Harmonized LASI Pilot Data Documentation

Version A

Chiaying Sandy Chien, Kevin Carter Feeney, Jenny Liu, Erik Meijer, Jinkook Lee

RAND Labor & Population

WR-1018

October 2013

This paper series made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the NICHD funded RAND Population Research Center (R24HD050906).

RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Labor and Population but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

RAND

***Harmonized LASI Pilot Data Documentation,
Version: A***

*Chiaying Sandy Chien, Kevin Carter Feeney, Jenny Liu, Erik Meijer,
and Jinkook Lee*

October 2013

*This project is funded by the National Institute on Aging,
National Institutes of Health (R03AG043052)*

Labor & Population Program

Preface

The Longitudinal Aging Study in India (LASI) is a multidisciplinary panel study designed to be nationally representative of India's population aged 45 and older. LASI collects information conceptually comparable to that gathered by the Health and Retirement Study (HRS) in the United States, and its sister surveys in Asia, Europe, and elsewhere. The LASI pilot was conducted in late 2010 and the data have been released for research use.

The HRS data is rich and complex. To make it more accessible to researchers, the RAND Center for the Study of Aging, with support from the National Institute on Aging (NIA) and the Social Security Administration (SSA), created the RAND HRS data files. The RAND HRS is a user-friendly version of a subset of the HRS. It contains cleaned and processed variables with consistent and intuitive naming conventions, model-based imputations and imputation flags, and spousal counterparts of most individual-level variables. The construction of the data is elaborately documented, with special attention to comparability of variables across survey waves. See Chien et al. (2013) for the documentation of the RAND HRS.¹

With funding and support from NIA/NIH, the Harmonized LASI file has been created to serve a similar purpose as the RAND HRS. The file includes the variables with names and definitions that mimic corresponding RAND HRS variables. This document describes the data. LASI data and further information is available at <https://www.g2aging.org>.

LASI is a joint project of three partnering institutions: Harvard School of Public Health, the International Institute of Population Sciences (IIPS), and the RAND Corporation. The LASI Principal Investigator (PI) team consists of David Bloom, the Clarence James Gamble Professor of Economics and Demography at the Harvard School of Public Health, Perianayagam Arokiasamy, Professor in the Department of Development Studies at IIPS, and Jinkook Lee, at that time Senior Economist at the RAND Corporation and Professor at the Pardee RAND Graduate School and currently Professor at the University of Southern California's Andrus School of Gerontology and Adjunct Senior Economist at the RAND Corporation.

A team from Harvard University, IIPS, and the RAND Corporation also provided their expertise to the project, including Lisa Berkman, David Canning, Amitabh Chandra, Nicholas Christakis, Ajay Mahal, and S.V. Subramanian from Harvard University; Sulabha Parasuraman, P. Sanjay Mohanty, and T.V. Sekher from IIPS; and Bas Weerman and Adeline Delavande from RAND.

Several other researchers contributed to the LASI pilot, including Peifeng Hu, Associate Professor at the University of California, Los Angeles Medical School; Arun Risbud, Senior

¹ Chien, S., et al. (2013). *RAND HRS Data Documentation, Version M*. Santa Monica, CA: RAND Center for the Study of Aging. <http://hrsonline.isr.umich.edu/modules/meta/rand/randhrsm/randhrsM.pdf>. More general information about the RAND HRS can be found at <http://www.rand.org/labor/aging/dataproducts/hr-data.html>.

Deputy Director and Head of the Microbiology Division, National AIDS Research Institute, Pune, India; Arvind Mathur, pProfessor and Unit Head of the Department of Medicine, Shastri Nagar Medical College, and In-Charge, Geriatric Centre, MDM Hospital, Jodhpur, Rajasthan; and Steven Heeringa, Senior Research Scientist, University of Michigan.

Table of Contents

PREFACE.....	2
1. INTRODUCTION AND OVERVIEW	5
1.1. Gateway to Global Aging Data	6
1.2. Data File Structure	7
1.3. Variable Naming Convention	7
1.4. Missing Values	8
1.5. Sampling Plan	9
1.6. Weighting and Survey Design	9
2. IMPUTATION OF INCOME AND CONSUMPTION	11
2.1. Background	11
2.2. Imputation Process	11
2.2.1. Ownership Imputation	12
2.2.2. Amount Imputation	15
2.3. Derived Variables	16
2.3.1. Household income Measures	16
2.3.2. Household Consumption Measures	16
2.3.3. Individual Earnings and Pension Measures	16
3. STRUCTURE OF THE CODEBOOK	18
4. DATA DOWNLOAD	21
5. DATA CODEBOOK.....	23
Section A: Demographics and Identifiers	24
Section B: Health.....	43
Section D: Income	67
Section H: Family Structure	78
Section K: Consumption	81

1. Introduction and Overview

This codebook documents the Harmonized Longitudinal Aging Study in India (LASI) file. The Harmonized LASI file is a user-friendly version of the LASI pilot data specifically designed for harmonization with the RAND version of the Health and Retirement Study (RAND HRS) and its sister studies, including the Harmonized English Longitudinal Study on Ageing (Harmonized ELSA), the Harmonized Survey of Health, Ageing, and Retirement in Europe (Harmonized SHARE), the Harmonized Korean Longitudinal Study of Aging (Harmonized KLoSA), the Harmonized Japanese Study of Aging and Retirement (Harmonized JSTAR), and the Harmonized China Health and Retirement Longitudinal Study (Harmonized CHARLS).

The Longitudinal Aging Study in India (LASI) was designed to be a nationally representative sample of India's population aged 45 and older. Its primary objective is to gather longitudinal data on the social, economic, and health situation of older people throughout India, so as to provide policymakers with information needed to improve the lives of the elderly. It is supported by the National Institute of Aging.

The LASI pilot study, conducted in 2010 in four Indian states, is a sample of 1,683 individuals within 950 households. The survey instrument is conceptually comparable to that for the U.S. Health and Retirement Study (HRS), and to those for similar HRS-type surveys in South Korea, Japan, China, Indonesia, Mexico, England, Ireland, and more than fifteen continental European countries and Israel. The survey also captures population characteristics specific to India.

The HRS is a national panel survey of individuals in the US over age 50 and their spouses. Its main goal is to provide panel data that enable research and analysis in support of policies on retirement, health insurance, saving, and economic well-being. The survey elicits information about demographics, income, assets, health, cognition, family structure and connections, health care utilization and costs, housing, job status and history, expectations, and insurance. The RAND HRS is a user-friendly version of a subset of the HRS. It contains cleaned and processed variables with consistent and intuitive naming conventions, model-based imputations and imputation flags, and spousal counterparts of most individual-level variables.

The LASI survey instrument comprises the household survey and individual survey. The household survey was conducted with one selected key informant in each household. The individual survey was conducted with each respondent. The Harmonized LASI pilot data include all the individuals interviewed. This includes individuals who were age-eligible (at least 45 years old) and their spouse regardless of age.

1.1. Gateway to Global Aging Data

The Health and Retirement Study (HRS) has achieved remarkable scientific success, as demonstrated by an impressive number of users, and research studies and publications, utilizing the HRS. Its success has generated substantial interest in collecting similar data in other countries as population aging is experienced and is progressing in every region of the world.

Consequently, a number of surveys have been designed to be comparable with the HRS: the English Longitudinal Study of Ageing (ELSA), the Survey of Health, Ageing, and Retirement in Europe (SHARE), the Mexican Health and Aging Study (MHAS), the Korean Longitudinal Study of Aging (KLoSA), the Japanese Study on Aging and Retirement (JSTAR), the China Health and Retirement Longitudinal Study (CHARLS), the Indonesia Family Life Survey (IFLS), the Irish Longitudinal Study on Aging (TILDA), the Study on Global Ageing and Adult Health (SAGE), and the Longitudinal Aging Study in India (LASI). The overview of this family of surveys, including their research designs, samples, and key domains can be found in Lee (2010).²

As these surveys were designed with harmonization as a goal, they provide excellent opportunities for cross-country studies. The value of comparative analyses, especially the opportunities they offer for learning from policies adopted elsewhere, is widely recognized. Yet, there are only a limited number of empirical studies exploiting such opportunities. This is partly due to the difficulty associated with analyzing multiple, complex surveys and the policies and institutions for each country.

