

# **DICTIONNAIRE DES VARIABLES SHARE - VAGUE 2**

**Dernière mise à jour : Octobre 2013**

<b>AVERTISSEMENTS AUX UTILISATEURS DE CE DICTIONNAIRE .....</b>	<b>3</b>
Données brutes.....	3
Spécificités nationales.....	3
<b>GUIDE D'UTILISATION .....</b>	<b>4</b>
Objectif .....	4
Éléments de contexte vague 2 .....	4
Contenu des fichiers de données.....	5
Comment se présente l'information ? .....	6
<b>REMARQUES METHODOLOGIQUES .....</b>	<b>10</b>
Données financières.....	10
Répondant.....	10
Filtres .....	11
Code de variable .....	12
Boucle .....	13
Cahier des cartes-codes .....	13

# AVERTISSEMENTS AUX UTILISATEURS DE CE DICTIONNAIRE

## Données brutes

Les chiffres présentés ici sont destinés à fournir une première information descriptive aux utilisateurs de l'enquête SHARE. Il s'agit d'informations statistiques brutes sans application d'aucune pondération. Ces informations ne sont donc pas représentatives des différentes populations concernées et ne peuvent être utilisées pour effectuer des interprétations sur ces populations.

Le seul traitement appliqué aux données brutes est la procédure SHARETOM, développée par l'équipe de coordination européenne à Mannheim. Ce programme automatique (`sharetom.ado`) est téléchargeable avec les données et son utilisation est fortement recommandée. Il permet de transformer certaines valeurs associées aux valeurs manquantes (-1 pour « Ne sait pas » par exemple) en codes qui ne rentrent donc pas dans le calcul des statistiques (.a pour « Ne sait pas »).

Par ailleurs, les quantiles concernant les variables quantitatives ont été arrondis au nombre entier le plus proche, à l'exception des variables `ws011` et `ws013` dont les statistiques sont affichées avec deux décimales.

Un premier nettoyage des données a déjà été effectué au niveau européen. Néanmoins, quelques erreurs peuvent subsister. En plus de ce nettoyage, certaines variables ont été corrigées pour ce *codebook* (`pf003` par exemple). Une remarque a été ajoutée sous les tableaux concernés.

## Spécificités nationales

Les données issues de SHARE sont harmonisées *ex-ante* et *ex-post* (comme pour les données financières converties en euro par exemple). Néanmoins, du fait de la diversité des systèmes nationaux, certaines questions portant sur les prestations ou aides perçues, les types de contrat de travail ou encore sur les études suivies fournissent des résultats non comparables sans traitement supplémentaire.

Les résultats présentés dans ce document proviennent des données brutes sans distinction des spécificités nationales puisque les statistiques sont présentées tous pays confondus. Ainsi, pour certaines questions, les réponses sont présentées sous l'appellation générique « Catégorie spécifique pays ». Il est alors indispensable de se reporter au site de SHARE (rubrique *Item Correspondence*) pour l'analyse des données.

Par ailleurs, les libellés utilisés dans ce document proviennent de la version française des questions et réponses telles qu'elles ont été utilisées en France. Les catégories spécifiques à la France subsistent donc dans ce document (par exemple RMI ou CDD).

# GUIDE D'UTILISATION

Pour des informations détaillées sur la méthodologie et l'utilisation des fichiers de données, vous devez consulter le guide intitulé *Release Guide 2.5.0 Waves 1 & 2* concernant les vagues 1 et 2 de l'enquête. Ce guide est disponible, en anglais, à l'adresse suivante : [www.share-project.org](http://www.share-project.org) (onglet *Documentation*). Nous en reprenons ici certains éléments et précisons l'originalité et l'objectif de ce dictionnaire des codes.

## Objectif

Le but principal est de présenter, pour l'ensemble des modules de l'enquête, l'intégralité des questions posées lors de la vague 2 (2006-2007) ainsi que la distribution des réponses qui y ont été apportées (version 2.5.0 des données).

Ce dictionnaire des codes regroupe des informations statistiques (fréquence, moyenne, quartiles...) concernant la vague 2 de SHARE et offre ainsi aux utilisateurs une vision d'ensemble des données de l'enquête. Il permet notamment de vérifier l'effectif d'un sous-échantillon avant d'effectuer une analyse plus détaillée.

Les informations statistiques présentées ici concernent l'ensemble des pays européens ayant participé à la vague 2 de SHARE. Il est important de noter que ces informations ne sont **pas représentatives** de l'ensemble de la population des 50 ans et plus de ces pays. Il s'agit de **données brutes non pondérées**.

