

## Preface

Korea is experiencing a fast process of population ageing. The urgency triggered the Korea Labor Institute to create a panel survey on middle/old-aged population(45 or older). The panel survey, known as the Korean Longitudinal Study of Ageing (KLoSA), is to have a better understanding of Korea's population ageing and to offer an insight for policy-making and academic studies. In 2006, a panel was created with approximately 10,000 middle/old-aged adults and the first baseline survey was conducted. A survey on job history is being conducted currently.

KLoSA is significant both in terms of what is surveyed and how it is surveyed. First of all, KLoSA is designed to provide basic data on population ageing in Korea for policy-making and cross-disciplinary studies. The survey deals with social, economic, physical, and mental aspects of life, thus encompasses an array of academic fields such as sociology, economics, healthcare, gerontology, psychology, family studies and demography. Also it is designed while keeping in mind comparative studies with that of the US and European countries. It is also significant in that the survey is conducted face-to-face while it is computer-assisted. Such an attempt is the first of its kind in Korea. Of course, there have been numerous surveys done utilizing computers in Korea, but in those surveys data was collected either through telephone interviews or through web interviews. Computers have not been incorporated in collecting data from face-to-face interviews prior to KLoSA.

This user's guide presents basic elements of KLoSA, describes sampling and weight, provide details of each category and points out matters that require attention. KLoSA's data can be downloaded through the Korea Labor Institute's Data Service System. For details, please refer to Chapter VI of this user's guide.

Details of each chapter shall be updated in a form of technical reports and various documents related to KLoSA shall be made available at <http://klosa.kli.re.kr>. As Labor Review has published KLoSA's survey questionnaire, preliminary research, and a plan for the first baseline survey, please refer to Labor Review for more information on KLoSA's survey design.

November, 2007

**KLI**

**KLoSA Team**

# Table of Contents

## I. Overview of KLoSA / 3

1. Background and Purpose .....	3
2. Uniqueness in Survey Design .....	3
3. Who is Surveyed .....	4
4. Survey Categories .....	4
5. Survey Interval and Period .....	6
6. Progress .....	6
7. Survey Methodology .....	7
8. Data Access .....	8
9. Funding, Survey Organizer and Survey Agency .....	9

## II. Sample and Weight / 10

1. Sample Process .....	10
1) Stratification of EDs .....	11
2) Sampling Procedure of ED .....	12
2. Fieldwork and Response Rate .....	12
1) Fieldwork Process .....	12
2) Final result .....	14
3) Response Rate .....	15
3. Weight and Parameter Estimation .....	17
1) Necessity of Using Weight .....	17
2) Weight Computing Process .....	18

3) Calculating Estimation Equation Using Weight .....	19
---	----

### **III. Details and Features of KLoSA Data / 21**

1. Basic Response Unit .....	21
2. Definition .....	22
3. Survey Flow and Definition by Category .....	23
1) Demographics .....	23
2) Family and Family Transfer .....	24
3) Health .....	27
4) Employment .....	32
5) Income .....	34
6) Assets and Debts .....	35
7) Expectations and Life Satisfaction .....	36
4. Rules for Data Variable Name .....	38

### **IV. Item Non-Response and Imputation / 41**

1. Imputation of Missing Value .....	41
2. Application of Hotdeck Imputation .....	42
3. Characteristics and Details of the Imputation Module in Each Section .....	44

### **V. Generated Variables / 47**

### **VI. Data Access / 51**

## Overview of KLoSA

### 1 Background and Objectives

KLoSA was originated from the critical mind that policy and institutional studies are required to be properly geared up for population ageing which is taking place in a rapid speed in Korea and that basic data should be accumulated on status and behavior patterns of old-aged adults for such studies. Therefore, the purpose of KLoSA is to create research resources to help predict individual behaviors in the process of population ageing and help develop and implement effective social and economic policy through fact-finding on economic activities of middle/old-aged adults. Especially, panel survey on middle/old-aged adults will prove more valuable than one-off cross-sectional survey in a sense that panel survey allows time-series research on income changes after retirement, effects of social system and policy changes on individual decision-making.

### 2 Uniqueness in Survey Design

The survey was designed mindful of two elements. First, data collected shall be used not only in labor studies but also in various disciplines. KLoSA collects data as broad as to include employment status, income, asset, family relations, health and subjective judgement. This shall provide a set of data that some researchers in labor economics, sociology, social welfare, healthcare, family studies and gerontology may find useful. Second, it was attempted to design a survey that will allow comparative studies with that of advanced countries such as US, UK and other European countries that already have longitudinal studies in place. The comparison shall shed some light on what older citizens in Korea are faced with.

### 3 Who is Surveyed

The survey sample is approximately 10,000 adults who are 45 or older (born in 1961 or earlier) and reside in Korea not inclusive of Jeju Island. The 2006 main survey included only those living in households. Many surveys of elderly people in other countries include those aged 50 and over. But KLoSA extends its population to those aged 45-49 because career changes in these middle ages have become an important social issue since the financial crisis in late 1990s, with many people in their late 40s having been laid off from their career jobs. This extended population also enables examination of the relevance between economic activities in the middle ages with those in older ages, and with later life in general.

### 4 Survey Categories

KLoSA's survey categories are as follows:

<Table I-1> Description of Major Areas

Area	Description
CV. Coverscreen	<ul style="list-style-type: none"> <li>Information on household members (gender, age, marital status, relationship with the respondent, etc.)</li> </ul>
A. Demographics	<ul style="list-style-type: none"> <li>Date of birth, gender, academic background, religion, marital status</li> <li>Not living together/the spouse aged below 45: date of birth, academic background, employment status</li> <li>Religion, social activities and relationship</li> </ul>
Ba. Family (Child & Grandchild)	<ul style="list-style-type: none"> <li>Children: gender, age, academic background, employment status, house ownership, marital status, number of children</li> <li>Not living together with children: access, frequency of contact, financial aid given to/from children</li> <li>Number of grand children, care of grand children, time/period of care</li> </ul>
B. Family (Parents, Siblings & others)	<ul style="list-style-type: none"> <li>Sibling: relationship with the respondent, age, marital status</li> <li>Parents: age, academic background, presence (if dead, at what age?), employment status, house ownership</li> <li>Not living together with parents: access, frequency of contact, financial aid given to/from parents</li> <li>Other family members: financial aid given to/from them</li> <li>Presence of a family member with difficulty in ADL/IADL, time and period of care-giving</li> </ul>
Ca. Health Status	<ul style="list-style-type: none"> <li>Subjective health judgment, chronic disease, treatment, experience of traffic accident/injury from a fall/fracture of a bone</li> <li>Signs of aging (vision, hearing, pain), BMI, health-related habits (smoking, drinking), depression measurement</li> </ul>

Area	Description
Cb. ADL/IADL, & Caring	<ul style="list-style-type: none"> <li>• ADL and IADL</li> <li>• Caregiver for ADL (first/second/third): relationship with the respondent, days and hours of care-giving, payment</li> </ul>
Cc. Health Utilization	<ul style="list-style-type: none"> <li>• Health insurance subscription: health insurance, reimbursement, private health care insurance, payer of premium, premium, premium in arrears, delayed period</li> <li>• Health care facility usage: hospitalization, dental clinic, public health center, oriental medical center, outpatient, home call, prescription drugs, purchase and price of health care equipment</li> </ul>
Cd. Cognitions	<ul style="list-style-type: none"> <li>- MMSE-K application of a measurement set tailored to Korea</li> </ul>
Ce. Physical Performance Assessment	<ul style="list-style-type: none"> <li>- Measurement of grasping power</li> </ul>
D. Employment (Work type) (D001~D010)	<ul style="list-style-type: none"> <li>• Distinguishing work types</li> </ul>
D. Employment (Employed) (D100~D200)	<ul style="list-style-type: none"> <li>• Details of employment (workplace information, type of employment, position, industry, labor contracts, working hour, days of leave, salary determination, etc.)</li> <li>• The four social insurance scheme, benefit programs, labor union, etc.</li> <li>• Perception of and satisfaction with work</li> <li>• Job position desired, side job, part-time job, etc.</li> </ul>
D. Employment (Self-employed) (D300~)	<ul style="list-style-type: none"> <li>• Reason of running a business, workplace information, working hour, working day, perception of working hour, days of leave, perception of and satisfaction with work, job desired, side job, part-time job, etc.</li> </ul>
D. Employment (Unpaid Business) (D400~)	<p style="text-align: center;">Family</p> <ul style="list-style-type: none"> <li>• Reason of doing the work, workplace information, status of other workers, revenue, working hour, working day, days of leave, perception of and satisfaction with work, job position desired, side job, part-time job, etc.</li> </ul>
D. Employment (Job Seekers) (D500~)	<ul style="list-style-type: none"> <li>• Job seeking activity in the last one week/month, willingness to find a job, employability, reason of losing a job, reason of seeking a job, job seeking activity, difficulty in job seeking, retirement plan, work experience, the latest job, etc.</li> </ul>
D. Employment (Retiree) (D600~)	<ul style="list-style-type: none"> <li>• Time and reason of retirement, spouse's economic activity at the time, pastime items, perception of retirement and economic activity, the latest employment, etc.</li> </ul>
D. Employment (the last job) (D700~)	<ul style="list-style-type: none"> <li>• The last job information for job seekers or retiree</li> </ul>
E. Income	<ul style="list-style-type: none"> <li>• Earned income (salary, one's own business, agricultural &amp; fishery, side job)</li> <li>• Annuity (national/workplace/private)</li> <li>• Social benefits (status of grants, type)</li> <li>• Other income, total household income in the last 12 months</li> </ul>
F. Assets	<ul style="list-style-type: none"> <li>• Residence (type of ownership)</li> <li>• Real estate assets excluding residence (real estate ownership, lease,</li> </ul>

Area	Description
	factory/farm) • Financial asset (including money from the mutual financing association) • Other non-financial asset • Inheritance/donation, debts, household asset
G. expectation & satisfactions	Subjective & Life • Subjective judgment about future economic conditions (inheritance/donation, estimated working hour, economic activity), life expectancy, living standards, expectations on the government • Satisfaction with life

※ The first wave 1.0 version provides data collected in the CV category in a user-friendly way.

## 5 Survey Interval and Period

Panel survey is conducted on a specific phenomenon at regular intervals to track the changes over time. KLoSA, which began in 2006, conducts two types of surveys, baseline and special. Baseline survey is conducted every even-numbered year starting from 2006 and uses the same survey categories. The first baseline survey was conducted over a six-month period beginning in August 2006. Special survey takes place in every odd-numbered year under special topics not included in baseline survey. Job history will be the special theme for 2007.

## 6 Progress

In 2005, the Korea Labor Institute started preparatory studies for panel survey on middle/old-aged adults to support policy-making for population ageing and to provide research resource for a wide range of disciplines. The panel survey was named Korean Longitudinal Study of Ageing (KLoSA).

The progress made before the creation of KLoSA is as follows: As more attention was paid on societal changes that population ageing might bring about, the importance of relevant studies was recognized in early 2004 by the Presidential Committee on Ageing and Future Society. In 2005, the Ministry of Labor decided to finance the Korea Labor Institute with contributions of Employment Insurance Fund for a panel survey on population ageing. Preparatory works for panel survey started in earnest beginning early 2005.

As part of preparatory studies for the first KLoSA baseline survey, preliminary surveys were conducted for three times. The first preliminary survey was carried out in October 2005 to test some of survey questionnaire and scales that need to be verified during actual interview process. It was face-to-face survey with pencil and

paper approach on approximately 500 middle/old-aged adults. The second preliminary survey, held in February, 2006, incorporated computers in the interviewing process on a sample of 30 middle/old-aged adults residing in Seoul. The attempt was the first of its kind in Korea and to test the feasibility of computer assisted personal interviewing approach in the face-to-face interview environment. The third preliminary survey took place for two months from April 2006 and was the final and more thorough testing, duplicating every process of the first baseline survey from sampling to contacting households to be surveyed. Enumeration districts of Population and Housing Census serve as a framework for KLoSA sampling.

The fieldwork for the first KLoSA baseline survey took five months starting from August 2006. Baseline survey will be conducted every even-numbered year and special survey every odd-numbered year under a special theme not included in the baseline survey.

The fieldwork for the first baseline survey was completed in December 2006. After the first data cleaning, beta version of KLoSA's first baseline survey results was published on March 19, 2007, particularly for researchers who are keen on studies of population ageing.

The second data cleaning and multiple imputation process soon ensued to release 1.0 version of KLoSA's first baseline survey in November, 2007.

All research materials including data, user's guide, questionnaire and code book can be downloaded at KLoSA's web site (<http://www.klosa.kli.re.kr>) or at the Korea Labor Institute's web site (<http://www.kli.re.kr>). Visit the KLI's web site and click on the banner for KLoSA.