Identifying comparable questions across surveys is the first step toward cross-country analyses. Our team has developed the Gateway to Global Aging Data (G2G) to help users understand and utilize these large-scale population surveys on health and retirement. G2G includes several tools to facilitate cross-national health and retirement research. It includes a digital library of survey questions for all participating surveys. Its search engines enable users to examine cross-country concordance for each question. And using these tools, researchers can identify all questions related to particular key words or within a domain of interest.

G2G can be accessed at <https://www.g2aging.org>. For more information about obtaining the Harmonized LASI data from G2G see Chapter 4.

² Lee, J. (2010). Data set for pension and health: Data collection and sharing for policy design, *International Social Security Review*, 63(3-4), 197-222.

1.2. Data File Structure

The Harmonized LASI Data is distributed as a single file, which includes the pilot wave of data. The unit of observation is an individual. Each individual is identified by a unique identifier HHIDPN. Households are identified by HHID. It is important to note that unlike the HRS, households in the LASI may include multiple couples and/or singles living together. This file may be merged with other LASI data using HHIDPN.

1.3. Variable Naming Convention

With few exceptions, variable names in the Harmonized LASI Data follow a consistent pattern. The first character indicates whether the variable refers to the reference person (“R”), spouse (“S”), or the household (“H”). The second character indicates the wave to which the variable pertains: “0”, or “A”. The “0” indicates pilot wave. The “A” indicates “all,” i.e., the variable is not specific to any single wave. An example is RABDATE, the birth date of the respondent. The remaining characters describe the concept that the variable captures.

For example (see below), variable R0HLTHLM captures whether the respondent experiences an impairment or health problem that limits the kind or amount of paid work he/she can do.

Variable names in the Harmonized LASI are generally based on the variable name used in the RAND HRS for the same measure. Measures which are exactly or near-exactly comparable between the Harmonized LASI and RAND HRS use the exact same name. For instance RABYEAR is the variable name for the respondent birth year in both the Harmonized LASI as well as the RAND HRS. If the Harmonized LASI measure is deemed only somewhat comparable with the RAND HRS version of that measure, the variable name in the Harmonized LASI will often end in “_L.” This variable name suffix indicates some LASI-specific difference with RAND HRS version of this measure.

Other reasons for Harmonized LASI-specific variable names include: differences in survey questions, scale, survey routing, and imputed values. Harmonized LASI-specific variable names are used to notify the user that there are substantial differences between the RAND HRS and the Harmonized LASI measure, and clean harmonization between these measures is not possible.

For wealth and income measures, the Harmonized LASI do not use LASI-specific variable names even though wealth and income measures in the Harmonized LASI are expressed in Rupees while income and wealth measures in the RAND HRS are expressed in nominal dollars, and question texts and components of income and wealth differ from the RAND HRS.

Users should always check the “Differences with RAND HRS” section of each measure before comparing any Harmonized LASI measure to the RAND HRS version of the same measures or any other Harmonized Dataset version of the same measure.

1.4. Missing Values

Variables may contain missing values for several reasons. SAS and Stata offer the capability to distinguish multiple types of missing values, and we have attempted to record as much information as possible. Generally, the codes adhere to the classification in Table 1.

Table 1. Missing Codes

Cod	Reason for missing
.	Reference person did not respond to this wave
.D	Don't know
.R	Refused
.N	N/A
.U	Reference person is not married (for spouse variables)
.V	Spouse did not respond this wave (for spouse variables)
.M	Other missing

The coding scheme varies across variables. Consult the Data Codebook for details on individual variables.

Stata introduced the ability to distinguish multiple types of missing values in its Version 8. The Harmonized LASI files in Stata format are constructed to be used with Stata Version 8 or later.

1.5. Sampling Plan

The survey was conducted in four Indian states—Rajasthan and Punjab in the north and Kerala and Karnataka in the south—and fielded in the dominant language of each state. These four states were chosen to capture the demographic, economic, health, and cultural diversity of India.

The sampling plan was based on the 2001 Indian Census. Two districts were randomly chosen from each state as shown in Appendix Figure 1. Within these districts, eight primary sampling units (PSUs) were chosen to be surveyed. Primary sampling units were stratified across urban and rural districts within each of the four states to capture a variety of socioeconomic conditions. Rural PSUs with fewer than 500 households were then selected through a two-stage sampling procedure, while urban PSUs and rural PSUs with more than 500 households were selected through a three-stage procedure. This type of sampling plan introduces more sample-to-sample variability than simple random sampling, which has important implications for analysis.

Eligible households were defined as those with at least one member 45 years of age or older. Eligible individuals were 45 years of age or older and their spouses, regardless of the spouse's age. LASI randomly sampled 1,546 households from these stratified PSUs, and among them, 950 households with a member at least 45 years old were interviewed. Among these households, LASI data was collected from 1,683 individuals between October and December, 2010.

1.6. Weighting and Survey Design

The Harmonized LASI includes variables to allow users to produce weighted estimates with survey design adjusted standard errors. Weights are created using the inverse probability of selection combined with household and individual response rates. To calculate weights for the data, we used forecast estimates of the 2010 population based on the 2001 census round. Weights are based on rural and urban population counts of the population 45 years and older within each state.

There are two sets of weights in the data: one for households and one for individuals. For the household weight, there are two levels: one (HH_WT_POOLED) that is used to create a sample representative of the population across the four LASI states; and one (STATE_HH_WT) that is used to examine households within a given state.

For the individual weight, there are also two weight variables: one (INDI_WT_POOLED) that is used to make a sample representative across the four states; and one (STATE_INDI_WT) that is used to examine a representative group of individuals within a particular state.

In addition to weights, LASI also provides some stratification and cluster variables to account for LASI's survey design. They are STATE, DISTRICT, PSU, and RESIDENCE (rural/urban status). There is also a STRATA variable that the LASI team derived from these

sampling variables. This variable should be used to stratify standard errors on state, district, and rural/urban residence in any analysis.

2. Imputation of Income and Consumption

2.1. Background

Most LASI questions on household income, household consumption and individual earnings follow the same pattern as in the HRS. Consider income from any household members as an example. First, the interviewer asks whether any household member received any wages or other income from employment. If affirmative, the interviewer asks the value of total earnings. If the respondent is unable or unwilling to provide an exact amount, the interviewer asks whether it is more than 3,400 Rupees. Depending on the response, income ranges are progressively introduced to narrow the final range to one of the following: 0-1,700; 1,700-3,400; 3,400-6,000; 6,000-13,500; 13,500-32,000; 32,000 or more. These ranges are known as “brackets;” the sequence of probes into increasingly narrow ranges is known as “unfolding brackets” questions. The brackets vary by income, asset, and consumption questions.

The respondent may opt out of the question sequence at any time. As a result, the raw data contain valid zero-value responses, exact amounts, complete bracket responses, incomplete bracket responses, and claims of ownership without a corresponding value. An incomplete bracket happens if the respondent provided some information but was unable or unwilling to respond through to the last unfolding bracket probe. For example, a respondent can indicate that the income amount is more than 13,500, but then refuses to tell whether the amount is more than 32,000. In this case, the final range is open-ended at 13,500 or more. A claim of ownership without a corresponding value results if the respondent indicated that that household member has income, but did not reveal the exact amount or a range. A claim of ownership without a corresponding value is a special case of an incomplete bracket, namely an open-ended bracket of greater than zero dollars.

In summary, the data contain valid responses and several types of responses that require imputations. These situations are listed below in decreasing order of informational content:

- Case 1: We may know a “complete” range of values;
- Case 2: We may know that the household has the income type, but have no information on its value, or only coarse information in the form of incomplete brackets; or
- Case 3: We may not even know whether the household has the income type, much less its value.

