



**Disability, Care and Participation
Technical Report: Life Opportunities Survey
ESRC: Secondary Data Analysis Initiative**

ESRC 2674 – February 2017

Eleni Kampanellou

Contents

Live Opportunities Survey.....	1
Two modules.....	1
Identifiers.....	1
Module 0: Sample Selection.....	2
STEP 1: Identifying eligible individuals.....	2
STEP 2: Final sample.....	2
Module 1: Individual Level Demographic and Socio-economic Variables	3
Sex.....	3
Age	3
New age - two versions of age.....	3
Marital status - DVAge > 15	4
New marital status - DVAge > 15	4
Ownership – applies to HRP.....	4
New ownership – applies to HRP (two versions).....	5
Economic status - DVAge > 15	5
New- economic status - DVAge > 15.....	5
Higher education - age >15 AND Age <70.....	6
New higher education - age > 15 AND Age <70.....	6
Ethnicity- applies to all.....	6
New ethnicity - applies to all.....	6
Module 2: New Individual Level Variables	7
Carers (individual).....	7
Module 3: Household Level Variables.....	9
Carers (household/dummy).....	9
Geography: deprivation-urban/rural	11
Module 4: Analysis	12
Factor and cluster analysis.....	12
Acknowledgements.....	14

Live Opportunities Survey

This report outlines the main steps of dataset preparation for the analysis of the Life Opportunities Survey-Wave 1. It includes all the original and new/recoded variables used for the subsequent analysis.

The recoding process took place on the original file of the Wave 1 only dataset named: losw1_special_licencev1_protect.dta [Data/LOS/Coding/original folders].

Two modules

Integrated Household Survey-HIS Core Household module.

Life opportunities module.

HIS Core Household questionnaire: Household information, accommodation type, nationality, country of birth, national identity, religion, period at current address, health smoking, Government training scheme, employment, looking for work and education.

Life opportunities questionnaire: ICF Participation restriction, Learning, Work, Economic Life, Transport, Community-Leisure-Civil Life, Accessibility-The Home, Accessibility-Outside Home, Services and Policies, Domestic Life-Wave1, Domestic Life-Wave2, Caring for others, Discrimination, Crime, Social Contact, Choice and control-Wave2, DDA, Impairment and Health conditions, Income, Benefits, Quality of Life-Wave1.

Identifiers

Person Identifier: Person/ PID

Household Identifier: HSERIAL

Module 0: Sample Selection

STEP 1: Identifying eligible individuals

Number of individuals in dataset (W1): 468000.

Number of children: 9223.

The next step is to use the variable IOut1 that describes whether the individual has given a full/partial/interview or refusal/ineligible etc. The do not apply category of this variable includes 25,351 cases, ineligible 29, refusal 288 and no contact 174. The number of children aged 0-15 are 3146.

Table 1: Interview outcome and group classification

Individual Outcome	Frequencies
N/A	3
Full	31,915
Partial	4,246
No interview <16	9,223
Ineligible	33
Refusal	837
No contact	543
Total	46,800

STEP 2: Final sample

Using the variable IOut1 I delete those in the n/a category, those who refused/ no contact and those who were ineligible. Those in the refused/no contact category do not have weights. For the analysis on adults >15 I keep those individuals who have given full interviews 31914 (one case did not have weights, so it was deleted). For the analysis on children I used the individuals <16, 9223 cases.

Starting sample

Households: 19951 - Population: 46800.

Final sample

Households: 19885(keeping only full interviews with weights).

Population: 31914 (keeping only full interviews with weights).

Module 1: Individual Level Demographic and Socio-economic Variables

Sex

The variable sex kept its original form: 1=Male 2=Female.

Age

DVAge: Variable label = DVAge.

Original variable DVAge as single numbers.

New age - two versions of age

nDVAge: Variable label = Age wave 1 only recoded

Value = 1	Label = 0 to 10
Value = 2	Label = 11 to 15
Value = 3	Label = 16 to 24
Value = 4	Label = 25 to 34
Value = 5	Label = 35 to 44
Value = 6	Label = 45 to 54
Value = 7	Label = 55 to 64
Value = 8	Label = 65 to 74
Value = 9	Label = 75 and over

nDVAge2: Variable label = Age wave 1 only recoded

Value = 1	Label = 0 to 10
Value = 2	Label = 11 to 15
Value = 3	Label = 16 to 24

Value = 4	Label = 25 to 34
Value = 5	Label = 35 to 44
Value = 6	Label = 45 to 54
Value = 7	Label = 55 to 64
Value = 8	Label = 65+

Other versions of age

Variable nWorkage: Values=1 "16-64", 2= "65+" -9= "N/A Children" Label "Working age".