Les données complètes sont téléchargeables gratuitement, après demande, sur le site [www.share-project.org](http://www.share-project.org).

## Éléments de contexte vague 2

### Échantillon vague 2

SHARE interroge un échantillon d'individus<sup>1</sup> de 50 ans et plus et leur conjoint éventuel (quel que soit son âge). Lors de la vague 2, près de 35 000 individus ont été

---

<sup>1</sup> Certains pays tirent directement des individus, d'autres tels la France, tirent des logements, donc des ménages dont au moins un des membres a 50 ans ou plus (Boersch-Supan et al (2005). The Survey of Health, Aging, and Retirement in Europe – Methodology, MEA - Université de Mannheim (Allemagne), [www.share-project.org](http://www.share-project.org)).

interrogés dans 14 pays européens (Allemagne, Autriche, Belgique, Danemark, Espagne, France, Grèce, Irlande, Italie, Pays-Bas, Pologne, République Tchèque, Suède, Suisse). Plus de 18 000 individus avaient déjà participé à la vague 1.

L'enquête a suivi en vague 2 les individus de la vague 1 qui sont entrés en institution.

### Thèmes traités en vague 2

Les données collectées explorent les domaines suivants :

- l'état de santé (santé physique, comportements à risque, capacités cognitives, santé mentale, système de soins, force de préhension, souffle, vitesse de marche...);
- la participation au marché du travail (emploi et retraite) ;
- les conditions de vie (richesse, revenu, consommation, actifs financiers, transferts, logement) ;
- la démographie (composition du ménage, enfants) ;
- l'environnement social (soutien familial et social, activité sociale, attentes).

### **Contenu des fichiers de données**

Lors de la collecte, les questions étaient organisées par modules qui respectaient l'ordre suivant :

CV	<i>Coverscreen</i>	<i>Coverscreen</i> (composition du ménage)
DN	<i>Demographics</i>	Démographie
PH	<i>Physical Health</i>	Santé physique
BR	<i>Behavioural Risks</i>	Comportement à risques
CF	<i>Cognitive Function</i>	Fonction cognitive
MH	<i>Mental Health</i>	Santé mentale
HC	<i>Health Care</i>	Soins médicaux
EP	<i>Employment and Pensions</i>	Emploi et retraite
GS	<i>Grip Strength</i>	Force de préhension
PF	<i>Peak Flow</i>	Capacité respiratoire
CS	<i>Chair Stand</i>	Lever de chaise
WS	<i>Walking Speed</i>	Vitesse de marche
CH	<i>Children</i>	Enfants
SP	<i>Social Support</i>	Soutien social
FT	<i>Financial Transfers</i>	Transferts financiers
HO	<i>Housing</i>	Logement
HH	<i>Household Income</i>	Revenus du ménage
CO	<i>Consumption</i>	Consommation
AS	<i>Assets</i>	Patrimoine
AC	<i>Activities</i>	Activités
EX	<i>Expectations</i>	Attentes
IV	<i>Interviewer Observations</i>	Observations enquêteur

XT	<i>End-of-Life Interview</i>	Entretien « Fin de vie »
Drop-off	<i>Drop-off</i>	Questionnaire auto-administré
Vignettes	<i>Vignettes</i>	Vignettes d'ancrage

A chacun de ces modules correspond un fichier<sup>2</sup> (au format STATA ou SPSS) contenant :

- d'une part, des variables communes à tous les modules et des identifiants (mergeid, hhid, hhid1, hhid2, country, waveid, split, language, cvid)
- d'autre part, des variables générées à partir de chacune des questions.

Les variables spécifiques à chacun des modules sont identifiées par un code contenant le numéro de la question. Ce code est constitué d'une abréviation sur deux caractères du nom du module en anglais (1<sup>ère</sup> colonne du tableau précédent) et d'un nombre à 3 chiffres. Par exemple, la variable « FT002\_ » correspond à la question « 002 » du module « FT – Transferts financiers ».

Au total, près de 2 000 variables sont disponibles dans l'ensemble des fichiers de données.

Les variables ont été classées en trois grands types (voir exemples pages suivantes) :

- (1) Variable qualitative : les valeurs prises par les variables ne sont pas quantifiables ;
- (2) Variable quantitative : les valeurs prises par les variables sont quantifiables ;
- (3) Variable indicatrice : pour les questions à choix multiples (QCM), plusieurs variables ont été créées qui correspondent chacune à une modalité de réponse ; chaque variable prend la valeur 1 si la modalité de réponse a été sélectionnée et 0 sinon.

