7 classes)
- Pas de temps de la fréquence de l'aide de l'entourage (jour/semaine/mois)
- Rang de l'aidant (premier cité/deuxième cité/troisième et plus)
- Age de l'aidant (5 classes)
- Cohabitation de l'aidant avec l'enquêté
- Relation de l'aidant avec l'enquêté (conjoint/enfant-parent/autre famille/autre)
- Nombre d'aides à la vie quotidienne (4 classes)
- Soutien moral
- Sexe de l'aidant
- Trois indicateurs d'aide à la vie quotidienne (5 classes d'aide selon le type et le nombre de tâches effectuées)
- Lieu de naissance (France métropolitaine/Autres)
- Nombre d'aidants professionnels et informels (4 classes)
- Nombre de personnes dans le ménage (4 classes)
- Aide de professionnels
- Rémunération de l'aide
- Répondant au questionnaire (la personne seule/la personne aidée d'un tiers/un tiers)
- Au moins une restriction d'activité forte
- Gérer ses ressources seul
- Revenu mensuel moyen (5 classes)
- Sexe de l'aidé
- Degré de dépendance (GIR).

Les variables « Revenu mensuel moyen » et « Gérer ses ressources seul » seront par la suite supprimées de l'analyse, car possédant un nombre important de données manquantes, elles conduisent à exclure trop d'individus de l'analyse.

Dans un deuxième temps, une régression logistique multivariée inclut les variables précédemment triées et une sélection de modèle pas-à-pas permet d'optimiser le modèle. La sélection du modèle est d'abord effectuée sur toute la population. Les variables sélectionnées sont ensuite introduites (forcées) dans le modèle stratifié sur l'âge de l'aidé (< 60 ans / ≥ 60 ans). La stratification sur l'âge permet d'introduire la mesure du degré de dépendance pour les plus âgés (variable non pertinente pour les moins de 60 ans). Enfin, la population est de nouveau stratifiée selon que l'aidant cohabite ou non avec l'aidé, en forçant les variables précédemment sélectionnées. Les résultats de la sélection de modèle sont donnés dans le tableau 3.

Tableau 3. Sélection du modèle multivarié expliquant la non-réponse

Variables explicatives	Age < 60		Age ≥ 60	
	Non-cohabitant	Cohabitant	Non-cohabitant	Cohabitant
Age de l'aidé	x	x	x	x
Besoin d'aide humaine	x	x	x	x
Ancienneté de l'aide		x		
Limité (pb de santé)	x	x	x	x
Difficultés psychologiques	x	x	x	x
État de santé général	x	x	x	x
Difficultés à se rappeler				
Diplôme le plus élevé	x	x		
Fréquence de l'aide par...	x	NA	x	NA
Rang de l'aidant	x	x	x	x
Age de l'aidant	x	x	x	x
Cohabitation de l'aidant	NA	NA	NA	NA
Relation avec l'aidant				
Nombre d'AVQ	x	x	x	x
Soutien moral	x	x	x	x
Sexe de l'aidant				
Indicateur d'aide à la VQ 1			x	x
Indicateur d'aide à la VQ 2			x	x
Indicateur d'aide à la VQ 3	x	x		
Lieu de naissance	x	x	x	x
Nombre d'aidants	x	x	x	x
Taille du ménage				
Aide de professionnels	x	x	x	x
Rémunération de l'aide	x	x	x	x
Répondant				
Restriction d'activité forte				
Sexe de l'aidé	x		x	
GIR de l'aidé	NA	NA		
% de concordance	72.4	62.7	66.0	61.6

NB : les variables sélectionnées sont désignées par un « x ».

4.2.2. Facteurs expliquant le nombre d'heures

Le nombre d'heures d'aide mensuelle est une variable de comptage, dont la distribution est poissonnienne. Une transformation logarithme lui est donc appliquée pour s'approcher de la normalité.

La sélection de modèle a été effectuée directement sur les sous-populations stratifiées de la même manière que précédemment (âge de l'aidé et cohabitation). L'inclusion de variables explicatives qualitatives a nécessité l'utilisation de la PROC GLMSELECT. Le R² ajusté a été choisi comme critère de sélection des modèles pour la procédure stepwise (option *select=adjrsq*). L'entrée et la sortie de variables ont été mises en compétition dans le processus de sélection (option *drop=competitive*).