Variable nChildage: Values=1 "Up to 11",2="12-15 only"-9= "N/A Adults" Label "Children".

Marital status - DVAge > 15

MarSta (Wave 1 only): Variable label = Legal Marital Status

Value = 1	Label = Single, that is never married
Value = 2	Label = Married and living with your husband/wife
Value = 3	Label = A civil partner in a legally recognised Civil Partnership
Value = 4	Label = Married and separated from your husband/wife
Value = 5	Label = Divorced
Value = 6	Label = Widowed
Value = 7	Label = In a legally recognised Civil Partnership and separated
Value = 8	Label = Formerly a civil partner the Civil Partnership now legally dissolved
Value = -9	Label = Does not apply (refers to those <16, or no interview ineligible)

New marital status - DVAge > 15

nMarSta (Wave 1 only): Variable label = Marital status wave 1 only recorded

Value = 1	Label = Single, that is never married
Value = 2	Label = Married, civil partnership (2 3)
Value = 3	Label = Separated/Divorced (4 5 7 8)
Value = 4	Label = Widowed

Ownership – applies to HRP

Ten1 (Wave 1 only): Variable label = Tenure

Value = 1	Label = Own outright
Value = 2	Label = Buying it with the help of a mortgage or loan
Value = 3	Label = Pay part rent and part mortgage (shared ownership)
Value = 4	Label = Rent it
Value = 5	Label = Live here rent-free

Value = -9 Label = Does not apply
Value = -8 Label = No answer

New ownership – applies to HRP (two versions)

nTen1 (Wave 1 only): Variable label = Tenure wave 1 only recoded

Value = 1 Label = Owns outright (1)
Value = 2 Label = Owns with mortgage/shared (2 3)
Value = 3 Label = Rent/Rent free (4 5)
Value = -9 Label = Does not apply
Value = -8 Label = No answer

nTen2 (Wave 1 only): Variable label = Tenure wave 1 only recoded

Value = 1 Label = Owns outright/mortgage/shared (1 2 3)
Value = 2 Label = Rent/Rent free (4 5)
Value = -9 Label = Does not apply
Value = -8 Label = No answer

Economic status - DVAge > 15

EconSt Wave 1 (only) Variable label = Wave 1 Economic Status

Value = 1 Label = Full-time Self-employed (self-employed all full-time)
Value = 2 Label = Full time employee
Value = 3 Label = Part-time employee
Value = 4 Label = Workless, retired from paid work
Value = 5 Label = Workless, ILO Unemployed
Value = 6 Label = Workless, sick or disabled
Value = 7 Label = Workless, other inactive
Value = 8 Label = Under 16
Value = -9 Label = Does not apply

New- economic status - DVAge > 15

nEconSt Wave 1 (only) Variable label = Economic Status wave 1 recoded

Value = 1 Label = Employed (1 2 3)
Value = 2 Label = Workless, ILO Unemployed (5)
Value = 3 Label = Workless, retired from paid work (4)
Value = 4 Label = Workless, other inactive/sick or disabled (6 7)
Value = 5 Label = Under 16
Value = -9 Label = Does not apply

Higher education - age >15 AND Age <70

Variable = DVEdAtt Variable label = Highest level of qualification

Value = 1	Label = Degree level qualification (or equivalent)
Value = 2	Label = Higher educational qualification below degree level
Value = 3	Label = A-Levels or Highers
Value = 4	Label = ONC National Level BTEC
Value = 5	Label = O Level or GCSE equivalent (Grade A-C) or O Grade/CSE equivalent
Value = 6	Label = GCSE grade D-G or CSE grade 2-5 or Standard Grade level 4-6
Value = 7	Label = Other qualifications (including foreign qualifications below degree level)
Value = 8	Label = No formal qualifications
Value = -9	Label = Does not apply
Value = -8	Label = No Answer

New higher education - age > 15 AND Age <70

nDVEdAtt : Variable label = Higher Education Wave 1 recoded

Value = 1	Basic (3 5 6) (A levels, O level and GCSE)
Value = 2	Higher (1 Degree level)
Value = 3	Other below degree (4 7 2) (Higher below degree, ONC. other)
Value = 4	No qualifications (8 No formal qualifications)
Value = -9	Label = Does not apply
Value = -8	Label = No Answer

Ethnicity- applies to all

Wave 1 only Ethnicity: EthRep - Variable label = Ethnicity report

Value = 1	White
Value = 2	Mixed
Value = 3	Asian or Asian British
Value = 4	Black or Black British
Value = 5	Chinese or other Ethnic group

New ethnicity - applies to all

Ethnicity was recoded as White vs Non-white due to small numbers.

