**GUIDANCE FOR USE OF QUANTITATIVE 10/66 INDEP DATA**

**Version 1.0**

**Prepared 30th October 2015**

**Martin Prince (****1066drg@iop.kcl.ac.uk****)**

**King’s College London**

**Overall design**

These notes should be read in conjunction with the INDEP study protocol paper (1), the standard operating procedures manual, and the INDEP study questionnaires. Additional information about the project, and linked resources can be found at project homepage <https://www.alz.co.uk/1066/indep.php>, and resource page <https://www.alz.co.uk/1066/resources.php>

The 10/66 Dementia Research Group INDEP study (The Economic and Social Effects of Care Dependence in Later Life) was funded by the ESRC/ DFID joint poverty alleviation programme. We planned to investigate the impact of care dependence upon social and economic functioning at the household level (1).

In a nested cohort study design, households were pre-selected as engaged in incident care, chronic care, or no care (control households) of older adults, on the basis of findings from our previous baseline and incidence wave population-based surveys in rural and urban sites in Mexico, Peru and China (2;3). All care households and an equivalent number of randomly selected control households (batch matched for the age of the oldest qualifying resident) were invited for the INDEP follow-up. We calculated design (sampling) weights based on the inverse of the probability of selection (by site, age group of the oldest resident, and exposure), and response weights based on the inverse of the probability of household response (by site and exposure). The overall weight (pweight) is the product of these design and response weights. This weight can be used to refer back to the overall composition of the population-based sample for the 10/66 incidence wave surveys in these sites.

The focus for INDEP was upon detailed ascertainment of household level economic and social indicators of impact. We also carried out (in Mexico, Peru, China and Nigeria) detailed qualitative case studies of care households with varying characteristics of interest (see accompanying document of guidelines for use of qualitative data)

**1. Levels of data**

The INDEP quantitative data set includes data assessed at the level of

a) Households

Household level data comprises household data from the INDEP household interview, individual data from the household grids in the household interview aggregated back to household level, and household level information from the 10/66 baseline and incidence wave surveys

b) Individual residents

Individual level data comprises information on individual residents from the household grids in the household interview (see below), and from index older adults who completed the older adult interview or proxy (with accompanying information from the key informant interview, which we attempted to complete in every case)

**2. Unique identifiers**

Households are uniquely identified by COUNTRY, AREA (or CENTREID) and HOUSEID. HOUSEIDs are unique in each CENTREID, or combination of COUNTRY and AREA. If you wish to merge household level data, you should do so by COUNTRY, AREA and HOUSEID, or by CENTREID and HOUSEID.

Individual residents are uniquely identified by COUNTRY, AREA (or CENTREID), HOUSEID and CODE. CODEs are unique in each combination of CENTREID and HOUSEID, or in each combination of COUNTRY, AREA and HOUSEID. To merge resident level data, you should do so by COUNTRY, AREA, HOUSEID and CODE (nesting in that order), or by CENTREID, HOUSEID and CODE (nesting in that order). The relationship between COUNTRY, AREA and CENTREIDs is summarised in Table 1 below

**Table 1**

**Relationship between COUNTRY/ AREA and CENTREID**

|  |  |  |  |
| --- | --- | --- | --- |
|  | COUNTRY | AREA | CENTREID |
| China Urban | 1 | 1 | 10 |
| China Rural | 1 | 2 | 11 |
| Peru Urban | 2 | 1 | 3 |
| Peru Rural | 2 | 2 | 4 |
| Mexico Urban | 3 | 1 | 6 |
| Mexico Rural | 3 | 2 | 7 |

**3. Basic data structure**

Every variable have been labelled, with value labels where appropriate

Items from the survey questionnaire use question numbers as the variable name. Derived variables are placed at the end of the each data set, and have been named appropriately.

a) The household interview data set comprises

* all variables from the original interview
* exposure status
	+ according to the original selection of households
	+ following reclassification of incident and chronic care households to ‘care exit’ status, where all those that needed care had died between the 10/66 incidence survey and INDEP follow-up
	+ with the control households as the reference category
	+ with control households where needs for care were identified at the INDEP follow-up reclassified as a separate category
* background household data from the 10/66 incidence wave survey (allowing control for confounding from the baseline of the nested cohort)
* derived variables for the main INDEP outcomes - household income, consumption, savings, debts, economic strain, health care costs, catastrophic healthcare spending, carers giving up work for care, and children of school age out of school or attending intermittently

b) the household grid data sets comprise data on all individual residents reconstituted from the household interview data set.