Tableau 4. Sélection du modèle multivarié expliquant le nombre d'heures

Variables explicatives	Age < 60		Age ≥ 60	
	Non-cohabitant	Cohabitant	Non-cohabitant	Cohabitant
Age de l'aidé			x	x
Besoin d'aide humaine		x		
Ancienneté de l'aide		x		x
Limité (pb de santé)		x		
Difficultés psychologiques			x	x
État de santé général		x	x	x
Difficultés à se rappeler	x			x
Diplôme le plus élevé		x	x	
Fréquence de l'aide	x	NA	x	NA
Fréquence de l'aide par...	x	NA	x	NA
Rang de l'aidant	x	x		x
Age de l'aidant	x	x	x	
Cohabitation de l'aidant	NA	NA	NA	NA
Relation avec l'aidant	x	x		x
Nombre d'AVQ	x	x		x
Soutien moral	x	x	x	x
Sexe de l'aidant	x	x	x	x
Indicateur d'aide à la VQ 1		x	x	
Indicateur d'aide à la VQ 2	x	x		x
Indicateur d'aide à la VQ 3	x	x	x	x
Lieu de naissance	x	x	x	x
Nombre d'aidants	x	x	x	x
Taille du ménage	x		x	
Aide de professionnels	x			x
Rémunération de l'aide			x	x
Répondant	x	x	x	x
Restriction d'activité forte		x		x
Sexe de l'aidé	x	x		x
GIR de l'aidé	NA	NA	x	x
R ²	0.7281	0.3574	0.6996	0.3334

NB : les variables sélectionnées sont désignées par un « x ».

4.2.3. Modèle final d'imputation

L'imputation multiple a été réalisée uniquement sur la population des personnes âgées de 60 ans ou plus. Des modèles différents ont été ajustés selon que l'aidant est cohabitant ou non.

Les variables expliquant la non-réponse ou le nombre d'heures sont incluses dans le modèle d'imputation. L'imputation multiple est réalisée via l'algorithme MCMC avec chaînes multiples, en incluant les variables qualitatives sous forme d'indicatrices. La variable « nombre d'heures » est log-transformée avant l'imputation et la transformation inverse est appliquée à la sortie. Vingt imputations sont effectuées. Une valeur maximale est imposée à 24 heures d'aide par jour.

4.3. Application sous SAS

L'implémentation de l'imputation multiple a été réalisée à l'aide des procédures MI et MIANALYZE de SAS.

Lors de l'utilisation de l'imputation multiple sous SAS, la procédure MI est la première à être appelée : elle effectue les imputations. Pour l'imputation de la variable nombre d'heures (NBH), le code suivant a été utilisé :

```
proc mi data=matable nimpute=20 seed=123456 out=tableimp maximum=720 . ;
  mcmc chain=multiple;
  var NBH VAR1 VAR2 VAR3;
  transform log(NBH);
run;
```

Le nombre d'imputations est défini par l'option « nimpute » ; par défaut, cinq imputations sont effectuées. L'option « seed » sert à spécifier le germe, c'est-à-dire le nombre qui initialise le générateur de nombres aléatoires. Par défaut, un nombre est défini à partir de l'heure du jour de l'horloge de l'ordinateur. Cependant, afin de répliquer des résultats sous des situations identiques, il est nécessaire de spécifier un même germe. L'instruction « var » définit l'ordre des variables ; la ou les variables à imputer sont à placer en début de liste. La table « tableimp » est récupérée en sortie. Elle contient les *M* fichiers de données imputés, identifiés par la variable « _Imputation_ ».

Une fois les imputations effectuées, l'analyse peut avoir lieu. Elle doit s'effectuer une fois par imputation. Pour l'estimation d'une moyenne, on utilisera le code suivant :

```
proc means data=tableimp ;
  var NBH ;
  by _Imputation_ ;
  output out=tablemoy mean=moy stderr=var ;
run ;
```

C'est l'instruction « by _Imputation_ » qui permet une analyse séparée pour chaque jeu de données imputé. La table « tablemoy » est récupérée en sortie ; elle contient la moyenne et son écart-type estimés pour chaque jeu de données imputé.

La dernière étape consiste à combiner ces estimations. Cela se fait à l'aide de la procédure MIANALYZE. Pour l'exemple utilisé ci-dessus, on utilise le code suivant :

```
proc mianalyze data=tablemoy ;
  modeleffects moy ;
  stderr var ;
run ;
```

Le paramètre d'intérêt est placé dans l'instruction « modeleffects » et son écart-type dans l'instruction « stderr ». Dans le cas de l'estimation de plusieurs paramètres, il faut spécifier les matrices de variance-covariance.