Pour chacun de ces types, un traitement statistique différent a été effectué.

## Comment se présente l'information ?

Le document est organisé par modules et suit l'ordre chronologique d'un entretien en vague 2 (voir liste des modules ci-dessus). Le premier chapitre présente les variables disponibles dans chacun des fichiers de données détaillées puis, correspondant à chaque module, les chapitres suivants contiennent l'intégralité des questions posées.

Pour chaque question, l'information se présente en deux temps.

---

<sup>2</sup> La partie intitulée *Coverscreen* a donné lieu à la création de 2 fichiers : CV-H au niveau ménage et CV-R au niveau individuel.

Tout d'abord, un tableau permet de visualiser la question posée :

<b>Variable</b>	<b>Numéro de variable ou Liste des variables pour les QCM</b>
Label	<i>Label de la variable dans STATA (en anglais)</i>
Question	<i>Libellé de la question telle que posée au répondant</i>
Instruction	<i>Instruction éventuellement donnée à l'enquêteur</i>
Filtre	Filtre appliqué à la question lors de l'entretien face-à-face. Dans certains cas, ceci peut expliquer le grand nombre de 'Non-renseigné'.

Puis suivant le type de variable, un tableau présente les réponses fournies par les répondants des différents pays européens ainsi que des statistiques.

### (1) Variable qualitative

Ces variables correspondent généralement aux questions pour lesquelles, le répondant avait plusieurs réponses possibles et ne pouvait en choisir qu'une.

Par exemple, pour la question « ph003 - Diriez-vous que votre santé est... »

<b>Réponses</b>	<b>Fréquence</b>	<b>Pourcentage<sup>(a)</sup></b>	<b>Pourcentage cumulé</b>
1. Excellente	3 170	9,2	9,2
2. Très bonne	6 516	19,0	28,3
3. Bonne	12 642	36,9	65,1
4. Acceptable	8 365	24,4	89,5
5. Médiocre	3 590	10,5	100,0
Ne sait pas	10	.	.
Refus	14	.	.
Non renseigné <sup>(b)</sup>	108	.	.
	<b>34 415<sup>(c)</sup></b>	<b>100,0</b>	

(a) Le pourcentage est calculé par rapport aux observations renseignées.

(b) Le champ « Non renseigné » contient les valeurs manquantes ainsi que les individus filtrés.

(c) La dernière ligne correspond au nombre total d'observations.

### (2) Variable quantitative

Pour certaines questions, les réponses attendues étaient des quantités. Par exemple, cela correspond aux questions relatives aux revenus, à l'âge d'apparition d'un phénomène, à la taille du répondant, à son poids, ...

« ph009\_10 - Quel âge aviez-vous lorsqu'un médecin vous a informé pour la première fois que vous aviez un cancer ? »

Statistique	Valeur
Moyenne <sup>(a)</sup>	58
Ecart-type <sup>(a)</sup>	13
Minimum	1
1 <sup>er</sup> quartile	50
Médiane	58
3 <sup>e</sup> quartile	67
Maximum	85
Nombre d'observations <sup>(b)</sup>	689
Ne sait pas	1
Refus	2
Non renseigné	33 723

(a) La moyenne et l'écart-type sont non pondérés.

(b) Le nombre d'observations correspond au nombre d'observations renseignées.

### (3) Variable indicatrice (QCM)

Pour les questions à choix multiples, le répondant avait la possibilité de sélectionner plusieurs réponses. Pour stocker ces informations, une variable indicatrice (*dummy variable*), correspondant à chacune des possibilités de réponse, a été créée.

Dans ce cas, nous présentons les fréquences relatives à chacune des modalités de réponse ainsi que le pourcentage calculé par rapport au nombre de répondants.