## 7 Survey Instrument

### ◆ Computer Assisted Personal Interviewing(CAPI)

KLoSA has adopted Computer Assisted Personal Interviewing (CAPI) as its interview method, rather than one of the conventional methods for social surveys, such as Paper and Pencil Interviewing (PAPI). Interviewers read out questions to respondents from screens and enter their responses immediately. CAPI is often used for large-scale and repetitive public surveys in many Western countries. KLoSA has a long and complicated questionnaire and conducts a repetitive survey for a long-term over a large-scale sample. CAPI is therefore considered to be most appropriate. The processing and interviewing system for CAPI is Blaise, which was developed by Statistics Netherlands in 1986. Detailed information about Blaise can be found at its homepage (<http://www.blaise.com/onlinehelp>).

## ◆ Strengths of CAPI

- CAPI provides automatic routing, according to the logic of the questionnaire. There is no need for an interviewer to look for the next question manually, because the system finds it automatically following a response by an interviewee. Routing problems within the questionnaire are therefore eliminated.
- CAPI checks for inadmissible or inconsistent responses. For example, if the respondent reported someone as his/her older brother in Family Relations yet at the same time reported the brother's age to be less than his or her own, a pop-up window would appear to query this before the programme proceeded to the next question.
- CAPI enables 'customised' questions. For example, the suggested age of life expectancy in the section on Expectations and Life Satisfaction changes according to the age of the respondent: life expectancy appears as 75 for the respondents aged between 45 and 64, but increases to 80 for those aged 65-69, and to 85 for those aged 70-74.
- CAPI allows automatic calculation and scoring for cognition-related questions. With PAPI, interviewers have to use a calculator to check whether the response is correct. But CAPI allows an instant judgment of right/wrong answers by entry of a number given by a respondent.
- CAPI also allows the randomisation of questions. For example, questions asking about income or assets are open-ended in principle. But in order to reduce response refusal, or 'don't know' responses, a range of amounts are asked in a close-ended question format. The total range of amounts may be five (e.g. A, B, C, D, E), but respondents are offered only one of the three middle ranges, i.e. B, C or D, chosen at random. Then, following a response, one more range can be offered.

## 8

## Data Access

For data, questionnaire, coding guide, user's guide, and related research papers, please visit KLoSA's web site at (<http://www.klosa.kli.re.kr>) or click on the KLoSA banner at the KLI's web site at (<http://www.kli.re.kr>).

User registration is required to download primitive data. Data users who registered as a member will receive conference or data updates and other information by mail or via email.

In order to download data, make sure that you have **pop-up blocker disabled** on your browser. If pop-up blocker is enabled, it will block access to the KLI's Data Service System where relevant data resides.

You can download the entire dataset for 2006 baseline survey in "Download entire dataset" section or download some portions of data in "Download selected variables".

## 9 Funding, Survey Organizer, and Survey Agency

KLoSA is funded by Employment Insurance Fund, a decision made by the Ministry of Labor. The KLI's KLoSA Team organizes research activities through active exchanges with advisory panel at home and abroad. TNS Korea is a field agency.

Survey questionnaire was developed by the KLI's KLoSA Team under the supervision of Professor Jiyeun Chang and Professor Jinkook Lee and about fifty experts in various fields were asked for advice. RAND, the Center of Human Resource Research of Ohio State University and Statistics Netherlands provided technical support for the development of computer assisted personal interviewing. Professor Kay-O Lee and Professor Youngwon Kim is responsible for sampling and weight and Professor Juwon Song for imputation.

## Sampling and Weights

※ For the specific research papers for Sampling and Weights, please visit KLoSA's web site at (<http://www.klosa.kli.re.kr>) or click on the KLoSA banner at the KLI's web site at (<http://www.kli.re.kr>).

### 1 Sampling

The population of KLoSA includes, in principle, all adults aged 45 and over. (Many surveys of elderly people in other countries include those aged 50 and over. But KLoSA extends its population to those aged 45-49 because career changes in these middle ages have become an important social issue since the financial crisis in later 1990s, with many people in their late 40s having been laid off from their career jobs. This extended population also enables examination of the relationship between economic activity in the middle ages with that in older ages, and with later life in general.) However, Jeju island was excluded for interviewing convenience sake. The first wave included only those living in households. Institutionalised persons, who were excluded from the first wave, will be included in the second wave of the main survey.

The sampling frame of KLoSA comprises enumeration districts (EDs), as identified by the National Statistical Office's 2005 Census. Under this frame, Apartment EDs and Ordinary housing (non-apartment) EDs totalled 261,237, excluding island areas and institutions (social welfare facilities). For Wave One, 1,000 sample EDs were selected, with the aim of securing a maximum valid sample size of 10,000 persons, from six households per sample ED. This was decided on the basis of the fact that the average number of household members aged 45 and over was 1.67 in the 2000 Census.

### 1) Stratification of EDs

Before selecting sample EDs, the population was stratified by type of area and by type of housing. The EDs, clustered in 15 metropolitan cities and provinces, were stratified first by type of area (urban/rural), then by type of housing (Apartment/Ordinary housing, i.e. non-apartment) EDs.

An effort was made to produce reliable data even from those cities and provinces with small populations. Of the total 1,000 sample EDs, 15 cities and provinces were allocated first, each with 15 sample EDs. The remaining 775 EDs were then allocated in proportion to the population of the 15 cities and provinces. Within each city or province, EDs were allocated subsequently by type of area, i.e. urban and rural areas again, in proportion to the population of the areas concerned.

The final numbers of sample EDs by area and by type of housing are shown in <Table II-1>. Of the 1,000 sample EDs, 409 EDs were allocated to Apartment EDs and the remaining 591 EDs were allocated to Ordinary EDs. In urban areas, 363 EDs were allocated to Apartment EDs and 440 to Ordinary EDs while in rural areas, 46 EDs and 151 EDs were allocated respectively.

<Table II-1> Allocation of Sample EDs in 15 Cities and Provinces

	Total ED			Urban ED			Rural ED		
	Apartment	Common	Total	Apartment	Common	Total	Apartment	Common	Total
Seoul	62	115	177	62	115	177	0	0	0
Busan	29	44	73	29	42	71	0	2	2
Daegy	24	32	56	22	30	52	2	2	4
Incheon	25	32	57	25	30	55	0	2	2
Gwangju	21	18	39	21	18	39	0	0	0
Daejeon	19	20	39	19	20	39	0	0	0
Ulsan	15	17	32	13	14	27	2	3	5
Gyeonggi	89	99	188	78	78	156	11	21	32
Gangwon	15	24	39	12	12	24	3	12	15
Chungbuk	16	23	39	13	11	24	3	12	15
Chungnam	16	30	46	9	8	17	7	22	29
Jeonbuk	19	26	45	17	14	31	2	12	14
Jeonnam	14	31	45	11	9	20	3	22	25
Gyeongbuk	19	40	59	14	17	31	5	23	28
Gyeonam	26	40	66	18	22	40	8	18	26
Total	409	591	1,000	363	440	803	46	151	197

## 2) Sampling Procedure of ED

<Table II-1> describes the allocation result of ED according to municipal cities and provinces-urban area and rural area-housing style. It shows the number of EDs which reflect the size of population according to municipal city and province-urban area and rural area-housing style. In the apartment section of the urban area in Seoul, the population size is 19,017 and 62 districts are allocated.

As economic living conditions are bound to differ according to apartment districts in Seoul, in order to fairly allocate the 62 EDs, the ED population was drawn up using the administrative code and later applying the systematic sampling method.

To prepare for possible change in the condition of the EDs, 20% of alternative EDs were additionally sampled and included as in <Table 4>. The systematic sampling method was applied in order to sample EDs (including alternative) from the working population. To secure sufficient number of compatible households (with a family member aged 45+), EDs where the number of such households did not surpass 20 were excluded. The alternative EDs were again categorized into main and alternative groups, using the systematic sampling method. For instance, a total of 74 EDs (62 main and 12 alternatives) from the apartment in the eastern part of Seoul was sampled from 19,017 EDs, based on the systematic method. Again 12 EDs from the 74 were sampled as alternatives, using the same method.

## 2 1<sup>st</sup> Baseline Survey Report on Process and Results

### 1) Fieldwork Process

Before going into the details of the survey process, it would like to go through the interview training. Interview training took place on July 2006 for three days including a day for ‘general training’ and two for ‘local training’. The general training dealt with introduction of KLoSA and stressed its significances, categorization of jobs/industry, and measurement of recognition. The local training involved how to work with the panel, use the laptop for the CAPI, the context of the survey, and mock interviews. Two local trainings took place in Seoul, one in Busan, one in Daegu, one in Gwangju, and one in Daejun. Interviewers who completed the three-day training received certificates of achievement.

A total number of 84 interviewers finished the first three-day training. Six among the group resigned from taking part in the survey, leaving 78 interviewers to do so. During the course of the survey, additional training

---

was conducted when deemed necessary to replace the interviewers who dropped out and interviewers were committed to areas where progress was slow. Additional training of a total of 31 interviewers has taken place eight times from August to November. The training was for one day but was more intensive as the group was smaller. A total of 115 interviewers enrolled in the training, and 107 were put into action after eight dropped out. Nine interviewers chose to leave while the interviews were in progress, leaving 98 in the program until the survey was completed.

Before the interviewers visited the households chosen as samples, letters were sent out by mail to the households from June to July 2006. The letter introduced the survey, included official statements (from the heads of Ministry of Labor, Korea Labor Institute, and School of Health at Seoul National University), notified them of the visit, and asked for cooperation. It also included notifications via registered mail from town, county, and dong offices where the sample households reside regarding the district survey.

The process of formulating a panel through visits to households was executed in the following manner. The interviewer visits each of the addresses according to a set order with a list of the households, a map of the sample Enumeration District, and a laptop with questionnaire documents at hand. If a person over the age of 45 is confirmed to be living at the address, the household is confirmed as an adequate sample and those over the age of the 45 are interviewed. If there is no household member over the age of 45, the household is confirmed as non-adequate and the interviewer moves on the next address.

12 to 15 households were visited in each Enumeration District according to region and a panel was composed of six households. If a panel is not established with the 12 to 15 households available in the Enumeration District, an additional list of five to eight households was provided to keep the interviews going. However, not all Enumeration Districts had panels established with six households, and there were panels made up of a minimal of one household to a maximum of 12 households according to the conditions in the Enumeration District. There were 422 Enumeration Districts in which a panel was composed of six households (42.2%) and there were a total of 812 Enumeration Districts in which the panel households amounted to four to eight.

There were 96 Enumeration Districts which had less than three households, which broke down to 41 with ordinary houses (non-apartment buildings) and 55 with apartment buildings. Among the 591 Enumeration Districts of regular houses, only 6.9% were households of less than three members, while 13.4% of the total 409 apartment districts were households of less than three members. This indicates that the survey was more challenging in apartment districts. In these districts, the interviewer could visit the homes only if they had prior approval from the mothers' society or the control office. It was also difficult to approach the interviewees as there were many cases in which the contact would be done at the guard room or through an interphone.

The Enumeration Districts with no eligible households, under dismantlement due to redevelopment projects, or where the visits had been turned down all together have been replaced with preliminary Enumeration Districts. <Table 1> reveals the reasons behind replacing 32 Enumeration Districts. The fact that there were no persons over the age of 45 at the sample households accounted for 21 cases and was the largest cause. The second most frequent cause for seven Enumeration Districts was that there were no sample households left due to redevelopment projects where the houses were all being dismantled. There were two cases in which there were insufficient number of eligible households in the Enumeration District, and two other cases in which the interviewer had no access to the households. There was an Enumeration District which was replaced as all the sample households were found to be ineligible, but the replaced samples were also inadequate which led to another replacement of Enumeration Districts. 969 Enumeration Districts out of 1,000 went forward without any changes from the initial selections.

<Table II-2> The reason for replacing KLoSA Enumeration Districts (Unit: figure, %)

Reasons for replacement	Number of Enumeration Districts	
Dismantlement & redevelopment	7	( 21.9)
All households ineligible	21	( 65.6)
Insufficient number of eligible households	2	( 6.3)
Denied access to EDs and households	2	( 6.3)
Total	32	(100.0)

## 2) Fieldwork Result

From August to December 2006, the panel was established with 10,254 middle/ old-aged population over the age of 45 from 6,171 households in the 999 sample Enumeration Districts (Table 2). One Enumeration District among the 46 in the Choongchungnam-do area was not surveyed bringing the total number to 999. The area with the largest number of panel households and individuals was Gyeonggi-do with 1,170 households and 1,935 individuals. The area with the smallest number was the city of Ulsan with 188 households and 318 individuals. The number of panel households was 6.2 on average per Enumeration District. The city of Incheon had the largest number with 7.0, and the province of Gangwon the lowest with 5.5. In addition, the number of panels was on average 10.3 per Enumeration District. The province of Choongchungnam-do had the largest number with 12.4, and the city of Incheon had the lowest number with 9.8. There were on average 1.7 members over the age of 45 in each sample household. Choongchungnam-do had the largest number with 1.9 on average, and the city of Incheon the lowest with 1.4 on average.