2.2. Imputation Process

As defined previously, there are three types of missing values that require separate types of imputation. The imputation process is progressive in the sense that we first impute ownership for those for whom nothing is known. Given ownership, we impute exact amounts. In all imputations, we use all available information. In particular, where brackets are known, we impute values in the given range. The process here is a simplification of the imputation method used for the RAND HRS. Apart from differences within each imputation stage, we omit the imputation of brackets altogether, because very few households answered the

unfolding bracket sequence, and thus there is not enough information to use this for imputation of the cases where we do not know the bracket.

2.2.1. Ownership Imputation

We define "ownership" of a given consumption or income item as a binary indicator in which the household (or individual) indicates whether it consumes the item, incurs the expenditures, receives the income, and so forth. In its simplest form, this is the answer to a yes/no question, but sometimes is the result of a sequence of such questions. We only impute ownership if it is asked separately from the amount. For many consumption items, the household is only asked how much they spent on them, without a separate question of whether the amount is nonzero. In such a case, we do not separately impute ownership.

We used RAND's imputation methods for the HRS as our starting point. For the HRS, a set of about 30 explanatory variables is constructed, which includes labor force participation and education of the husband and wife, whether they are in fair or poor health, the age and cognition score of the financial respondent, and various other characteristics. These are then analyzed by a principal components analysis (PCA), and the first 10 principal components are used as covariates in the imputation models. For ownership, the imputation model is a logistic regression (or binary logit, or simply logit) model. But if the number of households who report owning the item ($\text{ownership} = 1$) is less than 50, the covariates are not used, and the marginal probability of owning the item is used for imputing it.

Most consumption and income items in the LASI pilot have low ownership rates. Furthermore, the variables used in the PCA were selected for their predictive power for the types of assets and income in the HRS (e.g., financial assets and income from them), which may not be most suitable for predicting the consumption and income variables in LASI.

Therefore, we have taken the following approach. We have constructed a set of explanatory variables that aims to mimic the explanatory variables of the HRS, but that is adapted to the LASI context, and similarly computed their first 10 principal components. Table 5 lists the variables used. The 10 components have associated eigenvalues of 3.04 down to 1.13 and cumulatively account for only 55% of the variance. Thus, it suggests that the PCA is not able to summarize the 31 variables very well. For consumption and income items with very low rates of ownership (we did not use a formal cutoff, but typically less than about 20 "yes" responses), we imputed using only the marginal probability. For items with higher ownership rates, we considered various alternatives: logit regressions on the 10 principal components, possibly augmented or replaced by other covariates that seem more relevant to the specific context, splitting the sample by potentially relevant characteristics such as urbanicity and state, and combinations of these approaches.

The vast majority of necessary ownership imputations are caused by households not responding to the whole consumption module or to any financial module. This corresponds to the "no financial respondent" designator in the HRS and holds for 19 households for the consumption [CO] module and 21 households for the financial household modules (income [IN], agricultural assets and income [AG], and real estate and financial and non-financial assets

and debts [AD]). For individual income, there is no financial respondent (it is the respondent or the proxy respondent), but there are 10 respondents who did not answer any question in the work and employment [WE] and pension [PN] modules, and this is the majority of missings in these sections.

Table 2: Variables used in the principal components analysis

household gives financial help to family or friends
hh receives financial help from family or friends
any male with more education than a high school degree
any female with more education than a high school degree
any male who finished primary school but does not have more education than a high school degree
any female who finished primary school but does not have more education than a high school degree
any male with excellent or very good self-reported health
any female with excellent or very good self-reported health
any male with fair or poor self-reported health
any female with fair or poor self-reported health
any male agricultural worker
any female agricultural worker
any male non-agricultural worker
any female non-agricultural worker
any male who is unemployed
any female who is unemployed
any male who is disabled
any female who is disabled
any male who is retired
any female who is retired
number of males 0-17 years of age
number of males 18-44 years of age
number of males 45-64 years of age
number of males 65 years of age or older
number of females 0-17 years of age
number of females 18-44 years of age
number of females 45-64 years of age
number of females 65 years of age or older
financial respondent is a single female
financial respondent is married
financial respondent has a missing cognition score
financial respondent has a low cognition score

Household modules

For the consumption module, we imputed whether each of a set of food items was (partially) home grown or received in-kind. For these, we split the sample by state and urbanicity and used the marginal probabilities in each cell. The other ownership variable imputed is whether some of the consumption reflects business expenses, for which we used the overall marginal probability.

The household income section asks about earnings, self-employment earnings, and pension income of all household members separately. We converted this to an individual-level data set and, for each of these three income components, estimated a logit model with the following covariates: state, urbanicity, gender, whether married, an interaction between gender and being married, a cubic polynomial in age, and dummies for whether the household member was the first or second in the household listing. We imputed ownership from these models.

The income, agricultural income and assets, and real estate and non-financial and financial assets and debts sections have a large number of household-level income components. Most of these have very low levels of ownership and we therefore used only the marginal probability. This holds for joint non-agricultural self-employment activities; eight kinds of government transfers; gifts (other than remittances), donations, or inheritances from others or charitable organizations; other household income; rental income from agricultural land; rental income from nine kinds of agricultural machines or equipment; income from renting out real estate or part of their own residence; rental income from seven kinds of means of transportation; and interest income from loans to others.

A number of household income components are more common, and for these we split the sample by state and urbanicity and imputed using the marginal probability in each cell. This holds for having a ration or BPL card; receiving remittances; engaging in crop growing, forestry, or fishing; renting out livestock; and selling livestock products.

Finally, there is a single question asking for the total amount of interest and dividends received from financial assets, with a sequence of questions preceding this determining ownership. Ownership of financial asset income is the only variable studied here in which state and urbanicity are not highly predictive but the principal components do have predictive power. So for this variable, our imputations are based on a logit model with the 10 principal components as covariates.

Individual modules

LASI uses adaptive survey routing to field certain question only to the relevant or eligible respondents. For example, the question "did you ever work for pay" was only asked if the individual answered no to the questions about doing agricultural work and doing non-agricultural work. In our imputations, we implement the same logic and thus, for this item, we do not impute this question if after imputation of the first two questions, one of them is a yes. We use similar logic for other questions in the work and employment and pension sections as well.

For agricultural work, we split the sample by rural or urban residence. For rural residents, we impute from a logit model with gender and age group as covariates. After studying patterns across age, we concluded that the following age categorization reflects these patterns well: <65, 65-69, 70-74, 75+. For urban residents, we imputed using the marginal probability.

For nonagricultural work, we split the sample by urbanicity and agricultural work, and for individuals in rural areas who do not engage in agricultural work, we further split by gender. In each of the resulting cells, we imputed using the marginal probability.

There are three types of nonagricultural work: employment, self-employment, and unpaid family work. For the main work activity, these are mutually exclusive. Conditional on engaging in nonagricultural work, we split the sample by urbanicity and agricultural work and imputed the type from the marginal probabilities within the cells.

Wage and salary workers are asked about 10 kinds of fringe benefits. We imputed these based on the (very low) marginal probabilities, conditional on being a wage and salary worker.

All individuals who report doing nonagricultural work were asked whether they have any side jobs. We imputed this based on the marginal probability, conditional on engaging in nonagricultural work.

Individuals who do not currently engage in work were asked whether they ever worked for pay. We split the sample by urbanicity and for rural residents by gender, and imputed using the probabilities within the cells.

Receiving official or unofficial pensions is conditioned on not currently doing nonagricultural work and ever having worked for pay. For this subgroup, we split the sample by agricultural work for official pension and imputed using marginal probabilities in each cell. For unofficial pensions we used the overall marginal probability in the mentioned subgroup. For receiving a commercial pension, we imputed using the unconditional marginal probability. For survivor pension, we split by gender and imputed for males and females separately using their marginal probabilities. For receiving a government pension, we imputed from a logit model with gender and age group (with the same categories as for agricultural work) as covariates.

