



Enquête SHARE: présentation de la base



Louis ARNAULT, Thomas RENAUD

PSL, Université Paris-Dauphine

SHARE = Survey on Health, Ageing and Retirement in Europe

Une enquête :

- Produite « **par et pour la recherche** » (initiée par le monde académique, projet inscrit dans le cadre de la Statistique Publique)
- **Open Data** (accès aux données via le portail SHARE, ou via la TGIR PROGEDO)
- **Européenne**
- **Longitudinale** (une nouvelle vague tous les 2 ans environ)
- Sur les **50 ans ou plus et leur conjoint éventuel**
- Entretien en **face-à-face**
- Financée sur fonds **publics**
- **Harmonisée** malgré des variations nationales

Vagues de SHARE

1	2	3	4	5	6	7*	8	Cor 1	Cor 2	9	10	
04-05	06-07	08-09	10-11	12-13	14-15	16-17	19-20	Eté 20	Eté 21	21-22	2024	Année
11	15	15	16	15	18	27	28	26	26	27	?	Pays
32 K	38 K	43 K	58 K	65 K	70 K	80 K	50 K	58 K	49 K	73K	?	Entretiens
<i>SHARELife</i>				<i>+ SHARELife</i>				COVID-19		<i>SHARE-Corona v1 et v2</i>		<i>+ HCAP</i>

9 vagues d'enquête principales déjà réalisées depuis 2004

Structure du questionnaire :

- Un **questionnaire principal** individuel (SHARE classique) qui dure **environ 1h30**
- Des recueils **complémentaires ou spécifiques**
 - Questionnaires rétrospectifs sur les parcours de vie (vagues 3 et 7)
 - Questionnaire « fin-de-vie »

Structure des fichiers de données :

- SHARE est une enquête **pluridisciplinaire** et **multi-thématique**
- Environ **25 modules** à chaque vague, **46 fichiers de données pour la V8 uniquement**

!

sharew8_rel8-0-0_ac

sharew8_rel8-0-0_as

sharew8_rel8-0-0_ax

sharew8_rel8-0-0_br

sharew8_rel8-0-0_cf

sharew8_rel8-0-0_ch

sharew8_rel8-0-0_co

sharew8_rel8-0-0_cv_r

sharew8_rel8-0-0_dn

sharew8_rel8-0-0_dropoff

sharew8_rel8-0-0_ep

sharew8_rel8-0-0_ex

sharew8_rel8-0-0_ft

sharew8_rel8-0-0_gs

sharew8_rel8-0-0_gv_accelerometer_day

sharew8_rel8-0-0_gv_accelerometer_hour

sharew8_rel8-0-0_gv_accelerometer_total

sharew8_rel8-0-0_gv_big5

sharew8_rel8-0-0_gv_children

sharew8_rel8-0-0_gv_exrates

sharew8_rel8-0-0_gv_health

sharew8_rel8-0-0_gv_housing

sharew8_rel8-0-0_gv_imputations

sharew8_rel8-0-0_gv_isced

sharew8_rel8-0-0_gv_networks

sharew8_rel8-0-0_gv_weights

sharew8_rel8-0-0_hc

sharew8_rel8-0-0_hh

sharew8_rel8-0-0_ho

sharew8_rel8-0-0_it

sharew8_rel8-0-0_iv

sharew8_rel8-0-0_mh

sharew8_rel8-0-0_ph

sharew8_rel8-0-0_sn

sharew8_rel8-0-0_sp

sharew8_rel8-0-0_sr

sharew8_rel8-0-0_te

sharew8_rel8-0-0_technical_variables

sharew8_rel8-0-0_xt

sharew8ca_rel8-0-0_ca

sharew8ca_rel8-0-0_ca_at

sharew8ca_rel8-0-0_cv_r

sharew8ca_rel8-0-0_gv_imputations_ca

sharew8ca_rel8-0-0_gv_weights_ca

sharew8ca_rel8-0-0_xc

sharew8ca_rel8-0-0_xt

Les données de SHARE sont nombreuses, ...