<Table II-3> KLoSA 1<sup>st</sup> Baseline Survey Fieldwork Result

(Unit: ED, Housholds, Persons)

	Sample Enumeration Districts(EDs)	EDs with over one successfully interviewed HHs	Successfully interviewed sample HHs	Successfully interviewed sample Individuals	Successfullyinterviewed sample Individuals/HHs
Total	1000	999	6,171	10,254	1.7
Seoul	177	177	1,076	1,767	1.6
Incheon	57	57	400	556	1.4
Gyeonggi	188	188	1,170	1,935	1.7
Gangwon	39	39	215	391	1.8
Busan	73	73	450	743	1.7
Ulsan	32	32	188	318	1.7
Gyeongnam	66	66	390	676	1.7
Daegu	56	56	337	562	1.7
Gyeongbuk	59	59	361	602	1.7
Gwangju	39	39	233	401	1.7
Jeonbuk	45	45	292	485	1.7
Jeonnam	45	45	293	480	1.6
Daejeon	39	39	243	390	1.6
Chungbuk	39	39	235	392	1.7
Chungnam	46	45	288	556	1.9

### 3) Response Rate

A total of 13,602 eligible individuals (middle/old-aged over 45) resided in the 7,574 eligible households. This is the group of actual interviewees that will be put together in KLoSA panels. The number of eligible individuals per eligible household was 1.8 on average.

To reiterate, of the 13,585 excluding empty homes, 11,796 were confirmed directly to have middle/old-aged persons over 45. The confirmation rate was 86.8%. Among these figures 7,574 were eligible households and had middle/old-aged persons over 45 (eligible household members) with the rate of eligibility being 64.2%. A total number of 13,602 persons became interviewees with 1.8 middle/old-aged persons in each eligible household on average.

Of the 7,574 households with over one eligible candidate (over the age of 45), there were 6,171 successful in-

interviews with at least one person. The success rate of the interviews was 81.5% with the remaining 1,403 households declining to respond.

It will take look over the response rate in the first year's KLoSA Basic Research. The response rate can be calculated in various ways. For example, the most conservative means can be calculated by the response ratio of the entire targeted group. This includes cases in which the interviewer failed to meet with the candidate and therefore could not determine whether an eligible person lives there. This means dividing the respondent households with "eligible households (respondent households + non-respondent households) + non-confirmed households." This presumes that all non-confirmed households are eligible which is very unlikely. A more realistic method requires deducting the rate of eligible households among non-confirmed households. Generally, the rate of confirmed households is referred to. In this research, the rate of eligible households was 64.2% of confirmed households. If this is applied to the 1,789 non-confirmed households, it could be assumed that 1,149 are eligible with middle/old-aged persons over 45. The calculation is as follows and  $e$  refers to the assumed rate of eligibility.

$$HHR' Rate = \frac{Number\ of\ R'\ HHs}{Number\ of\ R'\ HHs + Number\ of\ Non - R'\ HHs + e \times Non - confirmed\ HHs}$$

*(R'=respondent, HH=Household, e= the assumed rate of eligibility)*

The above method of calculation puts the response rate of the 1<sup>st</sup> Baseline Survey at 70.7%. This is on par with UK's ELSA 69.9% when compared to the 1<sup>st</sup> round response rates of other panel surveys of middle/old-aged persons in other nations. It may be lower than the US HRS of 80.2%, but is remarkably higher than the 11 EU member states of 55.4%. Considering the conditions large-scale social surveys face today, a household response rate of 70.7% is a good result overall.

On the other hand, the response rate of individual candidates in interviewed households was 89.2%. This was not calculated by the above method, but by generated using the 10,254 successfully interviewed cases and the 11,499 candidates over 45. Such a method is in line with other panel surveys on middle/old-aged persons in other nations. The individual response rate overseas was the highest in UK's ELSA (97%), followed by Korea's KLoSA 89%, the US HRS 82%, the EU's SHARE 86%. KLoSA surveys all members of the household over 45, while HRS, ELSA, and SHARE only survey a couple even if several generations over 45 are in the household.

<Table II-4> Response rate of 1<sup>st</sup> year panel surveys on middle/ old-aged persons

(Unit: %)

1 <sup>st</sup> year panel surveys	Household response rate	Individual response rate
USA, HRS(1992)	80.2%	81.6%
UK, ELSA(2002)	69.9%	96.5%
Europe 11 countries, SHARE(2004)	55.4%	86.3%
Korea, KLoSA(2006)	70.7%	89.2%

The source: USA(HRS). Steven Heeringa and Judith Connor. *Technical Description of the Health and Retirement Survey Sample Design*. Institute for Social Research, University of Michigan. 1995.

UK(ELSA). Michael Marmot, James Banks, Richard Blundell, Carli Lessof and James Nazroo (eds.). *Health, Wealth and Lifestyles of the older Population in England: The 2002 English Longitudinal Study of Ageing*. Institute for Fiscal Studies. 2003.

EU(SHARE). Börsch-Supan, A., Brügiavini, A., Jürges, H., Mackenbach, J., Siegrist, J. and Weber, G. (eds.). *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe*. Mannheim Research Institute for the Economics of Aging, University of Mannheim. 2005.

### 3 Weight and Parameter Estimation

#### 1) Necessity of Using Weights

In general, sample surveys aim to correctly estimate the parameters of the population, e.g. its sum, mean, ratio, etc. However, the sample design of KLoSA was not made on an equal probability sampling. Instead, it was made on a systematic sampling after samples have been allocated first by region (city or province) and then by type of housing (apartment or ordinary housing, i.e. non-apartment). Under these conditions, parameter estimations based on simple mean or ratio of the sample can be biased. In order to produce unbiased estimators, we use weights, taking into account the selection probabilities and response rates of each sample. The benchmark adjustment based on other external information also needs to be considered.

Even if each sample has the same selection probability, we must consider unit non-response as well as item non-response rates. Additionally, if the information received at the time of the survey is not same as that considered initially when designing sampling, the benchmark weight adjustment must be considered so that the parameters can be properly estimated.

In short, KLoSA uses weights in order to produce unbiased parameter estimators. Its detailed processes are as follows.

## 2) Weight Computing Process

The weights for KLoSA consisted of three components, as below.

$$\text{Weight} = \text{inverse number of Selection Rate} \times \text{inverse number of Response Rate} \times \text{Benchmarking Weight}$$

Each component is explained below.

### ◆ Selection Rate

- KLoSA adopted a two-stage selection process for its sampling, first by Enumeration District (ED) and then by Household. Therefore, its selection rate also consisted of two rates, i.e. the selection rate of ED and the selection rate of household.

$$f = f_{1i} \times f_{2 \cdot 1i}$$

- $f$  denotes the total selection rate.  $f_1$  refers to the selection rate of the ED and  $f_{2 \cdot 1}$  to the selection rate of households within the ED. The samples of KLoSA were selected based on stratification by region (city/province), by area (urban/rural), and by type of housing (apartment/ordinary housing). The calculation formula taking this into account is as below.

$$f = f_{1i} \times f_{2 \cdot 1i}$$

$$f_{1i} = \frac{n_i}{N_i}, \quad f_{2 \cdot 1i} = \frac{m_{j \cdot i}}{M_{j \cdot i}}$$

- Here,  $i$  denotes the stratum.  $N_i$  refers to the number of the total EDs in the  $i$  th stratum and  $n_i$  to the number of sample EDs in the  $i$  th stratum.  $M_{j \cdot i}$  refers to the number of eligible households within the  $j$  th sample ED in the  $i$  th stratum, and  $m_{j \cdot i}$  to the number of the surveyed households within the  $j$  th sample ED in the  $i$  th stratum. The calculation formula for  $M_{j \cdot i}$  is as below.

$$M_{j \cdot i} = P_{j \cdot i} \times H_{j \cdot i}$$

- $P_{j-i}$  denotes the ratio of the eligible households within the  $j$  th ED in the  $i$  th stratum and  $H_{j-i}$  refers to the total number of households within the  $j$  th ED in the  $i$  th stratum.

#### ◆ Response Rate

- The response rate of KLoSA is the ratio of the number of respondents to the number of persons aged 45 or older within the eligible households. That is, if ... denotes the number of persons aged 45 or older within the eligible households and ... refers to the number of respondents within the eligible households, the response rate can be summarised as  $r_{ijk}/R_{ijk}$

#### ◆ Initial Weights

- Weights were initially designed as multiplied values of the inverse number of the selection rate and the inverse number of the response rate.  $i$  denotes the stratum,  $j$  refers to the ED and  $k$  to the household.

$$w_{ijk} = \frac{1}{f_1} \times \frac{1}{f_{2 \cdot 1}} \times \frac{R_{ijk}}{r_{ijk}}$$

#### ◆ Final Weights

- The weights were finalized by adding the benchmarking adjustment coefficient to the initial weights.

$$w_{ijkl}^* = BF \times w_{ijkl}$$

- Here,  $l$  refers to the respondent and  $BF$  to the benchmarking adjustment coefficient, which was calculated as below.

$$BF = \frac{\text{Estimating Settled Population of } i(\text{Region}), j(\text{Sex}) \text{ and } k(\text{Age}) \text{ groups}}{\text{Designed Weights Value } \sum \text{ of } i(\text{Region}), j(\text{Sex}) \text{ and } k(\text{Age}) \text{ groups}}$$

### 3) Calculating Estimation Equation of Using Weight

#### ◆ Sum

$$\widehat{Y} = \sum_{k=1}^H \sum_{i=1}^{n_k} \sum_{j=1}^m w_{hij} \cdot y_{hij} = \sum_{i=1}^H \widehat{Y}_h$$

$$\widehat{V}(\widehat{Y}) = \sum_{h=1}^H \widehat{V}(\widehat{Y}_h)$$

With the above formula, ... refers to the sum of  $h$  th stratum, and its estimated distribution is as below.

$$\widehat{V}(\widehat{Y}_h) = \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (y_{hi.} - \overline{y}_h)^2$$

here,  $y_{hi.} = \sum_{j=1}^m w_{hij} \cdot y_{hij}$  and finally  $\overline{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi.}}{n_h}$

◆ Mean (or Ratio)

The formulas for the national estimation of the mean and for its estimated distribution are presented below. The ratio estimation is one kind of mean estimation, with the only difference being that the response data  $y$  for the former is binomial, i.e. 0 or 1. Therefore, the following formulas can be used in the same manner for the ratio estimation.

$$\widehat{Y} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^m w_{hij} \cdot y_{hij}}{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^m w_{hij}}$$

$$\widehat{V}(\widehat{Y}) = \sum_{h=1}^H \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} (e_{hi.} - \overline{e}_h)^2$$

here,  $e_{hi.} = \left[ \sum_{j=1}^m w_{hij} (y_{hij} - \widehat{Y})^2 \right] / w_{hi.}$  and  $\overline{e}_h = \left( \sum_{i=1}^{n_h} e_{hi.} \right) / n_h$

The standard error can be denoted as below.

$$StdErr(\widehat{Y}) = \sqrt{\widehat{V}(\widehat{Y})}$$

## Details and Features of KLoSA Data

### 1 Basic Response Unit

#### ◆ **Basic response unit: Adults aged 45 or over as of 2006.**

- Respondents are adults aged 45 or over as of 2006. Anyone who is residing in households and was born in 1961 or earlier is qualified to become respondents of KLoSA.
- KLoSA is a survey that focuses on individuals. Conditions of households are collected to understand surroundings of individuals.
- For Income and Asset categories, respondents are asked to answer their own income and assets in their name. Total income and asset of a household will be calculated based on answers given and be provided as generated variables.

#### ◆ **Response unit that require attention**

- Information on children of a couple, both of whom are 45 years old or older and reside in a household, shall be identical. Since both of the couple are eligible for KLoSA survey, information on children given by the first respondent of the couple was used for the remaining member of the couple. The variable "W01Ba\_resp" indicates which member of a couple provided information on children.
- **"One year before"** indicates one year prior to survey and **"last one year (year 2005)"** indicates the period starting from January 1 throughout December 31, 2005.
- Monetary unit is KRW 10,000.