Variable: nEthRep Label: Ethnicity Wave 1 recoded

Value = 1	White
Value = 2	Non-White (Mixed, Asian/Asian British, Black/Black British, Chinese/other)

Module 2: New Individual Level Variables

Carers (individual)

The next step is to create a carers variable at an individual and household level. I used two variables CareHhd & CareLout that indicate whether the person is a carer in or out of household. These variables are binary 1=Yes 2=No.

Variable: carers (original CareHhd & CareLout) Label: Type of carers

Value = 1	Label = Not a carer
Value = 2	Label = Carer only in household
Value = 3	Label = Carer only out of household
Value = 4	Label = Carer in and out

Coding Missing: -7 I don't know -8 No answer -9 Does not apply

Participation restrictions score

I wanted to create a single participation score that calculates the sum of restrictions based on 8 different binary variables:

DVLearn = Learning
ICFLim02 = Work
DVEconL = Economic Limitations
DVCivic = Civic Life
DVSocial = Social Life
DVTrans = Transport
DVAccOut = Accessibility in the house
DVAccHome = Accessibility out of home

I had to define the -7,-8, -9 values as missing and calculate the total number of occurrences and then redefine the missing values (.) as -7,-8,-9. The variable is called ExBasic and its values vary from 0 for someone with no restrictions to 8. Subsequently I need to merge the categories due to small numbers for those with more than four participation restrictions. The new variable is called nExBasic1 Label=Participation score recorded and has merged the numbers above four in a category called 5 plus.

Impairment and severity score

I wanted to create a single summarized impairment/severity score that calculates the sum of impairments multiplied by the level of severity. There are 14 different types of impairment (learning, mobility, dexterity, pain etc). These are all summed up, so an

individual can have up to 14 different impairments. In order to calculate the sum of impairments I have used the following variables: DVSee DVHEAR DVSPK DVMOB DVDEX DVPAIN DVCOND DVBRTH DVLRN DVINTEL DVBEV DVMEM DVMENT DVOTH. The new variable is called ImpScore and is numeric. I used this variable and multiplied it by the level of severity (1 lowest – 4 highest). The level of severity is given by the variable DVSevImp. The final variable that contains this multiplication (14*4) is called ImpSevScore and has numeric values starting from 0 (no impairment) up to 56 (highest severity/number of impairments). However the highest score for adults in the dataset is 44. By using the ImpSevScore I divide the variable into 4 strata: 0: no impairment, 1-3: mild, 4-11: moderate, 12-44: severe. The new variable is named **ImpSevStrata (Label= Impairment and Severity Score Strata)**. For more information on how these strata is calculated look at the excel file named *stratifying impairment/severity scores*.

Health groups (individual)

I created a variable called: health that groups individuals into categories accruing to their impairment and participation restrictions status. For this I have used all the 8 binary restrictions variables (see above) as well as the binary variable DVImpair (0=No, 1=Yes) that indicates whether someone has impairment or not. The DVImpair variable covers 14 different impairments (see above impairment chapter).

Variable: health Label: Health groups

Value = 1	Label = None and limited
Value = 2	Label = None only (no restrictions)
Value = 3	Label = Impaired and limited
Value = 4	Label = Impaired only (no restrictions)

Coding Missing: -7 I don't know -8 No answer -9 Does not apply.

Using the health variable I create another called FeelingRestricted. It has two categories.

Variable: FeelingRestricted

Value = 1	Label = Restricted (1 & 3 from health)
Value = 2	Label = Not restricted (2 & 4 from health))

Coding Missing: -7 I don't know -8 No answer -9 Does not apply

Module 3: Household Level Variables

Carers (household/dummy)

After creating the individual level carers variable, I created a series of dummy variables called: carersdum1-carersdum7. These take values 1 or 0 and refer to every category of the carers variable plus the missing values (-7,-8,-9). This takes place by HSERIAL (the unique household identifier). I raised the dummy variables at a household level and summarize the frequency of the occurrence of every dummy variable by using the command collapse (sum).