In the household interview, these data were recorded in grids, with one row for each participant. However, in the household interview data set they are recorded in one row for each household. Preparation of these grid data sets has been a painstaking business with some cleaning of data to clarify the identity of resident across grids, and to ensure consistency of codes across grids. This cleaning was carried out on the grid data sets, and although initial cleaning for valid values was carried out on the main household data set, it would have been too laborious to transfer all of the cleaning from the grid data sets to the main data set. Therefore, any users wishing to use grid data should use the grid data sets. For analyses to be carried out at the level of individual residents the grid data can be used directly, appending household level data, as necessary, for mixed effects models taking account of household level clustering (residents nested within households). For analyses to be carried out at household level, household aggregate indicators can be derived from the grid data sets using the SPSS ‘aggregate’ commands. Many of these have already been provided within the household data set.

The household grids from which individual resident information has been derived are summarised in Table 2 below, and the number of residents upon whom information has been obtained is given in Table 3.

**Table 2**

**Summary of Household Interview Grids**

|  |  |  |  |
| --- | --- | --- | --- |
| Grid | Included | Focus | Specific measures |
| BA | Households followed through from incidence to INDEP. All residents identified at incidence wave | Changes in HH composition from incidence to INDEP. Outcome/ status of all residents from incidence wave | Characteristics. Present/ not present. Died/ moved away. Reasons for moving away |
| BB | Households followed through from incidence to INDEP. All INDEP residents who had moved into the HH since incidence wave | Changes in HH composition from incidence to INDEP. Residents that have moved in since incidence wave | Characteristics. Reasons for moving in |
| BC | All households. All current residents  | Characteristics of all residents at INDEP survey | Age, marital status, education |
| C | All current residents | Employment status | Employment status and reasons for not working |
| D | All current residents receiving any income | Income | Income, by source attributable to individuals, and individual propensity for income pooling |
| G | All current residents with long-term health problems, disabilities or needs for care | Health status and needs for care | Nature of illness/ disability, and use/ cost of health and social care |

**Table 3**

**Summary of Household Interview Grids – numbers of individual residents**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Households | Individual older person | AA12 | BA | BB | BC | D | G |
| Peru urban | 140 | 164 | 79 | 598(512)1 | 77 | 6122 | 441 | 115 |
| Peru rural | 56 | 61 | 19 | 214(171)1 | 44 | 2302 | 144 | 31 |
| Mexico urban | 189 | 191 | 124 | 700(503)1 | 143 | 6912 | 410 | 195 |
| Mexico rural | 167 | 175 | 83 | 515(391)1 | 130 | 6042 | 325 | 109 |
| China urban | 176 | 221 | 192 | 490(406)1 | 39 | 4552 | 396 | 242 |
| China rural | 144 | 130 | 196 | 375(251)1 | 97 | 5842 | 489 | 101 |
| TOTAL | 872 | 942 | 693 | 2892(2234)1 | 520 | 31762 | 2205 | 793 |

Footnotes

1. Numbers still resident

2. BA (numbers still resident) + BB (numbers of new residents) does not equal BC (numbers of current residents, as ‘new’ households (into which an index older person had moved since the incidence wave) were not included in BA and BB.