## 2 Definition

- ◆ **Loop:** If survey questions from a certain number to another deals with matters regarding one event or one person, the block of questions shall be repeated for the number of events or persons in question. Such question groups are called a loop. For an instance, 11 question items from Ba01 to Ba11 in the Family category ask basic background of the first child (i.e. date of birth, education, marital status, etc.). In this case, the same 11 question items should be repeated for the second child, the third and so forth for comparable dataset. Such a group of questions repeated for the number of events or children involved is a loop.
- ◆ **Unfolding Bracket Questions:** Unfolding bracket questions appear in a box in the questionnaire. When respondents answer "don't know" or "refuse to answer" to questions on income, assets or consumption, they are directed to unfolding bracket questions where an entry point amount is suggested so that they can indicate whether an actual amount is above or below the amount suggested. There appear five questions in the box but on the computer screen only one question will appear in random to ask a bracket sequence. By doing so, respondents will learn that they cannot skip to the next question when they refuse to answer. It helps reduce frequency of non-response. For those who don't answer because they don't recall, unfolding bracket questions provide a chance to think over once more. In any case, unfolding bracket questions collect partial information, at least, from respondents unwilling to provide information and such partial information proves useful for imputation of non-response.

**Generated variables** are added to the end of each survey category. They are variables that are adjusted from basic variables given in the questionnaire for the convenience of users. (※ Refer to Chapter V. List of Generated Variables or user's guide's "How to Use Generated Variables" at homepage.)

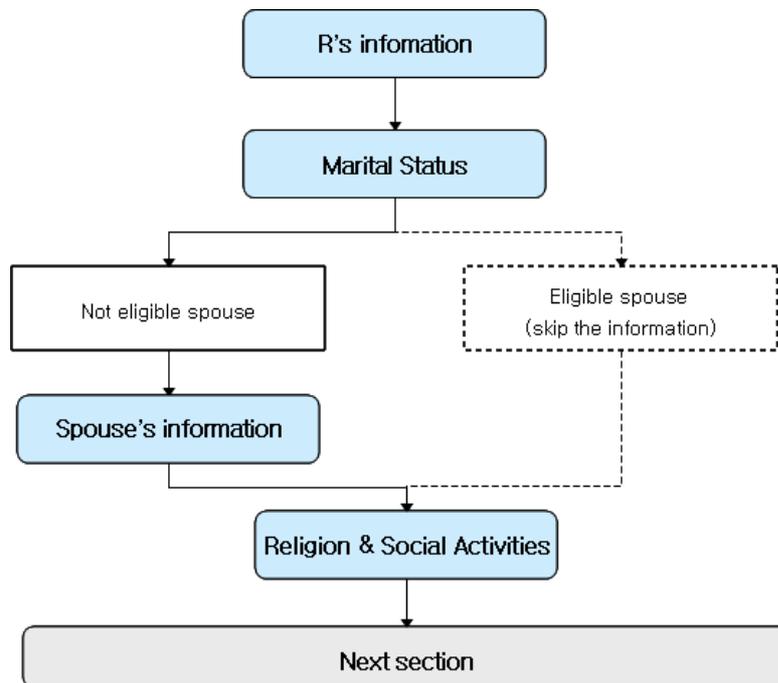
Dataset imputed through multiple imputation method are provided for questions on which respondents answer "don't know" or refuse to answer.

(※ Refer to Chapter IV or for further detail refer to user's guide's "Multiple Imputation Report" at home page)

3 Survey Flow and Definition by Category

1) Demographics

[Figure III-1] Survey Structure of the Demographics Category

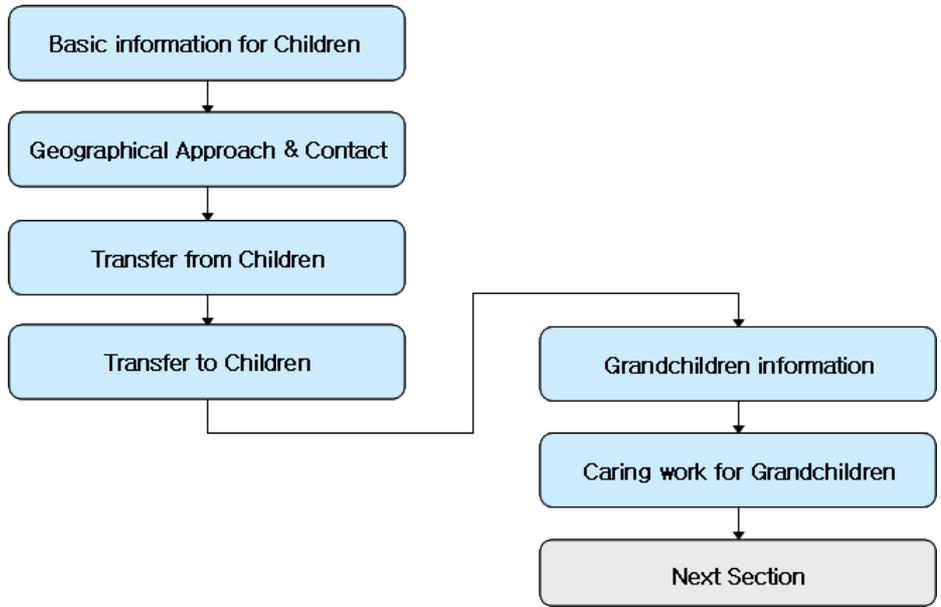


◆ Matters that require attention in terms of spouse's basic information

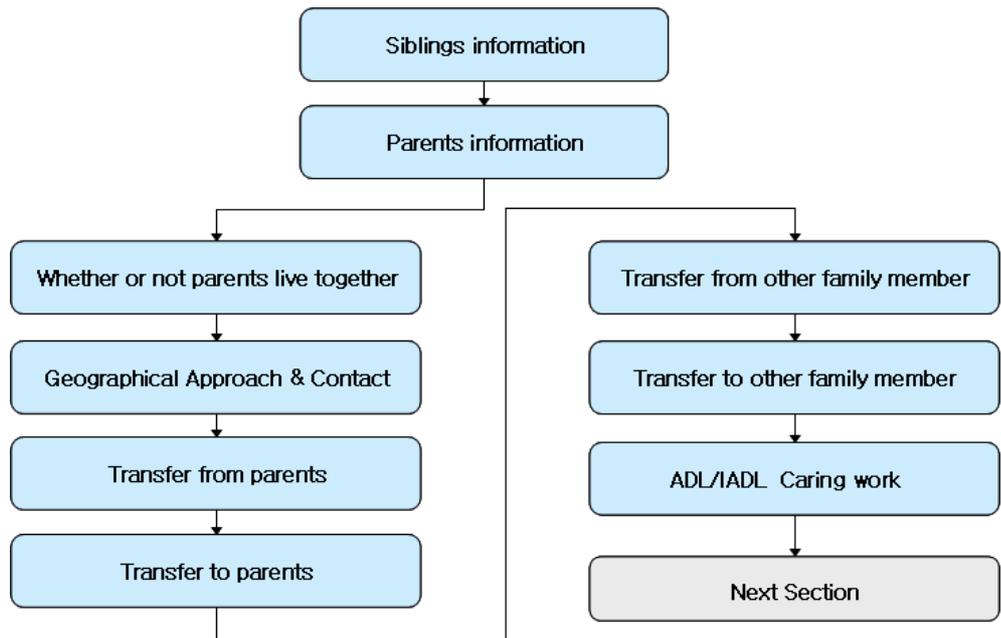
- Basic information is not collected on respondent's spouse aged 45 or over since the spouse is also eligible for the survey.
- Basic information (date of birth, education and employment status) is collected on respondent's spouse from respondent if the spouse has died, is missing or lives separately, or is excluded from survey because the spouse is younger than 45 years old.

## 2) Family

[Figure III-2] Family and Family Transfer(Children & Grandchildren)



[Figure III-3] Family and Family Transfer(Siblings & Parents)



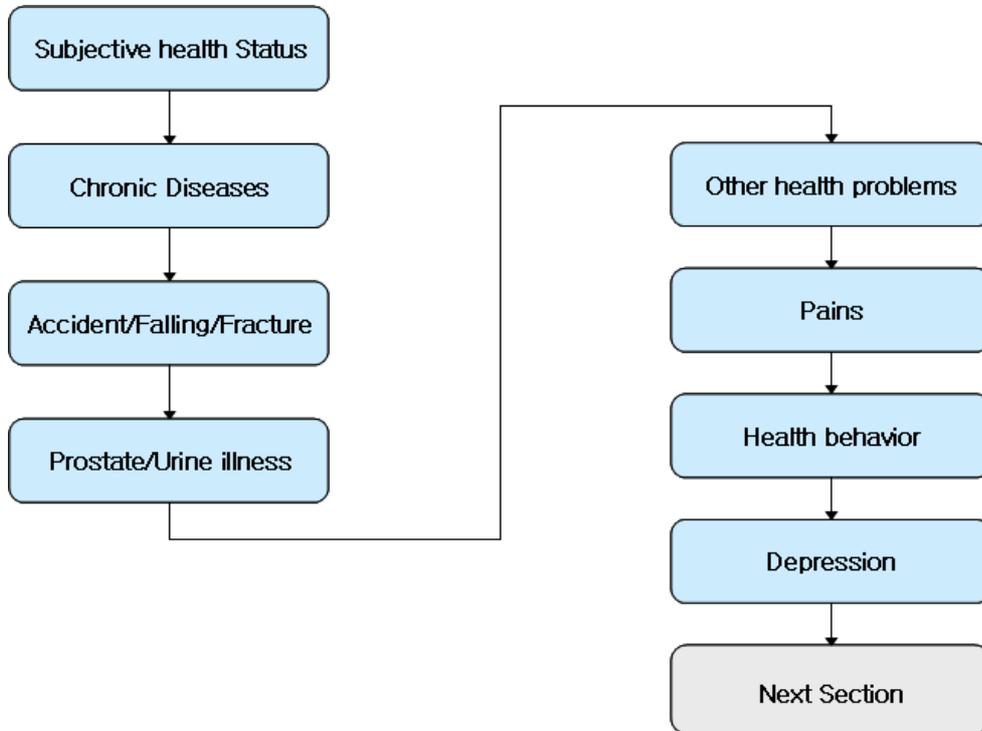
◆ **Definition and matters that require attention in the Family category**

- **Respondent's judgement on family relations is respected:** Respondent's judgement on family relations is fully respected irrespective of biological or blood connection or whether they are listed in the same family register. To take an instance, in relation to marital status, "currently married or living with a partner" shall include common-law marriage and "separated" shall include separation due to discord as well as other reasons including long-term institutionalization and long-term overseas stay. If respondent regards his/her non-biological father as his/her father and answers so, his/her response shall be respected. The same goes to a child (adopted, biological or step-child).
- **Monetary support** (it is further classified into regular and occasional): Monetary support includes cash as well as payment of various bills including medical expense, insurance premium payment, tuition, monthly payment for home purchase, deposit money for rent house and monthly rent.
- **Regular/occasional support:** Regular support refers to support that is provided on a regular basis ( once a month or once every two months) and occasional support refers to support for expense that occurs irregularly or unexpectedly (i.e. medical expense, tuition, living expenses offered irregularly) excluding holidays or birthdays.
- **Non-monetary support** (support in kind): Non-monetary support refers to support in a form of actual articles or gifts rather than cash. This includes services of going grocery shopping on behalf, or making kimchee or other side dishes for meals. Taking care of grandchildren, home care services and other care-taking activities are excluded and addressed separately.
- **Basic information of all living children:** This information is collected by question **Ba02 through Ba11**. Rules for variable code including putting W01 before question number and \_01 after question number are applicable. (\*For details, please refer to Paragraph 4 of Chapter III). For instance, variable code for Question # Ba02 is W01Ba02\_01 for the first child and W01Ba02\_02 for the second child and W01Ba02\_03 for the third.
- **Information on contact and monetary transfer with children who do not live together : Questions Ba12 through Ba46** deal with information on how often contacts or transfers are made between respondent and children who do not live together.
- **Question flow of monetary/non-monetary support to parents:** For questions asking whether respondent provides monetary/non-monetary support to parents who do not live together, a different set of questions are asked depending on the following conditions.

Question #	Condition
B033~B064	Both of parents are alive and live together.
B065~B102	Only father is alive or parents live separately even if both of parents are alive.
B103~B140	Only mother is alive or parents live separately even if both of parents are alive.