5. Résultats

5.1. Résultats de l'imputation

En moyenne, les résultats avec imputation multiple diffèrent peu des résultats sans imputation. En effet, les répondants ne présentaient pas un profil particulier de nombre d'heures d'aide. Cependant, l'imputation multiple a permis d'évaluer l'incertitude due à la non-réponse et de la chiffrer via l'intervalle de confiance. Cette incertitude touche particulièrement les nombres d'heures des aidants cohabitants, reflet à la fois d'un taux de non-réponse plus élevé et d'un modèle de prédiction moins performant.

5.2. Résultats de l'étude

Les deux tiers (66 %) des personnes âgées aidées par des professionnels reçoivent leur aide plusieurs fois par semaine : 30 % la reçoivent au moins une fois par jour et 36 % moins d'une fois par jour. Les trois quarts (77 %) des personnes âgées les plus dépendantes (GIR 1 et 2) aidées par un ou des professionnels, le sont au moins une fois par jour, 67 % des personnes âgées en perte d'autonomie plus modérée classées en GIR 3 sont dans ce cas, ainsi que 37 % de celles classées en GIR 4 et 17 % des personnes âgées les moins dépendantes (GIR 5 et 6).

La moitié des personnes âgées aidées par des professionnels, avec ou sans aide de l'entourage, reçoivent leur aide pendant au moins l'équivalent⁴ de 35 minutes par jour (voir tableau 6). Pour les personnes âgées très dépendantes (GIR 1 et 2), ce volume d'aide professionnelle médian⁵ équivaut à 2 heures 10 par jour, contre 25 minutes par jour pour les moins dépendantes (GIR 5 et 6). Le volume horaire d'aide des professionnels est également plus important lorsque l'aide est mixte : le niveau médian équivaut à 25 minutes d'intervention par jour pour les personnes âgées exclusivement aidées par des professionnels passe à 40 minutes par jour pour celles également aidées par leur entourage, l'aide mixte étant généralement associée à une perte d'autonomie plus importante.

Parmi les personnes âgées aidées par un ou plusieurs proches non cohabitants pour la vie quotidienne, les deux tiers (68 %) reçoivent leur aide plusieurs fois par semaine (41 % au moins une fois par jour et 27 % moins d'une fois par jour). Les personnes âgées les plus dépendantes (GIR 1 et 2) aidées par un ou plusieurs proches non cohabitants sont 74 % à recevoir une aide quotidienne, 61 % des personnes classées en GIR 3 sont dans ce cas et 41 % de celles classées en GIR 4. Un tiers (35 %) des personnes âgées les plus autonomes (GIR 5 et 6) reçoivent également cette aide au moins une fois par jour.

La moitié des personnes âgées aidées par leur entourage pour les tâches de la vie quotidienne le sont au moins l'équivalent d'une heure 40 minutes par jour (voir tableau 6). La durée de cette aide augmente avec leur degré de dépendance. La moitié des personnes âgées très dépendantes (GIR 1 et 2) aidées par un ou plusieurs proches reçoivent ainsi leur aide pendant au moins l'équivalent de 5 heures 15 par jour, alors que la moitié des moins dépendantes (GIR 5 et 6) la reçoivent durant au moins l'équivalent d'une heure 10 par jour. Le volume médian de l'aide apportée par les aidants cohabitants est quant à lui deux fois plus important que celui des aidants non cohabitants (respectivement l'équivalent de 2 heures 10 et d'une heure par jour).

⁴ Le nombre d'heures d'aide est rapporté en heures par jour, mais la fréquence des visites n'est pas forcément journalière. Par exemple, 14 heures d'aide par semaine équivalent à 2 heures par jour, mais l'aide peut être dispensée en deux visites de 7 heures (soit 2 fois par semaine).

⁵ La médiane est ici préférée à la moyenne, car les valeurs extrêmes influencent fortement la moyenne (des personnes peuvent en effet recevoir une aide 24 heures sur 24).