So I have the following equivalent possibilities by household:

- Carersdum1= not applicable
- Carersdum2=no answer
- Carersdum3=don't know
- Carersdum4= no carers
- Carersdum5=in hh carers
- Carersdum6=out of hh carers
- Carersdum7=in and out of hh carers

I created a household level variable based on the above dummy variables.

Variable: hh_carers (original carersdum1-carersdum7) Label: Type of carers household level

- | | |
|-----------|---|
| Value = 1 | Label = No carers in hh |
| Value = 2 | Label = At least one carer in household (not all) |
| Value = 3 | Label = At least one carer out of household (not all) |
| Value = 4 | Label = There are all carers (in &/or out) |

Coding Missing: -7 I don't know -8 No answer -9 Does not apply.

I have also created a second version of the hh_carers for the bivariate analysis.

Variable: hh_carers2 Label: Type of carers household level recoded

- | | |
|-----------|--|
| Value = 1 | Label = No carers in hh |
| Value = 2 | Label = At least one carer in household (not all in &/out) |
| Value = 3 | Label = There are all carers (in &/or out) |

Household size

I created a second version of the DVHSize variable for the bivariate analysis.

Variable: nDVHSize2 Label: Size of household recoded

Value = 1	Label = One
Value = 2	Label = More than one

Household children structure

I created a variable for the bivariate analysis named hh_childage.

Variable: hh_childage Label: Household Children Srtucture

Value = 1	Label = Up to 11 (includes also older children up to 15)
Value = 2	Label = 12-15 only (no younger children in household)
Value = 3	Label = No children<15

Household barriers

Likewise with carers in households, I created a series of dummy variables for seven barriers. For example DVLearndum1- DVLearndum4, ICFLim02dum1- ICFLim02dum5, DVCivicdum1- DVCivicdum4 etc. These take values 1 or 0 plus the missing values (-7,-8,-9). This takes place by HSERIAL (the unique household identifier. I raised the dummy variables at a household level and summarize the frequency of the occurrence of every dummy variable by using the command collapse (sum).

Example: series of dummy variables for learning restrictions.

DVLearndum1	= not applicable
DVLearndum2	= no answer
DVLearndum3	= noLearn (not any learning barriers)
DVLearndum4	= yesLearn (with learning barriers)

I then recode the positive answers (yesLearn) and create the new binary variable.

yesLearn2 – Label: Learning Barriers

Following the same steps I created similar variables for the rest participation restrictions

yesWork2	– Label: Work Barriers
yesCivic	– Label: Civic Barriers
yesAccHome	– Label: Accessibility at home Barriers
yesAccOut	– Label: Accessibility outside Barriers

yesSocial2 – Label: Social Barriers
yesTrans2 – Label: Transport Barriers

Note: Economic restrictions barrier was not necessary to recode since it is already a household level variable.

Household severity

Based on the individual level impairment severity ImpSevStrata I created a household level variable. Likewise with the household level barriers, I created a series of dummy variables:

ImpSevStratadum1 = not applicable
ImpSevStratadum2 = not impaired
ImpSevStratadum3 = mild
ImpSevStratadum4 = moderate
ImpSevStratadum5 = severe

I then recode the positive answers of the above binary variables and create new ones

notapplicableseverityscore2 - Label: Not applicable (severity)

notimpaired2 - Label: Not impaired
mild2 - Label: Mild
moderate2 - Label: Moderate
severe2 - Label: Severe

Variable: house_severity Label: Household Impairment Severity

Value = 1 Label = Everyone not impaired
Value = 2 Label = At least one mild/moderate (not those severe)
Value = 3 Label = At least one severe (including all severe)

Geography: deprivation-urban/rural

Wave 1 only Urban/Rural: DVURIND - Variable label = Urban Rural Indicator.

1) Urban 2) Rural 3) Pseudo (delete the pseudo).

I have used the Indexes of Multiple Deprivation for England, Wales and Scotland (IMDE_D, IMDW_D, IMDS_D). Then I created there variables that divide the index numbers from least to most deprived for every country.

For example for England the following variable is created

Variables: nIMDE_D, nIMDW_D nIMDS_D

Value = 1	Label = Most deprived (1 2 3 4 5)
Value = 2	Label = Least deprived (6 7 8 9 10)

I then combine all the above three new variables and I create a single one

Variable: nIMD_count Label: Index of Multiple Deprivation countries

Value = 1	Label = Most deprived England
Value = 2	Label = Most deprived Wales
Value = 3	Label = Most deprived Scotland
Value = 4	Label = Least deprived England
Value = 5	Label = Least deprived Wales
Value = 6	Label = Least deprived Scotland

Finally I merge the information of both the individual level and household level variables into one dataset.