3) Health

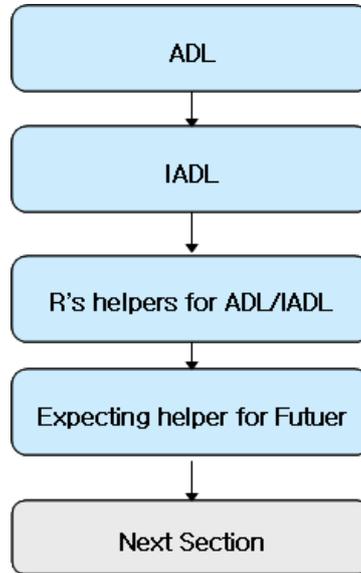
[Figure III-4] Health Status



◆ Subjective health status : C001 and C142 are placed in a random order.

- Two different versions of a question are asked for judgement on respondent's health status. C001 and C142 are five-point scale questions but with different scale structure. The former has Fair in the middle 3 and the latter has Excellent in 1 and Fair in 4. Which of the two questions comes first on the computer screen is decided randomly to eliminate response bias based on how question is placed in terms of order.
- That is, for some respondents, C142 is asked first followed by other health related questions and ended by C001 and for others C001 comes before C142. It is to compare responses on subjective health status depending on the item composition of the scale. Such technique, also adopted in longitudinal studies on ageing in other countries, will allow international comparative study.

[FigureIII-5] Functional Limitations and Helpers



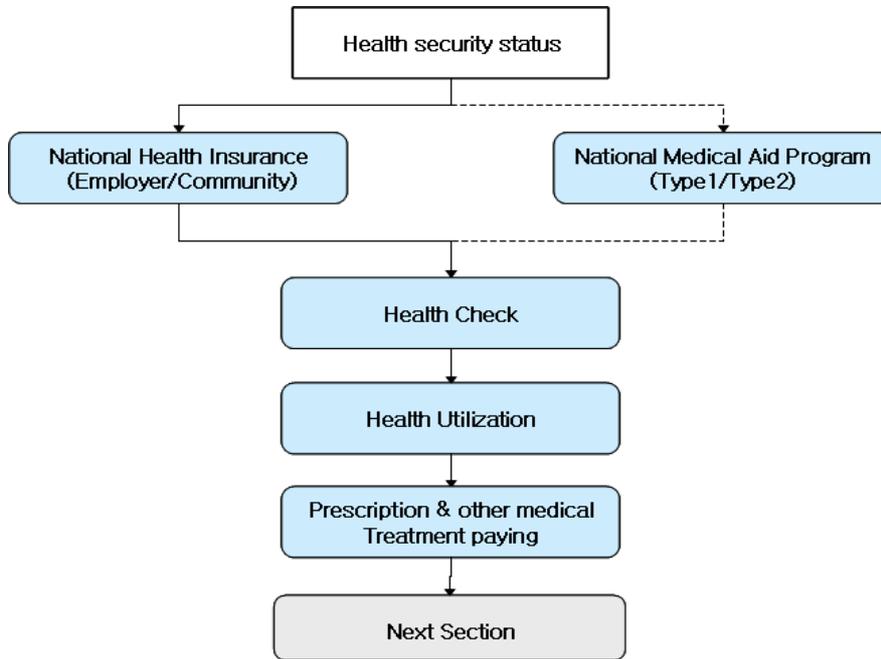
◆ **Response criteria on questions regarding ability of daily activity (ADL) and ability of instrumental daily activity(AIDL)**

- It is based on activities of the last one week. Difficulties in conducting daily activities due to temporary pain or injury that respondent expect to last less than three months are excluded as it is deemed that respondent shall carry daily activities without help after for some time.
- The fact that respondent is not engaged in IADL because there are helpers around or respondent is not willing to do does not qualify a situation, "I need help". Difficulties in conducting IADL because of physical, mental or cognitive problems qualify a situation that requires help.

◆ **ADL/IADL helper**

- B. Care-giver in the Family category is when respondent helps other family members having difficulties in conducting daily activities whereas **ADL/IADL helper in this category is a person who helps respondent in conducting daily activities** due to respondent's physical, mental and cognitive problems. Respondent may name helpers up to three in the order of the most often helping person to the lesser helping person.

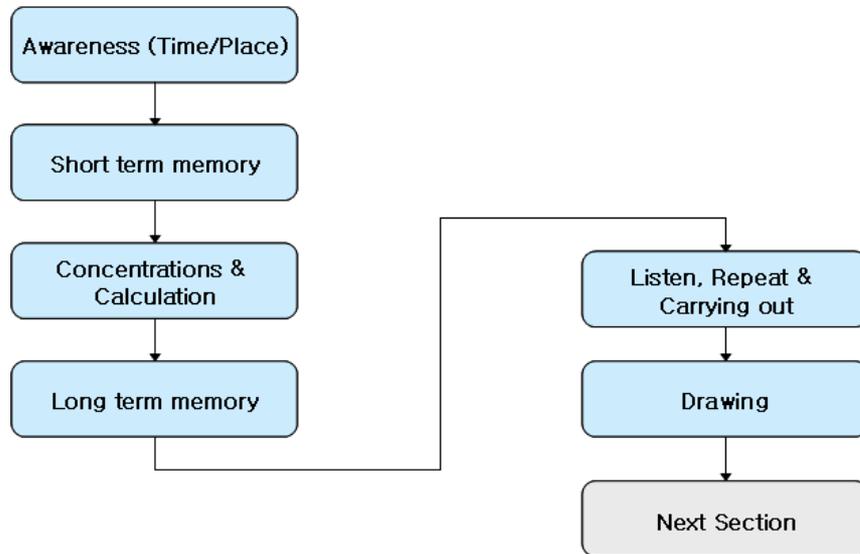
[Figure III-6] Health Insurances and Services



◆ **Matters that require attention in regard to medical expenditure: Amount that respondent paid is measured.**

- Medical expenditure is broken down to hospitalization expense due to injury, operation or diseases which cost the most, dental care expense, oriental clinic expense and other ambulatory care. For expenses for such medical care, it is asked how much respondent pay out-of-pocket excluding the amount covered by insurance plans or other family members.

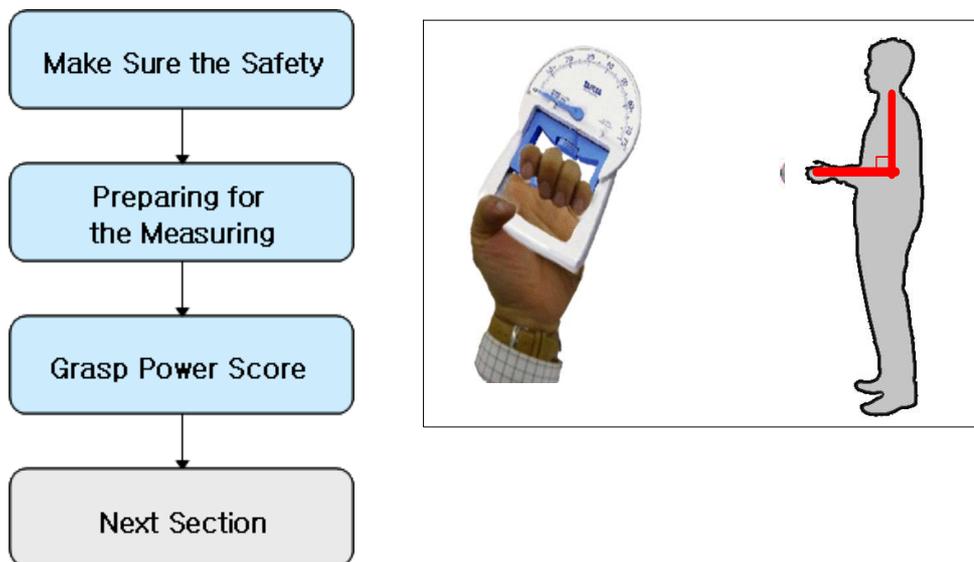
[Figure III-7] Cognition Ability



◆ Questions on cognition

- Cognition is one of the often used methods to check dementia. As dementia has emerged as a very important ageing issue, a periodic check on cognition of adults aged 45 or over will make a valuable resource for researches on dementia among Korean population. This is the reason cognition check is included in KLoSA's baseline survey.
- Training on interviewers assisted by experts is conducted before preliminary surveys and baseline surveys were carried out to promote accurate cognition check.
- In an effort to increase the accuracy of checking respondent's cognition, CAPI survey program is programmed to eliminate interviewers' interference in testing memory and calculation ability and let computers check time and accurate or inaccurate answers.

[Figure III-8] Grip Strength



#### ◆ Survey on physical measurement

- There are many kinds of physical measurement in relative to ageing. Grip strength was measured for KLoSA's baseline survey in 2006.
- Walking speed, as far as space allows, may be measured, balancing on one foot may be examined, or blood or saliva samples may be taken for test for longitudinal studies on ageing in other countries. Taking into account housing structure in Korea and practicality of measurement methods that interviewers can employ, grip strength was decided to be measured for the first baseline survey.
- Interviewers judge whether respondent is in a state that allows to measure grip strength. If respondent refuses measurement or either one of the hands is injured or aching, it will be abstained from measuring grip strength..

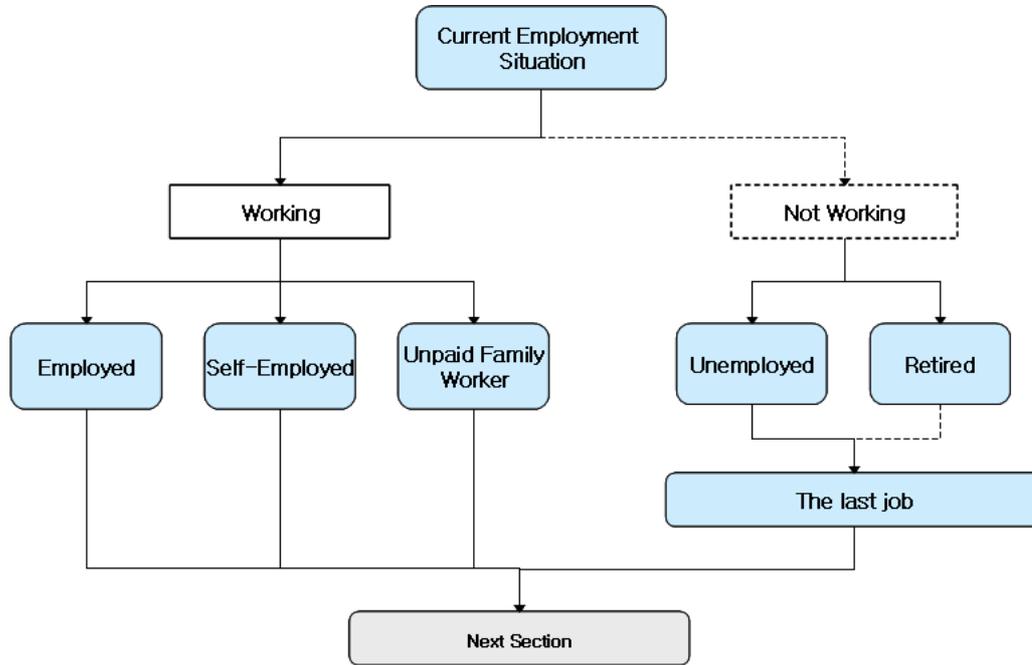
## 4) Employment

### ◆ Definition in Employment Category

- **Pay-earning job:** Pay-earning job means any work, including a side job, an irregular sideline, or pastime work as long as it earns pay. This may not meet the criteria of the employed definition in the Employment category.
- **Retirement:** A state that people stop doing career work. Either they are not engaged in any type of pay earning activities or even if they are engaged in such activities it is only for pastime. They are not intended to work any more seriously than they currently do unless there is a dramatic change in the future.
- **Full-time/Part-time worker:** Part-time worker refers to any employed person who works part-time or as a side job, whose normal hours of work are less than those of comparable full-time workers, or whose wage is calculated on an hourly basis. Full-time worker refers to employees who work full day.
- **Permanent worker :** Workers whose contract lasts one year or longer or who can remain employed as long as they wish without a pre-defined contract term.
- **Temporary worker :** Workers whose contract lasts one month or longer but shorter than one year, or whose work is scheduled to end within a year even if they do not have a pre-defined contract term. (Provided that any workers whose contract term is shorter than one year are deemed temporary workers even though they have been and will remain employed by the same employer for a long time.)
- **Daily worker :** Workers whose contract lasts shorter than one month, who are employed and paid on a daily basis, or who work and get paid based on their work without having a fixed workplace.
- **Temporary agency worker :** Workers who are paid by a temporary work agency, not by a user company and regulated by Temporary Work Act.
- **Workers provided by contract firms :** Workers who are paid by a temporary work agency, not by a user company and not regulated by Temporary Work Act.
- **Contract work-based payment arrangement :** It refers to an agreement made by and between a contractee who agrees to perform and complete certain work assignment and a contractor who agrees to make payment based on the result of the completed assignment. Completion of an assignment distinguishes it from employment or consignment and is a condition to make payment.
- **Survey structure based on current employment situation:** Respondent's current employment situation is clas-

sified by the questions **from D001 to D010.**

[FigureIII-9] Employment



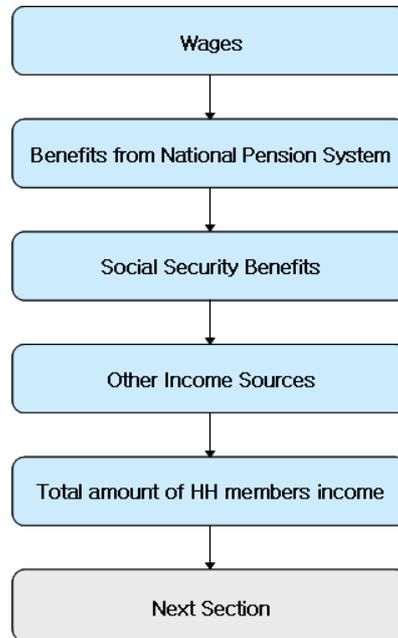
◆ Matters that require attention in terms of questions for employment situation classification (from D001 to D010)

- Employed refer to anyone who works under a contract with and receives wages from an employer.
- Self-employed refer to anyone who operates a business or works with tangible or intangible assets. Self-employed include writers and artists.
- Unpaid family workers refer to family members of the owners' who work. with no payment or allowances. Only those with 18 working hours or longer pursuant to the regulations of Economically Active Population Survey of the National Statistics Office are included. If a regular payment is made to family workers in any form, they are considered employed regardless of the 18 working hour requirement.
- Respondents who are engaged in multiple jobs are asked to answer "the most principal job".