- Plus de **400 000 réponses collectées** entre la première et huitième vague

... « éclatées » ...

- Au moins une **trentaine de fichiers de données** différents **par vague**
- **Coûteux** de récupérer les infos des différents modules aux différentes vagues

... et complexes

- **Différents niveaux de réponse** : individu, couple et ménage (notions de répondants famille / finance / ménage)
- **Filtres d'une vague à l'autre** pour les informations permanentes
- **Déviations d'une vague à l'autre** : questions nouvelles ou remaniées, questionnaires-papier spécifiques selon le pays...

➤ **Manipuler les fichiers de données brutes n'est pas évident**



Principes d'*easy*SHARE

Idée :

- Offrir un fichier unique, structuré, prêt à l'emploi sans avoir à réaliser les fusions de bases, et donc **utilisable immédiatement**

Caractéristiques essentielles :

- *easy*SHARE inclut... :
 1. ... **tous les répondants éligibles** à l'enquête...
 2. ... à **toutes les vagues** de participation : 1, 2, 3 (SHARELIFE), 4, 5, 6, 7 (SHARELIFE et classique) et 8
- *easy*SHARE n'inclut qu'un **nombre restreint de variables clés** (une centaine malgré tout) parmi l'ensemble des variables disponibles dans SHARE

Ce qui a guidé le choix des variables incluses :

- Comparabilité **internationale**
- Comparabilité **maximale entre vagues** ? mesures longitudinales (exception : histoires de vie rétrospectives collectées en vague 3 dans SHARELIFE)
- Faible taux de **valeurs manquantes**
- Indicateurs **agrégés, standardisés**, plutôt que variables brutes
- **Pas de filtres complexes**, ou alors documentation des filtres des variables cibles

1. Réintégration de l'ensemble des informations connexes

- Transfert des informations stables des vagues antérieures
- Remontée des information fournies par le répondant ménage/financier ou famille, à l'autre individu éligible du ménage

2. Variables recodées pour une utilisation immédiate

- Uniformisation du codage des valeurs manquantes
- Inclusion de variables générées à partir des variables de la base brute (identifiant du conjoint du répondant, âge, scores de santé ou de perte d'autonomie)
- Adaptation de certaines variables pour les rendre comparables avec HRS

3. Documentation très détaillée

- Mise à disposition du programme Stata qui génère *easy*SHARE à partir des données brutes
- Exemples d'analyses sous forme de « fiche cuisine » dans le guide d'utilisation

3 grandes perspectives d'utilisation potentielles d'*easy*SHARE

1. Se **familiariser** avec les données SHARE (perspective pédagogique)

- Sélection de variables qui illustrent le potentiel de SHARE
- Facilite l'entrée des utilisateurs novices dans l'utilisation des données

2. Socle de départ pour construire **sa propre base de travail** sur SHARE

- **Tous** les répondants à **toutes** les vagues dans *easy*SHARE : aucune perte d'observa^o
- Le do-file de construction permet de comprendre les mécaniques **de réaffectation des données d'une vague à une autre**, de **recodage**...
- Facile **d'inclure ensuite de nouvelles variables** à partir des données brutes

3. Parfaitement mobilisable pour mener **des travaux de recherche** sur SHARE



Contenu d' *easy*SHARE

Socio-démo :

- Âge
- Genre
- Pays
- Pays de naissance
- Zone urbaine de résidence
- Nationalité
- Niveau d'éducation (ISCED) et nombre d'années
- Religion à laquelle le répondant se sent le plus "attaché" (vague 1)
- Statut matrimonial
- Âge/genre du partenaire

Composition du ménage :

- Taille du ménage
- Présence d'un partenaire dans le ménage
- Nombre d'enfants vivant dans le ménage

Conditions de vie dans l'enfance (à 10 ans) :

- Nombre de livres dans le domicile
- CSP du principal pourvoyeur de ressources
- Capacités relatives en maths
- Capacités relatives en langue

Relations sociales / entraide :

- Mère/père en vie
- Nombre de frères et sœurs en vie
- Nombre d'enfants en vie
- Nombre de petits-enfants
- Proximité géographique des enfants (enfant cohabitant ?, enfant vivant à proximité ?)
- Activités sociales (bénévolat, sport, formation, politique...)
- Aides reçues et fournies (à/par qui?)