2.2.2. Amount Imputation

Where respondents answered “don’t know” or “refused to answer” to specific income amounts, they are channeled into unfolding brackets. Respondents are given up to a series of three choices within these brackets. When the final value from the bracket questions is reported to be one single suggested bracket value (e.g. 2000 rupees), this value is used as the income amount. When the final value from the bracket questions is reported to be within a range (e.g. higher than 2000 rupees and lower than 5000 rupees), then the final amount is replaced with the median value of amounts reported within this range derived from all other non-missing values.

However, brackets were rarely used in LASI and thus the number of these median imputations is small. Thus, for most households, we only know whether or not they owned

the item, or if even this is not reported, we have imputations of whether or not the household owns the item. When an item is owned (reported or imputed) and no bracket information is available, we used a conditional hotdeck procedure to impute the value. As mentioned above, we did not impute the bracket in an intermediate step, unlike the RAND HRS. The SAS “hotdeckvar” procedure was used for imputing the missing values. This is a procedure creating imputed variables through single hotdeck imputations. The missing values are imputed from randomly selected similar records in the database that share attributes related to the incomplete variables. The algorithm identifies all donor observations that have no missing values for specified variable. Missing values from the same observation are replaced with values randomly selected from a same donor observation to preserve correlations.

2.3. Derived Variables

2.3.1. Household income Measures

Summary measures of income are aggregated to the household level. A household respondent answers all questions on income for each member of the household. For each household member, individual earnings from a wage/salary job, profits from self-employment, agricultural income, and pension income received in the last 12 months are asked. These individual-specific amounts are summed across all household members and are then combined with all other income sources, such as remittance and other family transfers, government subsidies, and asset income. For households with agricultural and business income, LASI identifies which household members engage in those activities and avoids double-counting.

2.3.2. Household Consumption Measures

Aggregate measures of consumptions are created at the household level. The Household Consumption [CO] module in LASI asks one respondent from each household about expenses incurred. For food items and other frequently purchased items, such as utilities, fuels, transportation, and etc., respondents are asked to report expenditure during 30 day time period with an option to report expenditure per day or week. For food items, respondents are also asked about in-kind transfers and home-production in the last 30 days. For less frequently purchased items, such as health expenditures, education, etc., respondents are asked to report expenditure during the past 12 months with an option to report the expenditure during the past week and month. From these questions, we create summary yearly measures for food and non-food expenditures, and sub-categories of items are summarized below. Unfolding brackets are not used for expenditure questions.

2.3.3. Individual Earnings and Pension Measures

We created individual earnings measured based on self-reported information collected from the work [WE] and pension [PN] modules. All questions regarding actual earnings and pension receipts include unfolding brackets to more precisely capture missing information.

Median bracket value replacement is used where bracket bounds are known. Otherwise, remaining missing data are imputed according the “hotdeckvar” method described above.

missing values .d, .r, .m respectively. RWSHLT is set to special missing (.p) if the health status question was skipped because the interview was by proxy. RWSHLT is set to plain missing (.) for respondents who did not respond to the current wave.

SWSHLT is the respondent's spouse's self-reported general health status taken directly from spouses' values of RWSHLT. In addition to the special missing codes used in RWSHLT, SWSHLT employs two other missing codes, .u and .v. Special missing value .u is used when the respondent does not report being coupled in the current wave. Special missing value .v is used when the respondent reports being coupled in the current wave but their spouse is not interviewed.

8 → Cross-Wave Differences in LASI

The first scale of self-reported general health status is used in the pilot wave of the LASI.

9 → Differences with the RAND HRS

Unlike the HRS, the LASI varied when the respondent was asked RWSHLT inside the health module during Wave 1.

10 → LASI Variables Used

Wave 0 Core:	
HEHELF	Would you say your health is ... ? {start of section}
HEHELFB	How is your health in general? Would you say it was ...? {end of section}
HEHELF	self-reported general health

- 1 *Title:* The variables are documented in groups according to the concept that they measure. For example, there are eight variables related to self-reported health, corresponding to four waves and respondent/spouse. The title is often followed by a short description of the concept that is captured.
- 2 *Variable Names:* This entry shows the waves of variables in the group.
- 3 *Variable Labels:* This entry shows the Stata variable labels. As discussed above, the labels typically include the name of the variable, the file on which it is present, and a description of its contents.
- 4 *Variable Type:* This entry indicates the type of variable. It may be continuous (Cont), categorical (Categ), or character (Char).
- 5 *Descriptive Statistics:* This entry shows descriptive statistics on each variable. They include the number of nonmissing values, the mean, standard deviation, minimum value, and maximum value.
- 6 *Categorical Value Codes:* This entry shows the value label codes. These are only relevant for categorical variables. The first character(s) of the value labels indicate the value to which each label has been assigned. For example, value "1" is mapped into "1. Excellent" (not just "Excellent"). The entry also indicates which labels are

assigned to which variables, and shows frequency tabulations for all categorical variables.

- 7 *How Constructed:* This entry provides background on the manner in which variables were constructed.
- 8 *Cross-Wave Differences in LASI:* This entry briefly describes differences in question wording or contents between interview waves.
- 9 *Differences with the RAND RHS:* This entry describes any differences between the RAND HRS version of the variable and the Harmonized LASI version of the variable. It is imperative these differences are understood when using harmonized measures.
- 10 *LASI Variables Used:* This entry provides the names and labels of raw LASI variables that were used to construct the new variables.

4. Data Download

The LASI team released the data through the Gateway to Global Aging Data (G2G). G2G is a web-based host of harmonized longitudinal studies including the several other HRS sister surveys. Users must create a log-in name and password to download the data from

<https://www.g2aging.org/>

Once logged in, a user can navigate to the LASI home page under the “Browse Studies” tab shown in Figure 1.

Figure 1. Data Download

The screenshot shows the G2G website interface. At the top, there's a navigation bar with tabs: Introduction, Surveys, Topics, Data, Obtain, Create, Read, and Help. The 'Create' tab is selected. Below the navigation bar, there's a section for 'Longitudinal Aging Study in India'. A table lists the survey details:

survey name	description	data status	options
LASI pilot	Pilot Study for the Longitudinal Aging Study in India		

Below the table, there's a button labeled 'LASI micro data is now available for download.' A red arrow points to this button.

Clicking on the “download” link will navigate to a new page that users can download the Harmonized LASI data and this codebook.

There are several other files on the page:

- LASI Pilot instrument (PDF): This is the survey questionnaire. This contains the original survey questions in English and associated variables' names. This document helps users to match the questions to variables in the data set. It also allow users to observe survey skip patterns in the questionnaire.
- LASI pilot household macro data: This is the household level data in STATA format.
- LASI pilot individual macro data: This is the individual level data in STATA format.
- LASI Pilot User guide (PDF): This contains the detail information about data collection, sampling design, data structure, and weight/sampling variables.
- LASI Pilot Fat file: This is a user friendly merged version of the household and individual dataset. The file contains all the variables from household and individual level files. Users can use this file instead of downloading two macro data (household and individual level files) and then merge them.
- LASI Pilot Fat file codebook (PDF): This is a corresponding codebook for the fat file that lists variables names, labels, and summary statistics.
- Harmonized LASI data: The data contains the derived variables.
- Harmonized LASI codebook (PDF): This is a corresponding codebook for the Harmonized data that lists variables names, labels, and summary statistics.

5. Data Codebook

Section A: Demographics and Identifiers

Person Specific Identifier

Wave	Variable	Label	Type
1	HHID	hhid: hhold id / num	Cont
1	PN	pn: person id (last digit prim_key)/ num	Cont
1	HHIDPN	hhidpn: unique person identifier (prim_key) / num	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
HHID	1683	1.4E14	2.2E13	1.1E14	1.8E14
PN	1683	1.65	0.86	1.00	7.00
HHIDPN	1683	1.4E14	2.2E13	1.1E14	1.8E14

How Constructed:

HHID is the numeric version of the household identifier.