All individual residents are identified by a CODE number, which is unique within each COUNTRY/ AREA/ HOUSEID, or each CENTREID/ HOUSEID. The CODE remains constant across all grid data sets. The original CODE numbering plan was that numbers for the BA grid were provided in advance (1-10 etc), and that numbers for the BB grid would be generated by the field sites (101-110 etc). These numbers would then be transferred across to the BC grid so that one or two digit CODEs would imply a resident present at the incidence wave interview, while three digit codes would imply a new resident moving in since the incidence wave. This approach was followed correctly in the China and Peru centres, but a different approach was used in Mexico. For reasons of convenience, and to enable accurate linkage, the BC CODEs for Mexico (201-210 etc) have been manually recoded back to the BA and BB grids.

c) the individual older person data set comprises information from the older person interview (items numbered BB) and the older person proxy interview if carried out (items numbered CC) and the key informant interview regarding the older person (items numbered AA). These interviews were carried out on all surviving older participants in the 10/66 baseline and incidence wave surveys in the incident care, chronic care or control households. This allowed us, among other things, to update exposure status since some older people within control households might have developed needs for care. We also assessed disability, life satisfaction, autonomy and met and unmet needs.

The key informant interview generated one further grid file, comprising answers to the AA12 (key informant) questions

**Table 4**

|  |  |  |  |
| --- | --- | --- | --- |
| Grid | Included | Focus | Specific measures |
| AA12 | Incident care, chronic care and control households. Any older persons currently (i.e. in the INDEP key informant interview) rated as requiring care | Residents previously or currently involved in providing care to one or more of the older adults requiring care | Identity of the resident providing care. Caregiving role. Coincident involvement in childcare. Giving up work or leaving education to provide care.  |

The IDs within the individual older person data set link their Participant Code Number (CODE - their ID for the INDEP study) to the PARTICID that, together with HOUSEID and CENTREID uniquely identified them within the 10/66 baseline and incidence wave surveys. These provide richer individual data on their health and social circumstances over a 7-9 year period prior to the INDEP survey. These data are available on request from the 10/66 open access data archive <https://www.alz.co.uk/1066/1066_public_archive_baseline.php>

**4. Summary of documents and data files uploaded onto the ESRC Data Store**

1. Guidance on the data resource (this document – guidance.doc)

2. The project standard operating procedures manual (SOP.doc)

3. The full survey questionnaires (questionnaire.doc)

4. The nine data files (summarised in Table 5 below)

**Table 5**

**Summary of quantitative data resource**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Level (inclusion) | Questionnaire section | Name of data file |
| Household Interview | Household | Sections A-I | household.sav |
| Individual older person interviews | Individual (older adults who participated in 10/66 incidence wave) | Sections AA, BB and/or CC) | indiv\_OA.sav |
| GRIDIndividual resident data derived from older person interview grid | Individual (carers of older people) | AA12 | AA12grid.sav |
| GRIDS Individual resident data derived from household interview | Individual (residents at incidence wave) | BA | BAgrid.sav |
| Individual (new residents since incidence wave) | BB | BBgrid.sav |
| Individual (current residents) | BC | BCgrid.sav |
| Individual (current residents) | C | Cgrid.sav |
| Individual (current residents with an income) | D | Dgrid.sav |
| Individual (current residents with an long-term illness and/ or needs for care) | G | Ggrid.sav |

**6. Future development**

These data sets are also available on request through the 10/66 open access data archive <https://www.alz.co.uk/1066/1066_public_archive_baseline.php>

As further enhancements to the data are made, these will be updated through the 10/66 data archive. Please request updated data releases as appropriate.

If you derive variables that may be of use to other users, please communicate this to us via 1066drg@iop.kcl.ac.uk

Reference List

 (1) Mayston R, Guerra M, Huang Y, Sosa AL, Uwakwe R, Acosta I et al. Exploring the economic and social effects of care dependence in later life: protocol for the 10/66 research group INDEP study. Springerplus 2014 July 28;3:379. doi: 10.1186/2193-1801-3-379. eCollection;%2014.:379-3.

 (2) Prince M, Ferri CP, Acosta D, Albanese E, Arizaga R, Dewey M et al. The protocols for the 10/66 Dementia Research Group population-based research programme. BMC Public Health 2007 July;7(1):165.

 (3) Sousa RM, Ferri CP, Acosta D, Guerra M, Huang Y, Ks J et al. The contribution of chronic diseases to the prevalence of dependence among older people in Latin America, China and India: a 10/66 Dementia Research Group population-based survey. BMC Geriatr 2010 August 6;10(1):53.