◆ Once employment situation is determined in the introduction, employed will be directed to questions of D101 through D206, self-employed D301 through D362, unpaid family workers D401 through D452, unemployed D501 through D532 and retired D601 through D622. ***However, questions regarding the last job in D700s are added for unemployed in D500s with previous work experience and retired in D600s.***

## 5) Income

[Figure III-10] Income

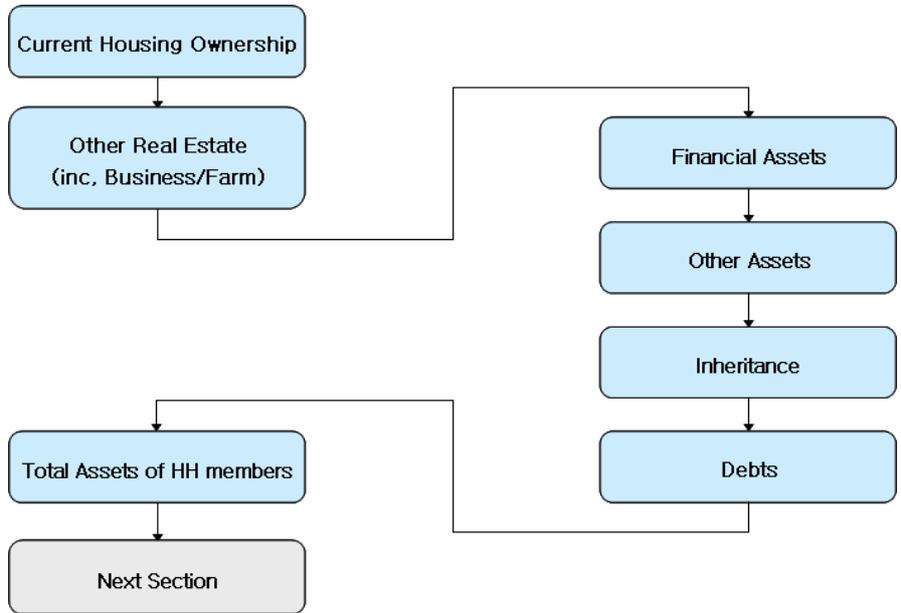


### ◆ Matters that require attention in terms of the Income category

- Respondent is asked to offer **his/her own income** as a response. More specifically, respondent shall not offer income of other household members such as spouse or children as his/her income. Respondent is also asked to offer **all types of income including labor income** as his/her income. That is, income in the Income category shall include labor income from all jobs that respondent is engaged in, including income from "the most principal job" in the Employment category, pension income and other income.
  - Income refers to "**income after tax**" and is in **ten million won**. "**Last one year(year 2005)**" refers to the period starting from January 1 throughout December 31, 2005.
  - There is a question about total income of a household in the last part of the Income category (E126). If a household has two eligible respondents or more, response on total income may vary depending on respondents. For reconciliation, please use "w01CV050\_r", a representative respondent variable for household total income.
- ◆ Respondents are most likely to answer "don't know" or "refuse to answer" in the Income category if they are unwilling to reveal their income. Therefore, dataset imputed through multiple imputation method may be useful. (※ Refer to Chapter 4 Item non-response and Imputation)

6) Assets and Debts

[Figure III-11] Assets and Debts



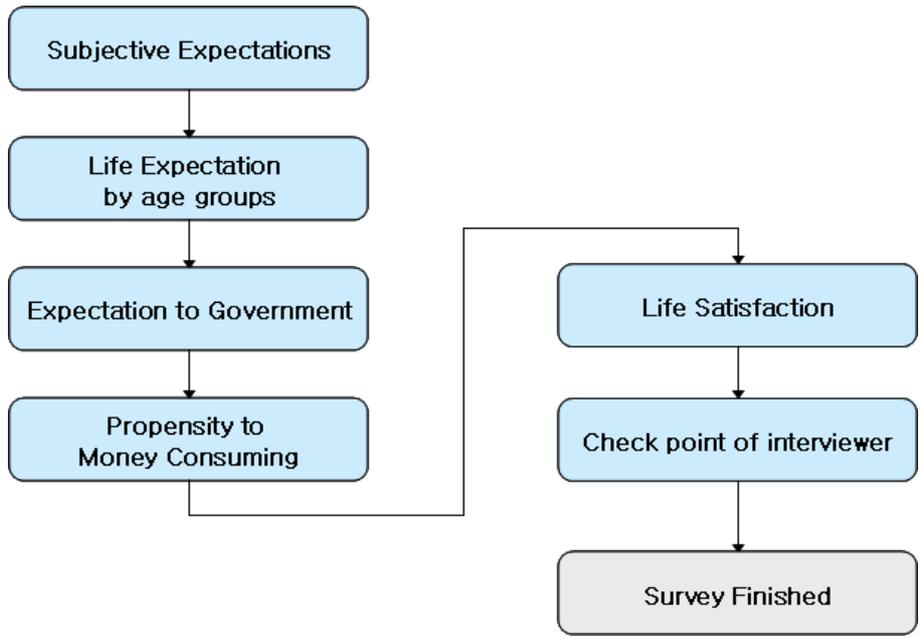
◆ Matters that require attention in the Asset category

- Asset in the Asset category includes real estate, business, financial asset, inheritance, gift and debts and respondents are asked to offer **assets under their name only**. For assets of which deed have multiple persons' name listed, the number of names is to be offered so that asset size information can be collected more accurately.
- Questions F001 through F048 deal with the asset value of current residence, followed by questions addressing other real estate. Therefore total asset value of real estate should be sum of current residence and other real estate.

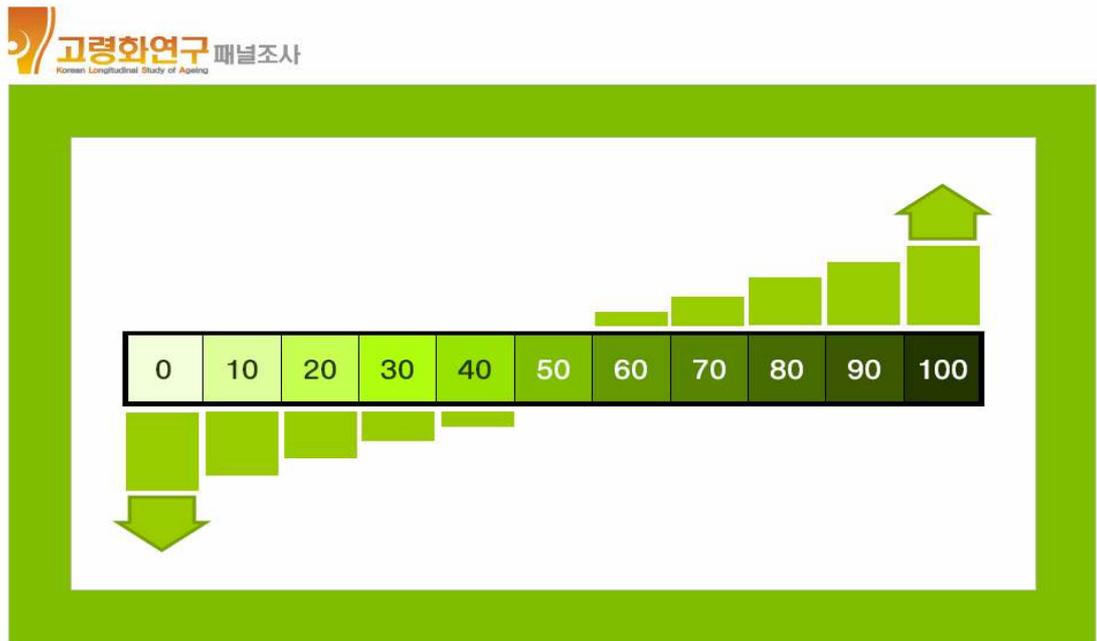
- ◆ As in the Income category, non-response such as "don't know" or "refuse to answer" are highly likely to occur for the Asset category. Dataset imputed through multiple imputation may be useful.  
(※ Refer to Chapter IV. For more detail, please download relevant documents in User's Guide at home page.)

7) Expectations and Life Satisfaction

[Figure III-12] Expectations and Life Satisfaction



[그림 III-13] Expectations and Life Satisfaction Showing Card



◆ **100-point scale and usage of picture cards.**

- Some expectation and life satisfaction surveys in advanced countries use percentage as a way to indicate respondent's opinion. However, 100-point scale picture cards are used to help respondent answering since it appears that some old-aged adults in Korea find it challenging to indicate their expectation and life satisfaction in percentage term due to lack of understanding.
- Expectation is expressed in a figure on **probability**. Respondent is asked to give interviewers a number from 0 to 100 where "0" means that respondent thinks there is absolutely no chance and "100" means that respondent thinks the event is absolutely sure to happen.
- The point given as a response to questions on life satisfaction is a figure on **satisfaction**. "0" means absolutely dissatisfied and "100" means absolutely satisfied.

## 4 The Rule of Variable Names

◆ Variable names basically follow the rule of "W01+question number".

◆ Rules for variable name of loop questions: 'basic variable name+sequence number' (see below box for illustration)

⇒ **Situation:** Respondent has a total of three living children and responds accordingly.

⇒ **Question number**

Ba01. Now I am going to ask you about your children. How many living children do you have?  
 \_\_\_\_3\_\_ [Range:0~20] (if the answer 0 then go to Ba47)

*[Loop: if Ba01 answer more than 1 then repeat the loop from Ba02 to Ba46, from Ba12 to Ba46 not living together child only]*

Ba02. What is your (the first/second/third ...) child's name?

[IWER: Please double check the spelling of the name, if the same name is used in Coverscreen. If R dose not provide the child's name, write down the birth order of child. e.g 'the first born' , 'the Second born'

\_\_Minna Kim\_\_\_\_\_

Ba03. Is your [Minna Kim] son or daughter?

- ① Son
- ⑤ Daughter

Ba04. How old is [Minna Kim]?

[IWER: For those who respond by zodiac sign, refer to the index card providing the according years]  
 \_\_\_\_18\_\_\_\_ [range:1~100]

⇒ **Rules for loop variable name:** For events or persons that fall under the loop description, **variable names shall be question number followed by "\_sequence number"**. In the above example where respondent has three living children, loop will run three times on questions Ba02 through 04. Variable names for the first, second and third child are generated as follows:

⇒ **Result**

- When respondent complete answers on **the first child, the first loop** will run.
  - The resulting variable names are W01Ba02\_01, W01Ba03\_01, and W01Ba04\_01.
- When respondent complete answers on **the second child, the second loop** will run.
  - The resulting variable names are W01Ba02\_02, W01Ba03\_02, and W01Ba04\_02.
- When respondent complete answers on **the third child, the third loop** will run.
  - The resulting variable names are w01Ba02\_03, W01Ba03\_03, and W01Ba04\_03.

◆ **Rules for variable name of multiple response questions: ‘Basic variable name + m + sequence number’**

- Multiple response questions will result in variables with as many 1 as responses and 0. Variable name will be created as question number followed by m1, m2, or m3. Here "m" represent multiple responses.