PN is the 1-digit person number. Together HHID and PN uniquely identify each individual.

HHIDPN is a numeric version of unique identifier for each respondent, which combining HHID and PN. HHIDPN is taken from "prim_key" from LASI raw data.

LASI Variables Used:

PRIM_KEY	PRIMARY KEY
----------	-------------

Spouse Identifier

Wave	Variable	Label	Type
1	S0HHIDPN	s0hhidpn: w0 spouse id (hhid+spousecvid) / num	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
S0HHIDPN	1683	1.0E14	6.8E13	0.00	1.8E14

How Constructed:

S0HHIDPN is the identifier of respondent's spouse. It was created by HHID and spousecvid.

If the spouse is not interviewed, S0HHIDPN is set to zero.

LASI Variables Used:

HHID	HHID: HH IDENTIFIER, NUMERIC
PRIM_KEY	PRIMARY KEY
SPOUSECVID_R	SPOUSE IDENTIFIER

Sampling State

Wave	Variable	Label	Type
1	R0STATE	r0state: w0 sampling state	Categ
1	S0STATE	s0state: w0 sampling state	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0STATE	1683	2.51	1.10	1.00	4.00
S0STATE	1208	2.47	1.12	1.00	4.00

Categorical Variable Codes

Value-----	R0STATE
1.punjab	402
2.rajasthan	417
3.kerala	462
4.karnataka	402

Value-----	S0STATE
.u=unmar	321
.v=sp nr	154
1.punjab	312
2.rajasthan	308
3.kerala	298
4.karnataka	290

How Constructed:

R0STATE is the sampling state variable.

The LASI survey was conducted in four Indian states - Rajasthan and Punjab in the north and Kerala and Karnataka in the south – and fielded in the dominant language of each state. These four states were chosen to capture the demographic, economic, health, and cultural diversity of India.

The S0STATE variable is taken from the spouse's R0STATE variable.

LASI Variables Used:

STATE	STATE
-------	-------

Analysis Weight

Wave	Variable	Label	Type
1	H0PWTHH	h0pwthh: w0 pooled household weight	Cont
1	H0SWTHH	h0swthh: w0 state household weight	Cont
1	R0PWTRISP	r0pwtrisp: w0 pooled person-level weight	Cont
1	R0SWTRISP	r0swtrisp: w0 state person-level weight	Cont
1	S0PWTRISP	s0pwtrisp: w0 pooled person-level weight	Cont
1	S0SWTRISP	s0swtrisp: w0 state person-level weight	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0PWTHH	1683	1.00	0.34	0.52	1.44
H0SWTHH	1683	1.00	0.07	0.81	1.09
R0PWTRISP	1683	1.00	0.35	0.52	1.44
R0SWTRISP	1683	1.00	0.06	0.82	1.07
S0PWTRISP	1208	1.00	0.35	0.52	1.44
S0SWTRISP	1208	1.00	0.06	0.82	1.07

How Constructed:

H0PWTHH and H0SWTHH are the household level analysis weights. H0PWTHH is the pooled household weight and is used to create a sample representative of the population across the four LASI sates. H0SWTHH is the state level analysis weight that is used to examine households within a given state.

R0PWTRISP and R0SWTRISP are the individual level analysis weights. R0PWTRISP is the pooled individual level weight and is used to create a sample representative of the population across the four LASI sates. R0SWTRISP is the state analysis weight that is used to examine households within a given state.

These weights are created using the inverse probability of selection combined with household and individual response rates. To calculate the weights, we used forecast estimates of the 2010 population based on the 2001 census round. Weights are based on rural and urban population counts of the population 45 years and older within each state.

The S0PWTRISP and S0SWTRISP are taken from spouse's R0PWTRISP and R0SWTRISP, respectively.

LASI Variables Used:

HH_WT_POOLED	POOLED HOUSEHOLD WEIGHT
INDI_WT_POOLED	POOLED INDIVIDUAL WEIGHT
STATE_HH_WT	STATE HOUSEHOLD WEIGHT
STATE_INDI_WT	STATE INDIVIDUAL WEIGHT

Number of Household Respondents
--

Wave	Variable	Label	Type
1	H0HHRESP	h0hhresp: w0 # respondents in household	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0HHRESP	1683	2.00	0.69	1.00	6.00

How Constructed:

H0HHRESP is the number of individuals in the house who actually responded at each wave. It counts the number of respondents sharing the same household id.

LASI Variables Used:

HHID	HHID: HH IDENTIFIER, NUMERIC
PRIM_KEY	PRIMARY KEY
RCVID_R	RESPONDENT IDENTIFIER

Whether Couple or not

Wave	Variable	Label	Type
1	R0CPL	r0cpl: w0 whether coupled	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0CPL	1683	0.72	0.45	0.00	1.00

Categorical Variable Codes

Value	R0CPL
0.no	475
1.yes	1208

How Constructed:

R0CPL indicates whether this respondent is a part of a coupled or not. Household in LASI can consist of more than one couple or single persons.

Please note this variable specifically refers to the RAND-HRS definition of a household consisting of one couple or one single person.

R0CPL is set to one if the respondent is coupled with another respondent. R0CPL is set to zero if the respondent is not coupled with any other respondent in the same household.

LASI Variables Used:

PRIM_KEY	PRIMARY KEY
RCVID_R	RESPONDENT IDENTIFIER
SPOUSECVID_R	SPOUSE IDENTIFIER

Financial Respondent, Family Respondent, and Consumption Respondent
--

Wave	Variable	Label	Type
1	R0FINR	r0finr: w0 whether financial resp	Categ
1	S0FINR	s0finr: w0 whether financial resp	Categ
1	R0FAMR	r0famr: w0 whether family resp	Categ
1	S0FAMR	s0famr: w0 whether family resp	Categ
1	R0CONSR	r0consr: w0 whether consumption resp	Categ
1	S0CONSR	s0consr: w0 whether consumption resp	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0FINR	1683	0.52	0.50	0.00	1.00
S0FINR	1208	0.47	0.50	0.00	1.00
R0FAMR	1683	0.52	0.50	0.00	1.00
S0FAMR	1208	0.47	0.50	0.00	1.00
R0CONSR	1683	0.52	0.50	0.00	1.00
S0CONSR	1208	0.47	0.50	0.00	1.00

Categorical Variable Codes

Value-----	R0FINR
0.no	801
1.yes	882
Value-----	S0FINR
.u=unmar	321
.v=sp nr	154
0.no	640
1.yes	568
Value-----	R0FAMR
0.no	806
1.yes	877
Value-----	S0FAMR
.u=unmar	321
.v=sp nr	154
0.no	639
1.yes	569
Value-----	R0CONSR
0.no	801
1.yes	882
Value-----	S0CONSR
.u=unmar	321
.v=sp nr	154
0.no	641
1.yes	567

How Constructed:

R0FINR indicates whether the respondent answered the financial questions, that is, the household income and asset questions.

R0FAMR indicates whether the respondent answered the household demographic questions for every individual in the household. These questions included basic demographic information about everyone living the household.

R0CONSR indicates whether the respondent answered the consumption questions.

The S0FINR, S0FAMR and S0CONSR are taken from the spouse's R0FINR, R0FAMR and R0CONSR variables.

LASI Variables Used:

CONSUMPTIONR_R	CONSUMPTION R
FAMILYR_R	FAMILYR
FINANCIALR_R	FINANCIAL R

Interview Dates

Wave	Variable	Label	Type
1	R0IWMONTH	r0iwmmonth: w0 interview month	Cont
1	R0IWYEAR	r0iwyyear: interview year	Cont
1	S0IWMONTH	s0iwmmonth: w0 interview month	Cont
1	S0IWYEAR	s0iwyyear: w0 interview year	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0IWMONTH	1668	11.36	0.52	10.00	12.00
R0IWYEAR	1683	2010.00	0.00	2010.00	2010.00
S0IWMONTH	1202	11.39	0.53	10.00	12.00
S0IWYEAR	1208	2010.00	0.00	2010.00	2010.00

How Constructed:

R0IWMONTH and R0IWYEAR indicate the interview month and year.