Tableau 5. Nombre moyen d'heures d'aide par jour fournie aux personnes âgées pour la vie quotidienne par les aidants l'entourage, avec et sans imputation, selon le degré de dépendance

	Aidants non cohabitants			Aidants cohabitants		
	% de données manquantes	Moyenne sans imputation (répondants)	Moyenne avec imputation multiple	% de données manquantes	Moyenne sans imputation (répondants)	Moyenne avec imputation multiple
Tous	20,3	2h00 [1h50 – 2h05]	2h00 [1h50 – 2h10]	37,2	3h35 [3h20 – 3h50]	3h40 [3h30 – 3h55]
GIR 1	20,7	7h25 [3h20 - 11h25]	6h45 [3h40 - 9h50]	33,5	7h55 [6h20 - 9h40]	7h45 [6h15 - 9h20]
GIR 2	17,5	4h35 [3h50 - 5h25]	4h25 [3h40 - 5h10]	38,0	7h40 [6h50 - 8h30]	7h20 [6h35 - 8h05]
GIR 3	15,4	3h15 [2h45 - 3h50]	3h25 [2h50 - 4h00]	34,1	5h35 [4h50 - 6h25]	5h30 [4h45 - 6h15]
GIR 4	18,7	2h15 [1h45 - 2h45]	2h15 [1h45 - 2h40]	44,7	3h20 [2h55 - 3h45]	3h30 [2h55 - 4h00]
GIR 5	18,5	1h20 [1h10 - 1h35]	1h25 [1h10 - 1h35]	32,0	3h10 [2h35 - 3h40]	3h15 [2h40 - 3h50]
GIR 6	22,5	1h30 [1h20 - 1h40]	1h30 [1h20 - 1h45]	37,5	2h25 [2h10 - 2h45]	2h40 [2h20 - 3h05]

NB : Nombre moyen d'heures, suivi (entre crochets) de l'intervalle de confiance à 95%

Champ : Personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un ou plusieurs proche(s).

Source : Enquête Handicap-Santé 2008, volet ménages, Insee.

Note de lecture : en moyenne, les personnes âgées aidées à domicile par au moins un aidant non cohabitant sont aidées l'équivalent de 2 heures par jour.

Tableau 6. Nombre médian d'heures d'aide par jour fournie aux personnes âgées pour la vie quotidienne par les professionnels et par l'entourage, selon le degré de dépendance

	Aide des professionnels	Aide de l'entourage	Aide de l'entourage non-cohabitant	Aide de l'entourage cohabitant	Au moins une aide des professionnels ou de l'entourage
Tous	0h35 [0h15 – 1h15]	1h40 [0h35 – 4h00]	1h00 [0h25 - 2h00]	2h10 [0h55 - 4h30]	1h20 [0h25 – 3h30]
GIR 1-2	2h10 [1h00 – 4h00]	5h15 [2h45 – 10h05]	3h00 [1h05 - 5h45]	5h40 [2h55 - 10h10]	7h35 [4h00 – 12h35]
GIR 3	1h25 [0h45 – 2h20]	4h00 [2h00 – 6h30]	2h00 [1h00 – 4h35]	4h15 [2h15 – 6h45]	4h15 [2h25 – 7h30]
GIR 4	0h50 [0h25 – 1h35]	2h00 [0h55 – 4h20]	1h00 [0h25 - 2h40]	2h20 [1h05 - 4h30]	2h05 [1h00 – 4h20]
GIR 5-6	0h25 [0h15 – 0h50]	1h10 [0h25 – 2h55]	0h50 [0h20 - 1h45]	1h40 [0h40 - 3h30]	0h55 [0h25 – 2h05]

NB : Nombre médian d'heures, suivi (entre crochets) de l'intervalle interquartile

Champ : Personnes âgées de 60 ans ou plus vivant à domicile et aidées régulièrement pour des raisons de santé ou de handicap dans les tâches de la vie quotidienne par un proche ou un professionnel.

Source : Enquête Handicap-Santé 2008, volet ménages, Insee.

Note de lecture : 50% des personnes âgées aidées par des professionnels reçoivent au moins l'équivalent de 35 minutes d'aide de professionnels par jour, 25% en reçoivent moins de 15 minutes par jour, 25% en reçoivent plus d'1h15 par jour. 50% des personnes âgées aidées par un ou plusieurs proche(s) pour les tâches de la vie quotidienne reçoivent au moins l'équivalent d'1h40 d'aide de ces proches par jour, 25% en reçoivent moins de 35 minutes par jour, 25% en reçoivent plus de 4h par jour. 50% des personnes âgées qui reçoivent une aide à la vie quotidienne par un (des) proche(s) ou un (des) professionnel(s) reçoivent au moins l'équivalent d'1h20 d'aide par jour, 25% en reçoivent moins de 25 minutes par jour, 25% en reçoivent plus de 3h30 par jour.

Conclusion

Dans l'enquête Handicap-Santé, la non-réponse à la question du nombre d'heures d'aide est importante et handicape l'analyse. L'imputation multiple a permis de prendre en compte une grande quantité d'information disponible et d'estimer au mieux l'incertitude provoquée par ces données manquantes. Dans le cas d'enquête en population générale où de nombreuses variables auxiliaires sont mobilisables, l'imputation multiple est indiquée pour traiter la non-réponse, d'autant qu'elle est facilement mise en œuvre via les logiciels statistiques standards.