⇒ **Situation:** Respondents attend ① religious gatherings, ② social gatherings

⇒ **Variable name of multiple responses**

A017. Are you a member of any of the following organizations, clubs or societies? Please select all that apply.

- ① Church or other religious groups → **Variable name W01A017m1**
- ② Social clubs (e.g. private savings clubs, senior citizen's club, etc.) → **Variable name W01A017m2**
- ③ Entertainment clubs(Sport, music, arts etc.) → **Variable name W01A017m3**
- ④ Alumni society, society for people the same hometown etc. → **Variable name W01A017m4**
- ⑤ Volunteer groups → **Variable name W01A017m5**
- ⑥ Political party, NGOs, interest groups → **Variable name W01A017m6**
- ⑦ others → **Variable name W01A017m7**

⇒ **Result**

- W01A017m1 and W01A017m2 shall each have variable value of 1 and A017m3 through A017m7 shall each have variable value of 0.

◆ **Combination of loop and multiple responses: ‘Basic variable name +\_sequence number +m sequence number’**

- The most complicated combination of variable name is when loop and multiple response questions are combined. For variable name, the aforementioned rules are applied but the order shall be basic variable name followed by loop variable name and then multiple response variable name.

⇒ For example: R received regular monetary transfer from the first born and occasional monetary transfer from the second one

⇒ Variable name and value for the combination of loop and multiple answers

Ba15. Not counting any shared housing or shared food, during the LAST CALENDAR YEAR(2005) did you and your spouse receive any financial help from (the child name)? Please select all that apply.

① Yes, received regular monetary transfer

→ the first born child then the variable name **W01Ba16\_01m1**, value 1

→ the second one then the variable name **W01Ba16\_02m1**, value 0

② Yes, received occasional monetary transfer

→ the first born child then the variable name **W01Ba16\_01m2**, value 0

→ the second one then the variable name **W01Ba16\_02m2**, value 1

③ Yes, received non-monetary transfer

→ the first born child then the variable name **W01Ba16\_01m3**, value 0

→ the second one then the variable name **W01Ba16\_02m3**, value 0

⑤ No, I did not receive any financial help

→ the first born child then the variable name **W01Ba16\_01m5**, value 0

→ the second one then the variable name **W01Ba16\_02m5**, value 0

## Item Non-Response and Imputation

※ For the specific research papers for Multiple Imputation, please visit KLoSA's web site at (<http://www.klosa.kli.re.kr>) or click on the KLoSA banner at the KLI's web site at (<http://www.kli.re.kr>).

The baseline data were collected in 2006. In 2007, multiple imputation was conducted to handle missing values of income and asset related variables in the KLoSA baseline survey. This report describes a multiple imputation project conducted in 2007.

### 1 Imputation of Missing Values

To multiply impute missing values in KLoSA, we considered three imputation methods and selected hotdeck based on a modified predictive mean matching as our imputation method, because the simulation showed the best results. Hotdeck based on the modified predictive mean matching is a hotdeck imputation method suggested by Little(1998) and programmed by Bell at RAND (1999) using a SAS Macro.

Imputation was conducted with an adoption of several special characteristics of KLoSA. First, income or asset related variables often included nonresponse, or inconsistency existed between related items. To avoid bias due to nonresponse or inconsistency, the survey included unfolding bracket questions to obtain at least partial information about missing or inconsistent values. Second, in items with a few respondents, there were only a few respondents so that we could not find donors within a range based on information from unfolding bracket questions. Third, it was possible for some items to answer more than one responses per person. For example, insurance questions asked

participants to list the amount paid for each insurance. When a participant had more than one insurances, he/she was asked to list each amount paid for all listed insurances. Since the amounts paid for insurances by one person would be correlated each other, the prediction of missing amounts of insurance should include the correlation among insurances paid by the same person. Finally, there were many related items in the same session. Imputation should be conducted to achieve consistency among related variables. The following approaches describe the details to handle each property of KLoSA when multiple imputation was conducted.

The survey included eight sessions, and each sessions included many survey items. These items in the same session were often correlated each other. For example, total yearly paid work income depended on the number of months the participant worked. In KLoSA imputation, related variables were imputed using n-partition imputation (Marker et al., 2002). That is, a major variable with missing values was first imputed using hotdeck based on the modified predictive mean matching, and missing values in related variables were imputed as a group with observed values of donors. For example, if total yearly paid work income was missing for a participant, its value was imputed using hotdeck imputation and his/her missing months of work was imputed using his/her donor's observed months of work.

Income and asset variables often included items that should be the same among family members. For example, total family income or owning status of house should be the same among family members. However, since the survey was conducted individually, there were families with inconsistent answers or with answers missing in a few family members. We imputed missing values with answers of a family representative selected from the Coverscreen session. If a family representative was not measured, we imputed missing values with answers of a family member with the highest income (or asset). In the family session, a family representative in each generation of the family was selected to conduct the survey. Since other family members' answers were missing, their values were imputed with the ones from the family representative.

Multiple imputation was conducted separately between man and women. The KLoSA study targeted Koreans aged over 45, and the distribution of having a job, the amount of income or assets were very different between man and women. Therefore, it was expected that separate imputation between men and women would provide better imputation performance.

## 2 Application of Hotdeck Imputation Based on a Predictive Mean Matching

### ◆ Model 1 (Hotdeck)

This model was the original version of hotdeck imputation based on the modified predictive mean matching. It first fitted a regression model of a variable with missing values on other related predictor variables. This regression included only observed cases in the response variable, even if it could include missing values in predictor variables. With the fitted regression model, we calculated predicted values for both observed and missing cases of the response variable, and subclasses with more than 10 donors (if possible) were constructed based on stratification of predictive values. Within each subclass, donors were selected for missing values, and they were imputed with the donors' observed values.

### ◆ Model 2 (Bounded Hotdeck)

This model extended Model 1 to incorporate information obtained from unfolding bracket questions. KLoSA items included five unfolding bracket questions, resulting in six unfolding ranges. Since participants who answer unfolding brackets did not provide the exact value, but they provided information about the range they belong, multiple imputation was conducted by incorporating this information. For example, if a person answered unfolding brackets about his paid work income as "more than 2400 and less than 6000," his imputed paid work income should exist between 2400 and 6000. Therefore, we first separated observed values as one of six unfolding ranges divided with unfolding brackets. In each range, the observed values served as a donor pool for missing values within the same range. If enough donors existed within a range, they were divided as several subclasses based on predicted values in the regression and donors were selected from the same subclass within the same range.

### ◆ Model 3 (Hotdeck - GEE)

If a participant can provide more than one responses, these values are not independent each other. However, the standard regression model assumes independence between observations. Therefore, we extended the regression model to include correlated responses within the same participant, and the regression parameters were estimated using Generalized Estimating Equation (GEE) to handle this correlated covariance matrix. This method revised Model 1 to implement GEE in the estimation of the regression model by using SAS GENMOD procedure instead of SAS REG procedure. It similarly revised Model 2 to include correlated responses with unfolding bracket questions.

#### ◆ Model 4 (Regression)

In some items, missing values occurred within a top open bracket, but there were no observed cases in that range to serve as donors. For example, a man answered as the top open bracket in paid work income, but there was no observed value within the top open bracket. Since no donor existed, imputation of this person's paid work income was impossible. Therefore, we imputed this missing income with a predictive value of the regression. Hotdeck imputation based on the modified predictive mean matching calculated predictive values for all observations to construct subclasses, and we used these predictive values for imputing missing values without any donor. This regression imputation was applicable with a simple modification of Model 1 without fitting additional regression models.

### 3 Characteristics and Details of the Imputation Model in Each Section

#### ◆ A. Demographics

The demographics session was consisted of variables measuring age, educational background, marital status, religion and others. Among these variables, age(w01age\_001), religion(w01a014), the number of family numbers(w01hhsz), the number of households(w01gen\_num), region(w01region1), type of residence(w01enu\_type) and others were used as explanatory variables in multiple imputation. The variables 'married' and 'region3' were generated and the variables 'w01region1' and w01A014 were revised for this purpose.

#### ◆ B. Family and Family Transfer

We imputed only children's demographic information and regular/irregular monetary transfer in the Family and Family Transfer session. The survey of this session was conducted to a family representative only in each generation of the family, and we filled in their answers to other family members. If a family representative did not respond, we first imputed their values, and other family members' answers were filled in using imputed answers from their representative. The variables about regular/irregular monetary transfer were consisted of indicator variables about the existence of any transfer, exact amounts of transferred money if there were any monetary transfer, and unfolding bracket questions if respondents refused to answer exact amounts of support. For example, the variable w01BA15 asked whether respondents had received monetary transfer from their children. If the answer of w01BA15 was 1, it meant "Yes, I received monetary transfer from children(s)" and if the answer was 0, it indicated "No, I don't receive." If respondents provided their exact transfer amounts, these observed values were used as donors for imputation. We imputed the amount of monetary transfer for each children independently, not the total amount of all children. If a person answered for unfolding bracket questions, we generated ranges in accordance with the answer from bracket questions. If there were respondents who did not provide both the exact amount and partial information using unfolding brackets, these

observations didn't have any bound, and we used all observed values as donors for imputation.

### ◆ C. Health

The Health session was consisted of health status(Ca), functional limitations and helpers(Cb), health insurance and services(Cc), cognition(Cd), and grip strength(Ce) questions. In this session, imputation was implemented for variables to be used as predictor variables in imputation of later sessions. The nonresponse rates were less than 1% for most variables, and there were many indicator variables about the diagnosis of diseases or a need for help. Those variables were imputed using design variables and basic predictor variables from the demographics session. After imputations were finished, predictor variables for later imputations were generated.

### ◆ D. Employment

Respondents of the employment session were divided into the wage earner, the self- employer, the unpaid family worker, the retiree, the job seeker and others. Unfolding bracket questions were not applied in this session. Depending on their job status, different survey questions were asked. For example, Since only wage earners were surveyed for monthly income(w01D155), the number of applicable cases of w01D155 was 1875. Similarly, since only self-employers were surveyed for monthly income (w01D315), the number of applicable cases of w01D315 was 1695. Therefore, we selected only applicable cases to conduct imputation for each variable and merged the imputed cases with other unapplicable cases. The others imply participants who were not a retiree and did not seek a job currently. They were asked about the current employment status(w01D001-w01D010) and skipped to the most recent job(w01D701-w01D722). After imputing major employment variables, we generated new variables 'wage1'(main income) and 'wage2'(income from side job) to use them as predictor variables in later imputations. The Variable 'wage1' measures the main income of the wage earner, the self-employer and the participants who quitted the most recent job in 2006. The Variable 'wage2' measures monthly income from side jobs of the wage earner, the self-employer and the retiree.

### ◆ E. Income

The Income session was constructed with several sources of income. Each source of income variables was composed of an indicator variable, a variable to measure the exact amount, and unfolding bracket questions. For example, w01E001 asked whether participants had any work for pay in the last year. The response of "1" indicated that participants had any paid work income. The response of "5" indicated that participants did not have any paid work income. Missing values in the indicator variable were imputed by using demographic and design variables as predictors. Under the existence of the paid work income, we used participants with observed exact income as donors to impute missing paid work incomes, If the amount of the paid work income was missing but bracket questions were observed, we created ranges using unfolding brackets. If bracket

questions were also missing, we imputed missing values within the whole range. Variables about pension were measured as receiving monthly income, receiving lump sum payment, and receiving both monthly and lump sum payment. Since the amount received depended on the type they received, we classified participants by the type they received. The amount was imputed separately by the type. The details of imputation were similar for other income variables.

#### ◆ F. Assets And Debts

The Assets and debts session was divided into six parts; housing and real estate, business or farm, financial assets, other financial assets, personal loan, and income. Each part was consisted of indicate variables, variables about the exact amount, and bracket questions. All variables about the exact amount, except interest(w01F166), were either answered or skipped by the answer of indicate variables. If the answer of an indicator variable was "1", then participants had to answer the exact amount, and if the answer of the indicator variable was "5", then participants skipped the rest of that part. In addition, if the exact amounts were missing, then participants were asked to answer unfolding bracket questions, Five unfolding brackets were used to make a range of possible response, and imputation was implemented with this range information. If the bracket questions were missing, the values were imputed without bounds. Observed values served as donors in the imputation. Family members should report the same amount for housing and real estate variables, Therefore, if a value was missing, then we imputed the value with the answer of a family representative. If we cannot find values of the family representative, we imputed them individually. The financial assets part was constructed of cash, bank savings, deposits, insurance and so on. Insurance variables were measured by very detailed formats, the nonresponse rate was very high, and the number of donors was very small. That values were imputed by Model 3 that reflects correlations of more than one amounts from one person. In this model, predictor variables were the number of insurances that each person insured and dependent variables were the amounts of money paid for each insurance. Term life insurance, whole life insurance and annuity insurance used this model.