S0IWMONTH and S0IWYEAR are taken from spouse's R0IWMONTH and S0IWYEAR variables.

LASI Variables Used:

TSEND_R	TIMESTAMP END
---------	---------------

Birth date: Month and Year

Wave	Variable	Label	Type
1	RABMONTH	rabmonth: birth month	Cont
1	RABYEAR	rabyear: r birth year	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
RABMONTH	1150	5.51	3.15	1.00	12.00
RABYEAR	1681	1954.47	12.27	1907.00	1995.00

How Constructed:

RABYEAR and RABYEAR are the respondent's reported birth year and month, respectively.

There are about 23% of respondents did not answer the birth year question. We use the self-reported age to calculate the respondent's birth year if birth year is missing.

LASI Variables Used:

DM007_MONTH	DATE OF BIRTH MONTH
DM007_YEAR	DATE OF BIRTH YEAR
DM008	AGE IN COMPLETED YEARS

Age at interview (in months and years)

Wave	Variable	Label	Type
1	R0AGEY	r0agey: w0 self-reported age (years)	Cont
1	S0AGEY	s0agey: w0 self-reported age (years)	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0AGEY	1681	55.59	12.27	15.00	103.00
S0AGEY	1207	54.18	11.18	15.00	96.00

How Constructed:

R0AGEY is the respondent's age in years at the time of the current wave's interview. Respondent's age is calculated by respondent's birth year and interview date.

There are about 23% of respondents did not answer the birth year question. We use the self-reported age to calculate the respondent's birth year if birth year is missing.

S0AGEY is the spouse age and is taken from spouse's R0AGEY variable.

LASI Variables Used:

DM007_MONTH	DATE OF BIRTH MONTH
DM007_YEAR	DATE OF BIRTH YEAR
DM008	AGE IN COMPLETED YEARS

Gender

Wave	Variable	Label	Type
1	RAGENDER	ragender: gender	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
RAGENDER	1683	1.56	0.50	1.00	2.00

Categorical Variable Codes

Value	RAGENDER
1.male	734
2.female	949

How Constructed:

RAGENDER is respondent's gender. RAGENDER is set to 1 for male and 2 for female.

LASI Variables Used:

DM002	GENDER OF RESPONDENT
-------	----------------------

Education: Years of Education

Wave	Variable	Label	Type
1	RAEDYRS	raedyrs: years of education	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
RAEDYRS	1681	4.50	4.89	0.00	22.00

How Constructed:

RAEDYRS is the years of education variable.

The value of RAEDYRS ranges from 0 to 22. If respondent never attended school, RAEDYRS is set to 0.

LASI Variables Used:

DM029	ATTENDED SCHOOL
DM030	HOW MANY YEARS SCHOOLING

Education: Highest level of education
--

Wave	Variable	Label	Type
1	RAEDUC	raeduc: highest level of education	Categ
1	RAEDUC_L	raeduc_l: lasi highest level of education categories	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
RAEDUC	1682	1.31	0.74	1.00	4.00
RAEDUC_L	1682	1.56	1.73	0.00	5.00

Categorical Variable Codes

Value-----	RAEDUC
.m=oth missing	1
1.lt hs	1356
2.hs grad	198
3.some college	55
4.college and above	73

Value-----	RAEDUC_L
.m=oth missing	1
0.no shcool	770
1.lt primary school	160
2.primary school completed	243
3.secondary school completed	183
4.hs grad	198
5.some college or above	128

How Constructed:

RAEDUC is defined using the RAND HRS categorical summary of education: Less than high school, high school graduate, some college, and college and above. LASI survey respondents as to their highest educational qualification.

RAEDUC_L is another education categorical variable that shows details for lower education levels.

LASI Variables Used:

DM029	ATTENDED SCHOOL
DM031	HIGHEST LEVEL OF SCHOOLING COMPLETED

Current Marital Status

Wave	Variable	Label	Type
1	ROMSTAT	r0mstat: w0 marital status	Categ
1	SOMSTAT	s0mstat: w0 marital status	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
ROMSTAT	1683	2.13	2.34	1.00	8.00
SOMSTAT	1208	1.00	0.00	1.00	1.00

Categorical Variable Codes

Value	ROMSTAT
1.currently married	1362
4.separated	11
5.divorced	8
7.widowed	281
8.never married	21

Value	SOMSTAT
.u=unmar	321
.v=sp nr	154
1.currently married	1208

How Constructed:

ROMSTAT indicates a respondent's marital status. A code of 1 indicates the respondent is married. A code of 4 indicates the respondent is separated. A code of 5 indicates the respondent is divorced. A code of 7 indicates the respondent is widowed. A code of 8 indicates the respondent has never been married.

SOMSTAT is taken from spouse's ROMSTAT variable.

LASI Variables Used:

RMARITALSTATUS R MARITAL STATUS

Place of birth

Wave	Variable	Label	Type
1	RABPLACE	rabplace: birth place	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
RABPLACE	1676	17.70	4.87	1.00	28.00

Categorical Variable Codes

Value-----	RABPLACE
.d=dk	1
.m=missing	6
1.andhra pradesh	11
2.arunachal pradesh	3
3.assam	4
4.bihar	3
6.delhi	1
8.gujarat	1
9.haryana	2
10.himachal pradeh	4
12.jharkhand	2
13.karnataka	381
14.kerala	458
16.maharashtra	1
22.punjab	387
23.rajasthan	413
25.tamil nadu	3
28.uttar pradesh	2

How Constructed:

RABPLACE indicates the respondent's birthplace.

LASI Variables Used:

DM022_STATE	PLACE OF BIRTH STATE TERRITORY
-------------	--------------------------------

Urban or Rural Residency

Wave	Variable	Label	Type
1	R0LVREG	r0lvreg: w0 urban or rural residency status	Categ
1	S0LVREG	s0lvreg: w0 urban or rural residency status	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0LVREG	1683	1.72	0.45	1.00	2.00
S0LVREG	1208	1.74	0.44	1.00	2.00

Categorical Variable Codes

Value-----	R0LVREG
1.urban	472
2.rural	1211
Value-----	S0LVREG
.u=unmar	321
.v=sp nr	154
1.urban	314
2.rural	894

How Constructed:

R0LVREG indicates whether respondents live in urban or rural area.

S0LVREG is taken from spouse's R0LVREG variable.

LASI Variables Used:

RESIDENCE	RESIDENCE
-----------	-----------

Caste in India

Wave	Variable	Label	Type
1	RACASTE	racaste: w0 caste in india	Categ
1	S0CASTE	s0caste: w0 caste in india	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
RACASTE	1635	2.91	1.06	1.00	4.00
S0CASTE	1175	2.91	1.04	1.00	4.00

Categorical Variable Codes

Value	RACASTE
.m=oth missing	48
1.scheduled caste	272
2.scheduled tribe	179
3.other backward classs obc	609
4.none of them	575
Value	S0CASTE
.m=oth missing	33
.u=unmar	321
.v=sp nr	154
1.scheduled caste	186
2.scheduled tribe	136
3.other backward classs obc	450
4.none of them	403

How Constructed:

R0CASTE indicates whether respondent is a caste or tribe.

S0CASTE is taken from spouse's R0CASTE variable.