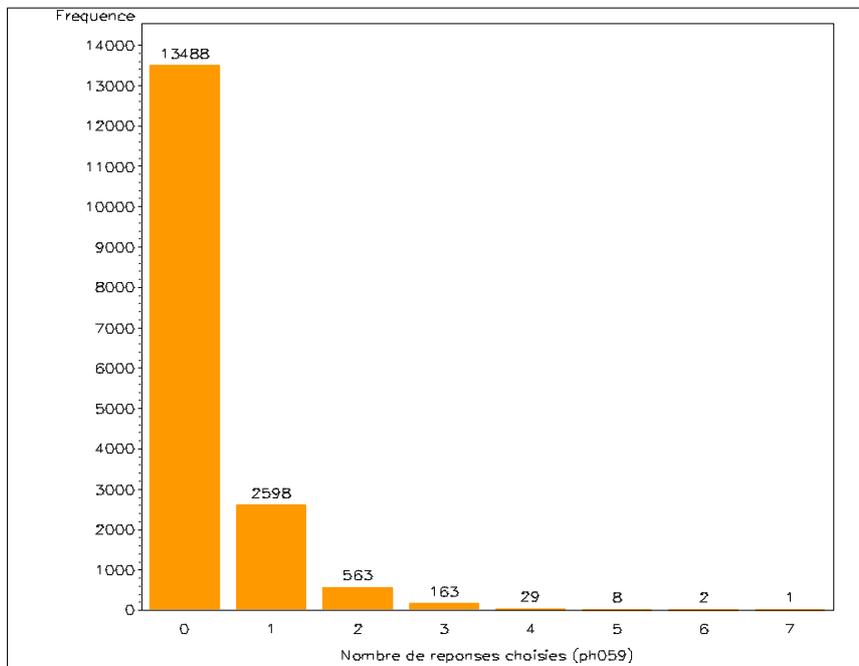
/!\ Dans le cas d'une question à choix multiples, la somme des pourcentages est généralement supérieure à 100 %.

« ph059 - Veuillez s'il vous plaît consulter la carte 13. Utilisez-vous l'un ou l'autre des objets listés sur cette carte? » (Seules les 3 premières modalités de réponses sont présentées ci-dessous)

Réponses	Fréquence	Pourcentage
1. Une canne	2 489	14,8
2. Un déambulateur	821	4,9
3. Une chaise roulante manuelle	378	2,2
...	79	0,5
Nombre d'observations	16 852	.
Ne sait pas	4	.
Refus	5	.
Non renseigné	17 554	.

Cette information est complétée par un graphique qui présente le nombre de modalité de réponses choisies par chacun des répondants.

Question	ph059 (suite)
Histogramme	du nombre de modalités choisies par chacun des répondants



Guide de	lecture
Ex. 1 :	13 488 personnes ont choisi la réponse 96. Aucun de ceux-ci
Ex. 2 :	2 598 personnes ont fourni une seule réponse à cette question

# REMARQUES METHODOLOGIQUES

## Données financières

Toutes les données financières (par exemple dans les modules EP et AS) ont été collectées en monnaie locale. Elles ont ensuite été converties en euro (taux de change disponibles dans les tableaux contenant les données détaillées).

## Répondant

Le répondant peut, dans certains intitulés de questions ou certaines instructions destinées aux enquêteurs, être désigné par « Le répondant » ou la lettre « R ».

Dans le cadre du questionnaire principal, selon la catégorie du répondant dans le foyer, différents modules de questions lui sont posés. Le tableau ci-après récapitule les modules posés selon les populations.

Il existe trois grands types de répondants :

- Le **répondant financier** est défini avant le début du module DN. Il répond au module FT et AS et est identifié par la variable dumfinr<sup>3</sup> ;
- Le **répondant ménage** est défini avant le début des entretiens individuels. Il répond aux questions concernant le ménage dans son ensemble (modules HO, HH, CO) ; il est identifié par la variable dumhhr<sup>8</sup> ;
- Le **répondant famille** répond au module CH et au début du module SP (SP001 à SP017) ; il correspond à la personne ayant répondu en premier dans un couple et est identifié par la variable dumfamr<sup>8</sup>.

Ceci explique que, selon les modules, il y ait plus ou moins de répondants par question.

---

<sup>3</sup> Ces variables sont disponibles dans le fichier de données CV-R.

## Qui répond à quel module ?

N°	Module	Question	Tous	Financier	Ménage	Famille	Non-proxy
0	CV	Coverscreen					
1	DN	Démographie	x				
2	PH	Santé physique	x				
3	BR	Comportement à risques	x				
4	CF	Capacités intellectuelles	x				x
5	MH	Santé mentale	x				x (partiel)
6	HC	Soins médicaux	x				
7	EP	Emploi et retraite	x				
8	GS	Force de préhension	x				x
9	PF	Capacité respiratoire	x				x
10	CS	Lever de chaise (< 75 ans)	x				x
11	WS	Vitesse de marche (> 75 ans)	x				x
12	CH	Enfants	x				
13	SP	Soutien social <sup>(a)</sup>	x (partiel)			x (partiel)	
14	FT	Transferts financiers		x			
15	HO	Logement			x		
16	HH	Revenus du ménage			x		
17	CO	Consommation			x		
18	AS	Patrimoine		x			
19	AC	Activités	x				x
20	EX	Attentes	x				x
21	IV	Observations enquêteur <sup>(b)</sup>					
22	XT	Entretien « Fin de vie »					

(a) Certaines questions concernant les aides reçues sont posées exclusivement au répondant famille.