Santé, perte d'autonomie, recours aux soins :

- Santé perçue (auto-déclarée)
- Nombre de maladies chroniques
- Scores de limitations fonctionnelles (ADL, iADL, capacités musculaires, capacités motrices)
- Scores aux tests de force de préhension
- Scores aux tests cognitifs
- Score global de dépression EURO-D + tous les sous-scores (dépression, pessimisme, idées suicidaires, sentiment de culpabilité, irritabilité...)
- Index de qualité de vie et bien-être CASP-12
- Recours aux soins et utilisation des services de santé (visites chez le médecin, hospitalisations, séjours en maison de retraite, aide professionnelle à domicile)

Facteurs de risque :

- IMC
- Consommation de tabac (actuelle et passée)
- Consommation d'alcool
- Activités physiques intenses
- Traits de personnalité (les « Big Five » : Ouverture, Conscienciosité, Extraversion, Agréabilité, Neuroticisme)

Emploi et revenus:

- Situation professionnelle courante
- Caractéristiques de l'emploi : type d'emploi, type de contrat, temps de travail hebdomadaire
- Satisfaction dans l'emploi
- Projet de retraite anticipée
- Capacité "à joindre les deux bouts financièrement" chaque mois
- Déciles de revenu du ménage (par vague et pays)



Travailler avec *easy*SHARE

<https://share-project.centerdata.nl/sharedatadissemination>

Il faut s'identifier (donc s'inscrire)



SHARE Research Data Center

Scientific Data Releases by the SHARE project

Please sign in

Username

Password

Sign in

Les données à votre disposition ?

Base Stata + documentation easySHARE

easySHARE Release
8.0.0

DOI: 10.6103/SHARE.easy.800

2022-02-10 01:00:22

sharewX_rel8-0-0_easySHARE_stata.zip
(20.46MB)

sharewX_rel8-0-0_easySHARE_spss.zip (35.11MB)

sharewX_rel8-0-0_easySHARE_R.zip (19.31MB)

Les données disponibles au format Stata, R ou SPSS (mais le code de construction de la base uniquement au format Stata)



easySHARE_Teacher_Statement
Adobe Acrobat Document
626 Ko



easySHARE_Release_8.0.0_ReleaseG
uide
Adobe Acrobat Document



easySHARE_rel8-0-0
Fichier DTA
73,9 Mo



easySHARE_rel8-0-0
Fichier DO
243 Ko

15
lire
Obs

Format long : 1 ligne par répondant et par vague de participation

Variables		mergeid	wave	country	wavepart
Filter variables here		AT-000327-01	1	11. Austria	12
Name	Label	AT-000327-01	2	11. Austria	12
mergeid	Person identifier (fix across modules a...	DK-395433-01	5	18. Denmark	567
hhid	Household identifier in respective wav...	DK-395433-01	6	18. Denmark	567
coupleid	Couple identifier in respective wave - ...	DK-395433-01	7	18. Denmark	567
wave	Wave	ES-831334-01	1	15. Spain	12345678
wavepart	Wave participation pattern	ES-831334-01	2	15. Spain	12345678
int_version	Interview version (baseline or longitud...	ES-831334-01	3	15. Spain	12345678
int_year	Interview year	ES-831334-01	4	15. Spain	12345678
int_month	Interview month	ES-831334-01	5	15. Spain	12345678
country	Country identifier	ES-831334-01	6	15. Spain	12345678
country_mod	Country identifier (ISO coded)	ES-831334-01	7	15. Spain	12345678
language	Language of questionnaire	ES-831334-01	8	15. Spain	12345678
female	Gender: female=1, male=0	ES-831334-01			
dn002_mod	Month of birth	ES-831334-01			
dn003_mod	Year of birth	ES-831334-01			
dn004_mod	Born in the country of interview				
age	Age at interview (in years)				
birth_country	Country of birth (ISO coded)				