Bibliographie

- [1] SOULLIER N. (2011). L'implication de l'entourage et des professionnels auprès des personnes âgées à domicile. *Études et Résultats*, **771**.
- [2] RUBIN D.B. (1978). Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse. *Proceedings of the Survey Research Methods Section (American Statistical Association)* : 20-34.
- [3] RUBIN D.B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons Inc., New York.
- [4] VAN DER HEIJDEN G.J., DONDEERS A.R., STIJNEN T., et MOONS K.G. (2006). Imputation of missing values is superior to complete case analysis and the missing indicator method in multivariable diagnostic research : a clinical example. *Journal of Clinical Epidemiology*, **59**(10) : 1102-1109.
- [5] RUBIN D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, **91**(434) : 473-489.
- [6] SINHARAY S., STERN H.S., et RUSSELL D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods*, **6**(4) : 317-329.
- [7] SCHAFFER J.L. et GRAHAM J.W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, **7**(2) : 147-177.
- [8] SCHAFFER J.L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC, Boca Raton: FL.
- [9] VAN BUUREN S., BOSHUIZEN H.C., et KNOOK D.L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, **18**(6) : 681-694.
- [10] WOOD A.M., WHITE I.R., HILLSDON M., et CARPENTER J. (2005). Comparison of imputation and modelling methods in the analysis of a physical activity trial with missing outcomes. *International Journal of Epidemiology*, **34**(1) : 89-99.
- [11] RUBIN D.B., STERN H.S., et VEHOVAR V. (1995). Handling "don't know" survey responses: the case of the slovenian plebiscite. *Journal of the American Statistical Association*, **90**(431) : 822-828.

Annexe 1 : La grille AGGIR (Autonomie Gérontologie Groupes Iso-Ressources)

Chaque personne interrogée dans le volet ménages de l'enquête Handicap-Santé s'est vue attribuer une « estimation-GIR » déterminée sur la base des réponses aux questions HSM et en s'appuyant sur la méthode d'attribution du GIR explicitée dans l'article R 232-3 du Code de l'action sociale et des familles, méthode qui s'impose aux évaluateurs des conseils généraux. La détermination du GIR s'appuie sur dix variables d'activité corporelle et mentale. Cette estimation GIR a été calculée pour toutes les personnes enquêtées âgées de 60 ans ou plus, que celles-ci aient ou non déclaré percevoir l'allocation personnalisée d'autonomie (APA). Le GIR attribué n'est donc pas ici le résultat de l'évaluation de professionnels ; il doit être considéré comme un indicateur statistique du degré de dépendance de la personne.

- Le GIR 1 correspond aux personnes âgées confinées au lit ou au fauteuil et dont les fonctions mentales sont gravement altérées, qui nécessitent une présence indispensable et continue d'intervenants.

- Le GIR 2 comprend deux groupes de personnes âgées : celles qui sont confinées au lit ou fauteuil, dont les fonctions mentales ne sont pas totalement altérées et qui nécessitent une prise en charge pour la plupart des activités de la vie courante et celles dont les fonctions mentales sont altérées mais qui ont conservé leurs capacités de se déplacer.

- Le GIR 3 correspond pour l'essentiel, aux personnes âgées ayant conservé leurs fonctions mentales, partiellement leur capacité à se déplacer mais qui nécessitent plusieurs fois par jour des aides pour leur autonomie corporelle. La majorité d'entre elles n'assument pas seules l'hygiène de l'élimination anale et urinaire.

- Le GIR 4 comprend essentiellement deux groupes de personnes. D'une part celles qui n'assument pas seules leurs transferts mais qui, une fois levées, peuvent se déplacer à l'intérieur du logement. Elles doivent parfois être aidées pour la toilette et l'habillement. La grande majorité d'entre elles s'alimentent seules. D'autre part celles qui n'ont pas de problèmes pour se déplacer mais qu'il faut aider pour les activités corporelles et les repas.

- Le GIR 5 correspond aux personnes qui assurent seules leurs déplacements à l'intérieur de leur logement, s'alimentent et s'habillent seules. Elles nécessitent une aide ponctuelle pour la toilette, la préparation des repas et le ménage.

- Le GIR 6 regroupe toutes les personnes qui n'ont pas perdu leur autonomie pour les actes discriminants de la vie courante.