Module 4: Analysis

Factor and cluster analysis

For the factor and analysis I keep only the adults who have weights, there are 31914 individuals. Then I define the missing values of the variables of interest. I replace the not applicable, no answer and I don't know categories with dots. The analysis has included all the variables that define the different participation restrictions. We have tried both principal component factor analysis and principal factor analysis using stata 14. After testing for the adequacy of the sample and level of correlations using factortest we have applied both commands factor, pcf (principal component) and factor, pf (principal factor). We have found that the principal-component factor analysis is more appropriate and gives more clear patterns of factors. We have used varimax rotation for the factors. The cluster analysis was conducted in spss 22 using the software's interface. A hierarchical simple matching method for binary variable was chosen as the most appropriate one using the same sample as the factor analysis. The code is saved in the file Wave1_analyses.do

Bivariate analysis (adults-children and young adults- households and regions)

As with the cluster analysis the bivariate analysis of adults includes the 31914 individuals aged 16 and over who have weights. For the young adults and children analysis, the sample comprises of individuals aged 11-15 and 16-18 only (10,284 total). For the household and regional bivariate analysis only the households of the 31914 individuals are kept in the sample (19885 households), thus the bivariate analysis is taking place at a household level. The bivariate analysis of different socio-economic and demographic indicators is taking place using the tab command. For more information look at the code files named Wave1_bivariate.do file and Wave1_children_youngadults.do file.

Logistic regression and multilevel modelling

Logistic modelling includes only the adults 31914 cases. Like with the cluster/factor analysis I define the missing values. The analysis focused on five different barriers to participation: work, civic, social, transport and accessibility outside home. These were the dependent variables. Confounding factors included information on age, gender, marital status, educational qualifications, ethnicity, economic status, tenure and economic restrictions. For the work barrier the model included only the working age population. Economic restriction is one of the barriers that functions as a proxy of economic affluence and in this case is one of the independent variables in the model. Educational qualifications are only reported for individuals up to 69 years old and therefore this variable is only included in the work barriers model. Two versions of logistic models are conducted for every barrier. The first version uses the DVImpair variable as an indicator of existence/absence of impairment. The second version includes only the impaired population and uses only the severity of impairment as an independent factor (ImpSevStrata).

Likewise the multilevel modelling uses the same indicators with the addition of household level and geographical level variables. Since multilevel modelling requires that the different levels of the variables are defined in the model, therefore I have used the HSERIAL (household id) variable for household level variables. The household level variables include indicators such as household impairment level, children in households and household size. The HSERIAL variable contains a unique id number for every household. For the geographical level variables I had to create my own unique identifier. In multilevel models the most detailed geographical level that can be used as a unique identifier is a lower output area variable. This is important since lower output area identifiers group areas with similar characteristics like deprivation levels together. However in this dataset the most detailed information is at a local authority level and this can be problematic since this level contains heterogeneous areas. Since deprivation, together with a rurality indicator, is one of the explanatory geographical variables and since it is calculated at a lower output area, then it can be used together with the local authority identifier for the creation of a unique proxy identifier, which is higher than deprivation and lower than local authority. This identifier groups together households with the same deprivation levels in the same local authorities.

This is achieved by using the egen command and group function in stata. A new identifier is named geoid.

Because the rurality (DVURIND) variable has a category named pseudo, in addition to urban/rural, it is necessary to delete this category and therefore have 31912 individuals instead of 31914. I also merge the different deprivation scores for England, Scotland and Wales into one single variable named IMD_Total. Because deprivation levels vary from 1 to 10, I recode IMD_Total into a five category variable QIMD_Total. Similarly to the logistic regression I define the missing values of the variables of interest (-7,-8,-9). Since the analysis explores whether people living in the same household in similar areas might also experience similar restrictions, I apply a multilevel random intercept model. I use the melogit command in Stata14 and define a three level random intercept model.

Acknowledgements

Data for analysis of the LOS were from:

Office for National Statistics. Social Survey Division. (2016). *Life Opportunities Survey, 2009-2014: Secure Access*. [data collection]. UK Data Service. SN: 8076, <http://doi.org/10.5255/UKDA-SN-8076-1>

We acknowledge with thanks the originators of the data sets and those who made them available for secondary analysis via the UK Data Service: www.ukdataservice.ac.uk