412 110 observations (= lignes)

- 1 observation correspond à **un individu à une vague donnée**
? **plusieurs lignes par individu**

De 30 000 à 77 000 observations selon les vagues (? sauf vague 8)

Wave	Freq.	Percent	Cum.
1	30,419	7.38	7.38
2	37,143	9.01	16.39
3	28,463	6.91	23.30
4	58,000	14.07	37.37
5	66,065	16.03	53.41
6	68,085	16.52	69.93
7	77,202	18.73	88.66
8	46,733	11.34	100.00
Total	412,110	100.00	

Hétérogénéité du nombre d'observations selon les pays

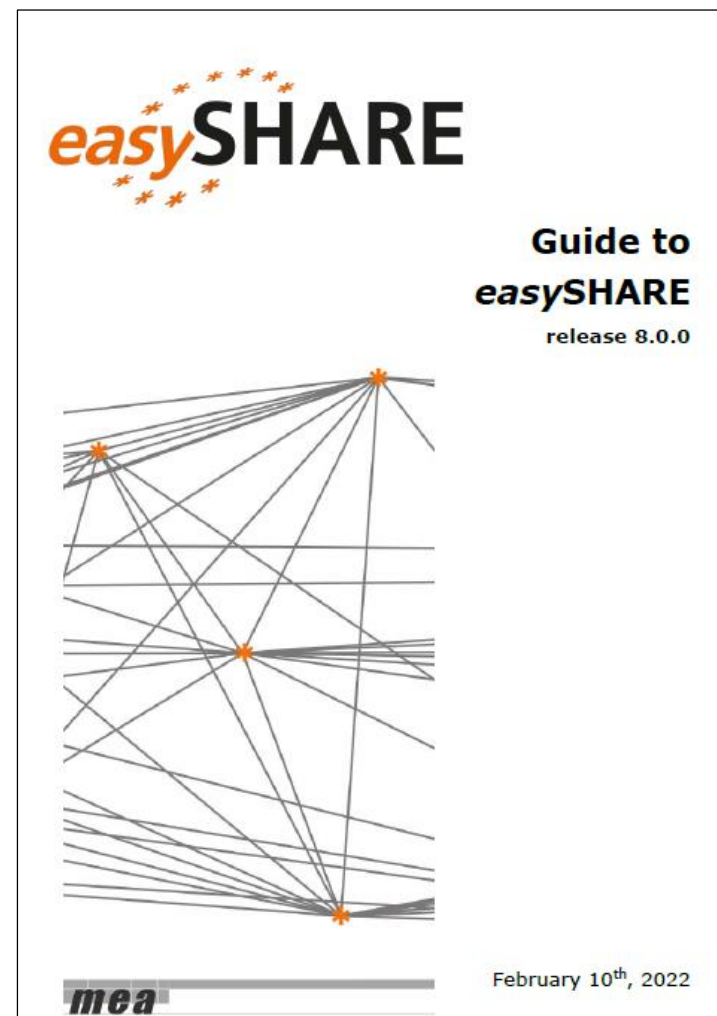
- Davantage d'observations sur les « anciens » pays SHARE (présents depuis la vague 1), comme la France
- Encore peu d'observations dans les « nouveaux » pays inclus dans SHARE uniquement depuis la vague 7