LASI Variables Used:

DM003	CURRENT MARITAL STATUS
DM034	CAST TRIBE KIND
DM034_OTHER	CAST TRIBE KIND OTHER

Section B: Health

Self-report of health

Wave	Variable	Label	Type
1	R0SHLT	r0shlt: w0 self-report of health	Categ
1	S0SHLT	s0shlt: w0 self-report of health	Categ
1	R0SHLTA	r0shlta: w0 self-report of health, european scale	Categ
1	S0SHLTA	s0shlta: w0 self-report of health, european scale	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0SHLT	1679	2.68	0.81	1.00	5.00
S0SHLT	1206	2.59	0.76	1.00	5.00
R0SHLTA	1677	2.48	0.75	1.00	5.00
S0SHLTA	1206	2.41	0.70	1.00	5.00

Categorical Variable Codes

Value-----	R0SHLT
.m=missing	4
1.excellent	62
2.very good	670
3.good	725
4.fair	180
5.poor	42

Value-----	S0SHLT
.m=missing	2
.u=unmar	321
.v=sp nr	154
1.excellent	50
2.very good	526
3.good	514
4.fair	97
5.poor	19

Value-----	R0SHLTA
.m=missing	6
1.very good	53
2.good	935
3.fair	551
4.poor	102
5.very poor	36

Value-----	S0SHLTA
.m=missing	2
.u=unmar	321
.v=sp nr	154
1.very good	40
2.good	726
3.fair	368
4.poor	54
5.very poor	18

How Constructed:

R0SHLT is the respondent's self-reported general health status using a scale ranging from Excellent to Poor. Codes range from 1 to 5. Missings to R0SHLT are assigned .m.

LASI also employs a second scale of self-reported general health status. R0SHLTA is the respondent's self-reported general health status using a scale ranging from Very Good to Very Bad. Codes range from 1 to 5.

S0SHLT and S0SHLTA are taken from the spouse's R0SHLT and R0SHLTA variables.

LASI Variables Used:

HT_RANDOM_HEALTH_A	RATING OF HEALTH
HT_RANDOM_HEALTH_B	RATING OF HEALTH European scale

Whether health limits work

Wave	Variable	Label	Type
1	R0HLTHLM	r0hlthlm: w0 whether health limit work	Categ
1	S0HLTHLM	s0hlthlm: w0 whether health limit work	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0HLTHLM	1677	0.40	0.49	0.00	1.00
S0HLTHLM	1206	0.37	0.48	0.00	1.00

Categorical Variable Codes

Value	R0HLTHLM
.m=oth missing	5
.r=rf	1
0.no	1003
1.yes	674

Value	S0HLTHLM
.m=oth missing	2
.u=unmar	321
.v=sp nr	154
0.no	757
1.yes	449

How Constructed:

R0HLTHLM indicates whether an impairment or health problem limits the kind or amount of paid work for the respondent. A code of 0 indicates that the respondent reports their work is not limited by a health problem.

A code of 1 indicates that the respondent reports their work is limited by a health problem. Don't know, not applicable, or refused values of R0HLTHLM are assigned special missing codes .d, .n, .r, respectively.

S0HLTHLM is taken from spouse's R0HLTHLM variable.

LASI Variables Used:

HT090	DOES HEALTH LIMIT ABILITY TO WORK
-------	-----------------------------------

Activities of daily living (ADLs): Raw recodes

Wave	Variable	Label	Type
1	R0WALKRA	r0walkra: w0 diff-walk across room	Categ
1	S0WALKRA	s0walkra: w0 diff-walk across room	Categ
1	R0DRESSA	r0dressa: w0 diff-dressing	Categ
1	S0DRESSA	s0dressa: w0 diff-dressing	Categ
1	R0BATHA	r0batha: w0 diff-bathing or showerng	Categ
1	S0BATHA	s0batha: w0 diff-bathing or showing	Categ
1	R0EATA	r0eat: w0 diff-eating	Categ
1	S0EATA	s0eata: w0 diff-eating	Categ
1	R0BEDA	r0beda: w0 diff-get in/out of bed	Categ
1	S0BEDA	s0beda: w0 diff-get in/out of bed	Categ
1	R0TOILTA	r0toilta: w0 diff-using the toilt	Categ
1	S0TOILTA	s0toilta: w0 diff-using the toilt	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0WALKRA	1673	0.06	0.23	0.00	1.00
S0WALKRA	1204	0.04	0.20	0.00	1.00
R0DRESSA	1675	0.04	0.20	0.00	1.00
S0DRESSA	1205	0.03	0.18	0.00	1.00
R0BATHA	1675	0.04	0.19	0.00	1.00
S0BATHA	1204	0.03	0.16	0.00	1.00
R0EATA	1675	0.04	0.19	0.00	1.00
S0EATA	1204	0.02	0.16	0.00	1.00
R0BEDA	1667	0.06	0.24	0.00	1.00
S0BEDA	1200	0.05	0.21	0.00	1.00
R0TOILTA	1675	0.05	0.22	0.00	1.00
S0TOILTA	1204	0.04	0.20	0.00	1.00

Categorical Variable Codes

Value-----	R0WALKRA
.m=oth missing	7
.x=dont do	3

0.no		1578
1.yes		95
Value-----		SOWALKRA
.m=oth missing		3
.u=unmar		321
.v=sp nr		154
.x=dont do		1
0.no		1156
1.yes		48
Value-----		R0DRESSA
.m=oth missing		7
.x=dont do		1
0.no		1605
1.yes		70
Value-----		S0DRESSA
.m=oth missing		3
.u=unmar		321
.v=sp nr		154
0.no		1164
1.yes		41
Value-----		R0BATHA
.d=dk		1
.m=oth missing		7
0.no		1611
1.yes		64
Value-----		S0BATHA
.d=dk		1
.m=oth missing		3
.u=unmar		321
.v=sp nr		154
0.no		1172
1.yes		32
Value-----		R0EATA
.m=oth missing		7
.x=dont do		1
0.no		1609
1.yes		66
Value-----		S0EATA
.m=oth missing		3
.u=unmar		321
.v=sp nr		154
.x=dont do		1
0.no		1174
1.yes		30
Value-----		R0BEDA
.m=oth missing		15
.r=rff		1
0.no		1563
1.yes		104
Value-----		S0BEDA
.m=oth missing		8
.u=unmar		321
.v=sp nr		154
0.no		1145
1.yes		55
Value-----		R0TOILTA
.m=oth missing		7
.x=dont do		1
0.no		1588
1.yes		87
Value-----		S0TOILTA
.m=oth missing		3
.u=unmar		321
.v=sp nr		154
.x=dont do		1

0.no		1154
1.yes		50

How Constructed:

ROWALKRA, RODRESSA, ROBATHA, ROEATA, ROBEDA and R0TOILTA are indicators that measure the respondents' difficulty with a select number of activities of daily life (ADLs): walking across room, dressing one's self, bathing one's self, eating, getting out of bed, and using a toilet.

These six ADL variables are derived based on 4 different answers:

1. Yes, have difficulty with..
2. No
3. Can't do
4. Don't want to

If the respondent answers "Yes" or "Can't do" to any difficulty question, the variables are set to 1 for some difficulty. If the respondent answers "no" to the any difficulty question, the variables are set to zero. A "don't do" response is recoded to missing values .X, since the respondent hasn't revealed whether he/she would have difficulty with the activity.