(b) A remplir par l'enquêteur après chaque entretien.

## Filtres

Comme dans toutes les enquêtes, certaines questions sont filtrées. Suivant les réponses fournies aux questions précédentes, selon l'âge du répondant ou selon qu'il vit ou non en maison de retraite, par exemple, une question sera posée ou ne le sera pas conduisant ainsi à un nombre important de valeurs non-remplées. Le fait de faire partie de l'échantillon longitudinal constitue également un filtre important. Un certain nombre de questions ne sont pas posées aux personnes ayant déjà été interrogées lors de la vague 1 de l'enquête.

**L'information sur les filtres est présentée telle qu'elle est fournie dans le questionnaire générique (anglais) de la vague 2 de SHARE.**

## Variables générées

Pour certains modules, des variables ont été générées afin de pouvoir comparer les résultats (cas des « Catégorie spécifique pays » mentionné dans la section Avertissements), effectuer des redressements de non réponses ou imputer des valeurs manquantes. De nombreuses variables générées sont disponibles (date de l'entretien, revenu total, patrimoine, santé, logement...). Elles seront décrites dans une prochaine version de ce dictionnaire des codes.

Des fichiers de données contenant les variables générées ou les imputations sont disponibles au téléchargement. En voici la liste :

- gv\_health.dta : variables générées relatives à la santé ;
- gv\_housing.dta : variables générées relatives au logement ;
- gv\_isced.dta : variables générées relatives à l'éducation ;
- gv\_weights.dta : pondération tenant compte des répondants uniquement ;
- gv\_weights\_nrp.dta : pondération tenant compte des conjoints n'ayant pas répondu ;
- imputations.dta : variables imputées.

## Code de variable

Le code d'une variable correspond généralement à celui de la question associée.

Ce code est constitué d'une abréviation sur deux caractères du nom du module en anglais (voir tableau liste des modules dans la partie Guide d'utilisation) et d'un nombre à 3 chiffres.

Cependant, trois exceptions existent :

- Les boucles : une même question peut être posée plusieurs fois avec des variations de certains champs. Le code de la variable contient alors le numéro de la question suivi d'un tiret bas (*underscore*) et de l'occurrence de la question.

*Exemple* : ph068\_3 correspond à la 3<sup>e</sup> occurrence de la question ph068\_.

- Les doubles boucles : les questions ep111\_ à ep116\_ sont placées à l'intérieur de deux boucles simultanées. La première boucle correspond au numéro d'une des réponses fournies à la question ep110\_ (allocations ou aides publiques reçues, allant de 1 à 6), et la deuxième correspond à la période de référence du versement reçu (allant de 1 à 20).

*Exemple* : ep112\_2\_3 correspond à l'année de début du versement du 2<sup>e</sup> type d'allocation ou aide publique (voir questionnaire pour détail) pour un 3<sup>e</sup> épisode de versement.

- Les questions à choix multiples : chaque variable correspond à une modalité de réponse à la question et son code correspond au numéro de question suivi de « d » et du numéro de la modalité.

*Exemple* : ph010d3 correspond à la 3<sup>e</sup> modalité de réponse de la question ph010\_.

## Boucle

Certaines questions ont été posées plusieurs fois. Dans ce dictionnaire, les variables supplémentaires sont présentées uniquement lorsque l'information qu'elles apportent est différente et non redondante.

Par exemple, dans le module « PH – Santé physique », les 4 variables associées à la question PH067 ont été présentées car celles-ci correspondaient à l'apparition d'un problème physique différent (crise cardiaque, maladie vasculaire cérébrale, cancer ou fracture de la hanche).

*A contrario*, dans le module « CH – Enfants », seule la 1<sup>ère</sup> variable (sur 14) associée à la question « CH005 - {nom de l'enfant} est un garçon ou une fille ? » a été présentée.

## Cahier des cartes-codes

Lors de la collecte, pour certaines questions, un cahier contenant les modalités de réponses a été mis à disposition du répondant par l'enquêteur.

Ce cahier est disponible en pdf en cliquant sur le lien « Cartes-codes » en bas du sommaire dans la partie gauche de chaque page.