Country identifier	Freq.	Percent	Cum.
11. Austria	21,228	5.15	5.15
12. Germany	26,019	6.31	11.46
13. Sweden	23,789	5.77	17.24
14. Netherlands	16,803	4.08	21.31
15. Spain	29,878	7.25	28.56
16. Italy	28,409	6.89	35.46
17. France	28,727	6.97	42.43
18. Denmark	22,061	5.35	47.78
19. Greece	20,399	4.95	52.73
20. Switzerland	17,763	4.31	57.04
23. Belgium	33,577	8.15	65.19
25. Israel	12,598	3.06	68.25
28. Czech Republic	27,488	6.67	74.92
29. Poland	14,746	3.58	78.49
30. Ireland	1,890	0.46	78.95
31. Luxembourg	5,375	1.30	80.26
32. Hungary	5,391	1.31	81.57
33. Portugal	4,969	1.21	82.77
34. Slovenia	16,121	3.91	86.68
35. Estonia	26,395	6.40	93.09
47. Croatia	6,097	1.48	94.57
48. Lithuania	3,472	0.84	95.41
51. Bulgaria	2,905	0.70	96.12
53. Cyprus	1,771	0.43	96.54
55. Finland	3,171	0.77	97.31
57. Latvia	2,529	0.61	97.93
59. Malta	2,067	0.50	98.43
61. Romania	3,396	0.82	99.25
63. Slovakia	3,076	0.75	100.00
Total	412,110	100.00	

Release Guide EasySHARE

Récapitulatif de toutes les informations nécessaires aux utilisateurs

- **Inclus** avec les données easySHARE
- **Téléchargeable** sur le site de SHARE
<http://www.share-project.org/special-datasets/easyshare.html>





Exemple d'utilisation *easy*SHARE

Mise à disposition du *do-file* servant à construire la base *easySHARE*, qu'il est donc possible de modifier / agrémenter

Ce qui permet notamment de ...

- Bien comprendre les recodages effectués et les bonnes pratiques pour **réaffecter à un répondant** l'ensemble des informations collectées auprès de son conjoint ou à une vague antérieure
- **Réintégrer** à *easySHARE* **certaines variables déjà existantes** dans les bases brutes
- **Construire ou modifier** certaines variables issues des bases brutes puis les intégrer dans la base *easySHARE*, **en fonction de la problématique**



Un code structuré, organisé par module



easySHARE_rel8-0-0



```
47  *****/
48
49
50  *----[ Overview of Contents ]-----
51
52  *----[ 0. Stata Version & Settings]-----
53
54  *----[ 1. Define paths and open log file]-----
55
56  *----[ 2. Define w, m, c_* globals]-----
57
58  *----[ 3. Copy Main SHARE Release to easySHARE directory]-----
59
60  *----[ 4. Extract & Recode Variables from cv_r]-----
61  *----[ 5. Extract & Recode Variables from DN / ST / RP]-----
62  *----[ 6. Extract & Recode Variables from AC ]-----
63  *----[ 7. Extract & Recode Variables from BR ]-----
64  *----[ 8. Extract & Recode Variables from CF]-----
65  *----[ 9. Extract & Recode Variables from CH ]-----
66  *----[10. Extract & Recode Variables from CO ]-----
67  *----[11. Extract & Recode Variables from CS ]-----
68  *----[12. Extract & Recode Variables from EP ]-----
69  *----[13. Extract & Recode Variables from GS ]-----
70  *----[14. Extract & Recode Variables from HC ]-----
71  *----[15. Extract & Recode Variables from HO ]-----
72  *----[16. Extract & Recode Variables from IV ]-----
73  *----[17. Extract & Recode Variables from MC ]-----
74  *----[18. Extract & Recode Variables from PH ]-----
75  *----[19. Extract & Recode Variables from SN ]-----
76  *----[20. Extract & Recode Variables from SP ]-----
77  *----[21. Extract & Recode Variables from HS ]-----
```


Cette partie de code permet de :