## GENERATED VARIABLES

General variables represents only respondent's answers. KLoSA team makes Generated variables, focusing on frequently using variables, based on general variables. The variables lists as follow and it is located the end of every section. If you want to know how dose it make from? Please see the Homepage.

A. DEMOGRAPHICS	
Variable Name	Variable explanation
w01edu	R's educational background arrange
w01respid1	Among family members respondent order (the number a respondent recognizes what number Jjae among 45 years old ideal, and informing during total household circles).
w01year	A respondent was born make.
w01a001_age	Respondent age.
w01gender1	Respondent sex.
w01respid2	The number it is to a spouse of a respondent, and informing (if there is a price, there is a spouse, and there is not a GyeolCheukChalMyeon spouse, or below 45 years old).
w01year2	A respondent spouse was born make.
w01age2	Age of a respondent spouse.
w01gender2	Sex of a respondent spouse.
w01hhsiz	A family member number.
w01gen_num	Form of household
w01e_num	A number circular the household whom you shall respond to to this questionnaire.
w01CID	Couple ID
w01c_num	Classify man and wife unit in a household.
w01region1	Regional 1: a local variable
w01region2	Region 2: village/town
w01region3	Region 3: Metropolitan/City/Town
w01enu_type	A residential form.
w01enu	enumeration district
w01mniw_y	The this questionnaire interview date (a year).
w01mniw_m	The this questionnaire interview date

	(a month).
w01mniw_d	The this questionnaire interview date (work).
w01wgt	Weights

Ca. Health Status	
Variable Name	Variable Explanation
w01Ca_list	A random grant.
w01bmi	BMI
w01body	According to BMI obesity level
w01smoke	Smoking duration(Unit: months)
w01smkterm	Smoking status
w01alc	Drinking status
w01alcterm	Drinking duration(Unit: months)
w01soju	Soju drinking or not
w01beer	Beer drinking or not
w01makgeolli	Macgeolli drinking or not
w01wisk	wiskey drinking or not
w01wine	wine drinking or not
w01addic	Drinking behavior
w01dep1	Depress status
w01dep2	CES-D10 Standard Depress status

Cb. Functional Limitations and Helpers	
Variable Name	Variable Explanation
w01adl	ADL index
w01iadl	IADL index

Cd. Cognition	
Variable Name	Variable Explanation
w01mmse	Cognition marks
w01mmseg	Cognitive function status

Ce. Grip Strength	
Variable Name	Variable Explanation
w01mgrip	Grasping power index

D. Employment	
Variable Name	Variable Explanation
w01ecoact	Labor force participation status
w01empdur	R's employment status in interviewing time (Wage workers/Self employed/non wage family worker)
w01lastempdur	R's last employment duration but R is not employed statue at the interviewing time
w01D103ind	Wage workers: Industrial classification(1 digit)
w01D308ind	Self employed: Industrial classification(1 digit)
w01D405ind	non wage family worker: Industrial classification(1 digit)
w01D518ind	job seeker: Industrial classification(1 digit)
w01D707ind	the last job: Industrial classification(1 digit)
w01D103indm	Wage workers Industrial classification (2 digits)
w01D308indm	Self employed: industrial classification(2 digits)
w01D405indm	non wage family worker: industrial Classification(2 digits)
w01D518indm	job seeker: industrial classification (2 digits)
w01D707indm	the last job: industrial classification (2 digits)
w01D109occ	Wage workers: occupational classification (1 digit)
w01D197occ_h	Wage workers: occupational classification (1 digit) of the desired job
w01D314occ	Self employed: occupational classification (1 digit)
w01D353occ_h	self employed: occupational classification (1 digit) of the desired job
w01D407occ	non wage family worker: occupational classification (1 digit)
w01D445occ_h	non wage family worker: occupational classification(1 digit) of the desired job
w01D519occ	job seeker: occupational classification (1 digit)
w01D610occ_p	retired: occupational classification (1 digit) of the pastime
w01D710occ	the last job: occupational classification (1 digit)
w01D109occm	Wage workers: occupational classification (2 digits)
w01D197occ_hm	Wage workers: occupational classification (2 digits) of the desired job
w01D314occm	Self employed: occupational classification (2 digits)
w01D353occ_hm	self employed: occupational classification (2 digits) of the desired job
w01D407occm	non wage family worker: occupational classification (2 digits)
w01D445occ_hm	non wage family worker: occupational classification (2 digits) of the desired job
w01D519occm	job seeker: occupational classification (2 digits)
w01D610occ_pm	retired: occupational classification (2 digits) of the pastime
w01D710occm	the last job: occupational classification (2 digits)

E. Income	
Variable Name	Variable Explanation
w01CV050_r	A person knowing incomes during the answer objects the best.
w01CV050_rn	A household total Income information source.(missing value replacement)
w01incfirst	the 1st rank of individual income within HH

F. Assets and Debts	
Variable Name	Variable Explanation
w01F235	Total individual asset for not interviewing person, even if aged 45 over in HH
w01CV051_r	A person knowing assets during the answer objects the best.
w01CV051_rn	A household total Assets information source. (missing value replacement)

Commenly Generated Variables	
Variablename_ct (e.g.:w01Ba16_01ct)	Unfolding Bracket Value of w01Ba16_01ct

## Data Access

All data relevant to KLoSA will be made available through KLoSA's web site (<http://www.klosa.kli.re.kr>) or the KLI's web site (<http://www.kli.re.kr>). If you visit the KLI's web site, please click on KLoSA's banner.

Please log in to download raw data after member registration. Other documents including user's guide, code book, questionnaire, fieldwork report, non-response imputation report do not require login to download from the web site. You will find questionnaire and code book also in Data Service System of the Korea Labor Institute where raw data resides and those are identical to files found in the KLoSA web site. These data will not be printed or made in the CD format for distribution. Users should download from their computer and make hard copy for their use.

Raw data, questionnaire, code book and user's guide shall be updated reflecting changes in data and addition of generated variables. A unique version number will be given to each update to help users identify the latest version at home page.

English version for questionnaire, code book, data, and some documents shall be provided for non-Korean speaking researchers.

◆ **SPSS data was prepared at 12.0 version and SAS at 9.0 version. Therefore, errors in variable names might occur at versions below.**

## ◆ Data download

- Complete member registration at home page and log in with appropriate ID(registered email) and password. When you click on the Data Download button, you will be automatically directed to Data Service System where data resides.
- **Note!!**  
Users should have **pop-up blocker disabled** on your browser to enter Data Service System.
- [Figure VI-1] When users log in for data download, pop-up window will appear to direct users to Data Service System of the KLI as indicated in the figure above. The screen is composed of three distinctive sections. The first section is a rectangular box on the left, the second is the largest part in the middle of the right, and the third is a horizontal box in the upper side. The description is as follows:
- Users can download either entire dataset or selected variables. If users wish to download the entire data, go to "Download entire dataset" in the third section. If users wish to download partial data by subject or keyword, go to "Download selected variables" in the second section.

[Figure VI-1] Data Service System Structure

The screenshot displays the 'Data Service System' interface. On the left (1), there is a 'contents of table' sidebar for the 'Korean Longitudinal Study of Ageing', showing a hierarchical tree of variables. At the top (3), there are navigation buttons for 'logout', 'korean', and 'help'. The main area (2) is titled 'Download Entire Dataset' and includes buttons for 'Data', 'Questionnaire', and 'Codebook'. Below this, the 'Download Selected Variables' section (2-1) features search filters for 'Survey Year' (2006) and 'Subject' (employment). It lists 'Downloadable Variables' (2-2) and 'Selected Variables' (2-3) with checkboxes and arrows for selection. The bottom of the page shows the KLI logo and copyright information: 'COPYRIGHT(C) 2005 BY KLI.RE.KR. ALL RIGHTS RESERVED BY KOREA LABOR INSTITUTE. WEBMASTER@KLI.RE.KR'.

### ◆ Download entire dataset

- Users who wish to download entire dataset may download data (raw data), questionnaire and code book in "Download entire dataset", marked 3 in the Figure VI-I. All data is in zip file both in English and Korean for user's convenience.

#### · Tips on raw data download!

Data provided through "Download entire dataset" is in SPSS or SAS and in Korean or English, all told, four different forms (Korean SPSS, English SPSS, Korean SAS, English SAS). It corresponds to 1 to 4 in terms of data name. This data reflects respondent's response where "don't know" or "refuse to answer" are left untouched. Generated variables are created by combining variables from survey questions. These variables are added at the end of each category. Data corresponding 5 to 8 is data after multiple imputation of non-response like "don't know" or "refuse to answer". The data does not include variables of survey questions, rather it is imputed dataset of non-response mainly in the Asset or Income category.

(※ Refer to Appendix for how to use multiple imputed data.)

Year	Data Name	Data Explanation
2006	<a href="#">w01_v1.0k(SAS).zip</a>	1st wave 1.0 version(A section~G section)_Korean SAS data
2006	<a href="#">w01_v1.0k(SPSS).zip</a>	1st wave 1.0 version(A section~G section)_Korean SPSS data
2006	<a href="#">w01_v1.0e(SAS).zip</a>	1st wave 1.0 version(A section~G section)_English SAS data
2006	<a href="#">w01_v1.0e(SPSS).zip</a>	1st wave 1.0 version(A section~G section)_English SPSS data
2006	<a href="#">w01_i_v1.0k(SAS).zip</a>	1st wave 1.1 version(Missing Value Imputation & Generated Variables)
2006	<a href="#">w01_i_v1.0k(SPSS).zip</a>	1st wave 1.1 version(Missing Value Imputation & Generated Variables)
2006	<a href="#">w01_i_v1.0e(SAS).zip</a>	1st wave 1.1 version(Missing Value Imputation & Generated Variables)
2006	<a href="#">w01_i_v1.0e(SPSS).zip</a>	1st wave 1.1 version(Missing Value Imputation & Generated Variables)

### ◆ Download selected variables

- Users can download selected variables of their interest. There are two ways to do so. First, as indicated 2-1 in Figure VI-1, users can search and download variables by subject or keyword in the search conditions. Second, users go to 1 in Figure VI-1, click open variables by category, select a category of their interest, read through variable explanation of actual data and download variables that interest them.

- **Download by search conditions:** Refer to 2-1 in Figure VI-1

For search by subject, users will find a list of subjects identified by researchers and select a subject of their interest to find downloadable variables in 2-2 Downloadable variable. For search by keyword, users shall enter keyword of their interest using "AND" or "OR" to create variables that suit users' interest. Once users select variables by subject in 2-1, all corresponding variables will be created in 2-2. Users shall

check the box in the left for variables of their interest and click a right arrow in the middle to move checked variables to the right. The format of data can be selected in 2-4. Code book for selected variables can be downloaded by clicking on 2-4 Download Code book.

· **Select variables from table list and download**

Make sure that Korean Longitudinal Study on Ageing is selected in the box of contents of table and click on folder to find "1st Wave 1.0 Version. Click on it and find a drop down menu of each category that goes from A Demographics to G. Expectation and Life Satisfaction. Folder and + indicate that there remain subcategories for users to click open and pencil and paper indicate no subcategories are remaining. Users will find the entire categories of data here. A click on a category of their interest will create all corresponding variables in 2-2 Downloadable variables. The rest is identical to what is described in download by search conditions above. That is, select variables at 2-2, move them to 2-3 and download at 2-4.

· **Function of 2-2 downloadable variables table**

2-2 Downloadable variables table is where variables selected through the drop down menu or through 2-1 search conditions are gathered. Check the square box on the left of downloadable variables at 2-2 that interest you. All of the variables gathered at 2-2 will be selected by checking a square box on the top. A click on arrows between 2-2 and 2-3 will move variables from 2-2 to 2-3 or vice versa. A double arrow means move all the variables from one place to another. A simple arrow means move only selected variables. Clear button will remove all the variables from the table.

· **Function of 2-3 Selected variables table**

2-3 Selected variables table is where variables selected from 2-2 Downloadable variables are gathered. Users can repeat the process as many times as they wish to accumulate all the variables of their choice. In doing so, users can start either from table contents or from search conditions. Clear button will remove all the variables from the table.

· **Function of 2-4 download selected variables and download code book**

Finally here users can download variables selected at 2-3 in the format of their choice: text, CSV, excel, or SAS. Users can also download code book for the variables selected.

◆ Any additional information regarding 2006 KLoSA's 1st Wave 1.0 Version will be updated in user's guide at homepage.