- 1) **recoder les valeurs manquantes** associées à certaines variables brutes du module PH (*Physical Health*) en vague 1
- 2) construire **des scores de mobilité** à partir de ces variables brutes

```
*----[18. Extract & Recode Variables from PH ]-----  
  
*>> w1:  
  
use $easy\data\release\sharew1_rel8-0-0_ph.dta, clear  
  
mvdecode ph048* ph049* ph006d? ph006d?? , mv(-1=.a \ -2=.b)  
  
gen      mobilityind = ph048d1 + ph049d2 + ph048d4 + ph048d5  
lab var mobilityind "Mobility index (high: has difficulties)"  
  
gen      lgmuscle = ph048d2 + ph048d3 + ph048d6 + ph048d8  
lab var lgmuscle "Large muscle index (high: has difficulties)"  
  
gen      adlwa = ph049d1 + ph049d3 + ph049d4  
lab var adlwa ///  
        "Activities of daily living w&h index (high: has difficulties)"  
  
gen      adla = ph049d1 + ph049d3 + ph049d4 + ph049d2 + ph049d5  
lab var adla "Activities of daily living index (high: has difficulties)"
```

Cette partie de code permet **de réaffecter** l'information sur le lieu de vie en vague 8, collectée auprès d'un seul répondant (le répondant ménage), **aux autres membres du ménage**

```
*-----  
*-----  
*-----  
  
*----[31. Other recodes per wave ]-----  
  
*>> w8:  
  
use $easy\data\temp\sharew8_merged_a.dta, clear  
  
*>> Assign information on area to other household members  
gsort hhid8 -hou_resp  
gen iv009_mod = iv009  
replace iv009_mod = -9 if iv009_==. & [ho001 == 5 | mn024 ==2]  
replace iv009_mod = iv009[_n-1] if iv009_mod==. & hhid8==hhid8[_n-1]  
lab var iv009_mod "Area of location"  
lab val iv009_mod areallocated  
lab def areallocated -9 "filtered: interview not at home", modify
```

Cette partie de code permet **de réaffecter** les informations sur le nombre d'années d'étude, le pays de naissance et la nationalité, collectées à **la première participation** seulement, à **l'ensemble des vagues de participation** de l'individu

```
*----[34. Transfer information collected once (in baseline interviews)]-----
*           to the next waves of the respondent

foreach var in eduyears birth_country citizenship {
    // isced1997_r is already transfered in gv_isced
    di "`var'"
    // First we check for deviations within person and set the
    // variable to missing if we have contradictory information.
    // This occurs rarely and there is no way to know which
    // information is correct.
    gen      `var'c    = `var'
    replace `var'c    = .   if `var'c < 0
    egen    `var'c_sd =   sd(`var'c), by(mergeid)
    replace `var'c    = .   if `var'c_sd > 0 & `var'c_sd < .

    // Here we start to copy the information to all lines of
    // the respondent; i.e. isced1997_r from wave 1 is written into
    // all lines of the same respondent.
    sort mergeid wave
    foreach i in $w {
        di "`var'"
        replace `var'c = `var'c[_n+'i'] if mergeid==mergeid[_n+'i'] & ///
            `var'c== . & `var'c[_n+'i'] !=.
        replace `var'c = `var'c[_n-'i'] if mergeid==mergeid[_n-'i'] & ///
            `var'c== . & `var'c[_n-'i'] !=.
    }
    // Here we copy back the missing codes, only into the line of the wave
    // the missing code occurred, if we did not find a valid answer in any
    // other wave.
    replace `var'c = `var' if `var'c == . & `var' < 0
    rename      `var' `var'_orig // to preserve the label stored in `var'
    rename `var'c `var'
}
}
```

Contacts de l'équipe SHARE-France pour toute question sur *easySHARE*

- **Thomas RENAUD et Louis ARNAULT** (Production, développement méthodologique et valorisation)

thomas.renaud@dauphine.psl.eu ; louis.arnault@dauphine.psl.eu

Responsabilité scientifique et conduite du projet SHARE-France

- **Florence JUSOT** (Responsable scientifique du projet)

florence.jusot@dauphine.psl.eu

- **Benjamin LEVY** (Conduite de projet)

benjamin.levy@dauphine.psl.eu