LASI Variables Used:

HT401	DIFFICULTY WITH DRESSING
HT402	DIFFICULTY WALKING
HT403	DIFFICULTY BATHING OR SHOWERING
HT404	DIFFICULTY EATING
HT405	DIFFICULTY GETTING IN OR OUT OF BED
HT406	DIFFICULTY USING TOILET

ADL Summary: sum ADLs where respondent reports any difficulty

Wave	Variable	Label	Type
1	ROADLA	r0adla: w0 some diff-adls / 0-5	Cont
1	S0ADLA	s0adla: w0 some diff-adls/ 0-5	Categ
1	ROADLWA	r0adlwa: w0 some diff-adls: wallace / 0-3	Cont
1	S0ADLWA	s0adlwa: w0 some diff-adls: wallace/ 0-3	Categ
1	ROADLAM	r0adlam: w0 missing of adls summary	Cont
1	S0ADLAM	s0adlam: w0 missing of adls summary/ 0-3	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
ROADLA	1662	0.23	0.80	0.00	5.00
S0ADLA	1197	0.17	0.67	0.00	5.00
ROADLWA	1673	0.12	0.47	0.00	3.00
S0ADLWA	1203	0.09	0.40	0.00	3.00
ROADLAM	1683	0.03	0.34	0.00	5.00
S0ADLAM	1208	0.02	0.26	0.00	5.00

How Constructed:

Two activities of Daily Living (ADL) summaries are derived. One uses the ADLs proposed by Wallace and Herzog in the paper Wallace and Herzog, 1995 to define an ADL summary: R_wADLWA: bathe, dress, and eat. The second includes these and adds getting in/out of bed and walking across a room: R_wADLA. In all waves the "some difficulty" versions of the individual measures are used to construct these measures, i.e., R_wWALKRA, R_wBEDA, R_wBATHA, R_wDRESSA, and R_wEATA variables are used. Each limitation adds one to the summary measure, that is:

$$R_{wADLWA} = \text{sum} (R_{wBATHA}, R_{wDRESSA}, R_{wEATA})$$

$$R_{wADLA} = \text{sum} (R_{wBATHA}, R_{wDRESSA}, R_{wEATA}, R_{wBEDA}, R_{wWALKRA})$$

R_wADLWA and R_wADLA are not computed for respondents with missing values of R_wBATHA, R_wDRESSA, R_wEATA, R_wBEDA, and/or R_wWALKRA.

R_wADLAM indicates the number of missing in summary measure.

Mental health (CESD score)

Wave	Variable	Label	Type
1	R0DEPRES	r0depresd: w0 cesd: felt depressed	Categ
1	S0DEPRES	s0depresd: w0 cesd: felt depressed	Categ
1	R0EFFORTD	r0effortd: w0 cesd: everything an effort	Categ
1	S0EFFORTD	s0effortd: w0 cesd: w0 everything an effort	Categ
1	R0SLEEP	r0sleeprd: w0 cesd: sleep was restless	Categ
1	S0SLEEP	s0sleeprd: w0 cesd: sleep was restless	Categ
1	R0WHAPPYD	r0whappyd: w0 cesd: was happy	Categ
1	S0WHAPPYD	s0whappyd: w0 cesd: was happy	Categ
1	R0FLONED	r0floned: w0 cesd: felt lonely	Categ
1	S0FLONED	s0floned: w0 cesd: felt lonely	Categ
1	R0FSADD	r0fsadd: w0 cesd: felt sad	Categ
1	S0FSADD	s0fsadd: w0 cesd: felt sad	Categ
1	R0GOINGD	r0goingd: w0 cesd: could not get going	Categ
1	S0GOINGD	s0goingd: w0 cesd: could not get going	Categ
1	R0ENLIFED	r0enlifed: w0 cesd: enjoy life	Categ
1	S0ENLIFED	s0enlifed: w0 cesd: enjoy life	Categ
1	R0CESDD	r0cesdd: w0 cesd score	Cont
1	S0CESDD	s0depresd: w0 cesd score	Cont
1	R0CESDMD	r0cesdmd: w0 number of missing in r0cesdd	Cont
1	S0CESDMD	s0cesdmd: w0 number of missing in r0cesdd	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0DEPRES	1587	0.46	0.74	0.00	3.00
S0DEPRES	1152	0.42	0.71	0.00	3.00
R0EFFORTD	1562	0.43	0.70	0.00	3.00
S0EFFORTD	1132	0.42	0.71	0.00	3.00
R0SLEEP	1590	0.63	0.85	0.00	3.00
S0SLEEP	1155	0.59	0.83	0.00	3.00
R0WHAPPYD	1585	0.95	1.04	0.00	3.00

S0WHAPPYD	1150	1.00	1.06	0.00	3.00
R0FLONED	1578	0.42	0.73	0.00	3.00
S0FLONED	1147	0.36	0.68	0.00	3.00
R0FSADD	1576	0.52	0.79	0.00	3.00
S0FSADD	1143	0.47	0.76	0.00	3.00
R0GOINGD	1517	0.32	0.66	0.00	3.00
S0GOINGD	1100	0.30	0.65	0.00	3.00
R0ENLIFED	1564	0.40	0.72	0.00	3.00
S0ENLIFED	1134	0.35	0.69	0.00	3.00
R0CESDD	1453	7.31	2.51	0.00	20.00
S0CESDD	1054	7.08	2.45	2.00	20.00
R0CESDMD	1683	0.54	1.74	0.00	8.00
S0CESDMD	1208	0.46	1.57	0.00	8.00

Categorical Variable Codes

Value-----	R0DEPRES
.d=dk	26
.m=oth missing	7
.p=proxy	58
.r=rff	5
0.rarely or none of the time < 1 day	1050
1.some or a little of the time 1-2 days	398
2.occasionally or a moderate amount of 3	91
3.most or all of the time 5-7 days	48

Value-----	S0DEPRES
.d=dk	19
.m=oth missing	3
.p=proxy	30
.r=rff	4
.u=unmar	321
.v=sp nr	154
0.rarely or none of the time < 1 day	789
1.some or a little of the time 1-2 days	279
2.occasionally or a moderate amount of 3	51
3.most or all of the time 5-7 days	33

Value-----	R0EFFORTD
.d=dk	51
.m=oth missing	7
.p=proxy	58
.r=rff	5
0.rarely or none of the time < 1 day	1052
1.some or a little of the time 1-2 days	389
2.occasionally or a moderate amount of 3	84
3.most or all of the time 5-7 days	37

Value-----	S0EFFORTD
.d=dk	39
.m=oth missing	3
.p=proxy	30
.r=rff	4
.u=unmar	321
.v=sp nr	154
0.rarely or none of the time < 1 day	767

1.some or a little of the time 1-2 days	283
2.occasionally or a moderate amount of 3	51
3.most or all of the time 5-7 days	31

Value-----	R0SLEEPRD
.d=dk	25
.m=oth missing	7
.p=proxy	58
.r=rf	3
0.rarely or none of the time < 1 day	895
1.some or a little of the time 1-2 days	456
2.occasionally or a moderate amount of 3	167
3.most or all of the time 5-7 days	72

Value-----	S0SLEEPRD
.d=dk	20
.m=oth missing	3
.p=proxy	30
.u=unmar	321
.v=sp nr	154
0.rarely or none of the time < 1 day	685
1.some or a little of the time 1-2 days	313
2.occasionally or a moderate amount of 3	105
3.most or all of the time 5-7 days	52

Value-----	R0WHAPPYD
.d=dk	28
.m=oth missing	7
.p=proxy	58
.r=rf	5
0.rarely or none of the time < 1 day	725
1.some or a little of the time 1-2 days	382
2.occasionally or a moderate amount of 3	307
3.most or all of the time 5-7 days	171

Value-----	S0WHAPPYD
.d=dk	23
.m=oth missing	3
.p=proxy	30
.r=rf	2
.u=unmar	321
.v=sp nr	154
0.rarely or none of the time < 1 day	513
1.some or a little of the time 1-2 days	262
2.occasionally or a moderate amount of 3	241
3.most or all of the time 5-7 days	134

Value-----	R0FLONED
.d=dk	35
.m=oth missing	7
.p=proxy	58
.r=rf	5
0.rarely or none of the time < 1 day	1097
1.some or a little of the time 1-2 days	343
2.occasionally or a moderate amount of 3	91
3.most or all of the time 5-7 days	47

Value-----	S0FLONED
.d=dk	27
.m=oth missing	3
.p=proxy	30
.r=rf	1
.u=unmar	321
.v=sp nr	154
0.rarely or none of the time < 1 day	838
1.some or a little of the time 1-2 days	229
2.occasionally or a moderate amount of 3	52
3.most or all of the time 5-7 days	28

Value-----	R0FSADD
.d=dk	37
.m=oth missing	7
.p=proxy	58
.r=rf	5
0.rarely or none of the time < 1 day	986
1.some or a little of the time 1-2 days	411

2.occasionally or a moderate amount of 3	123
3.most or all of the time 5-7 days	56
Value-----	S0FSADD
.d=dk	28
.m=oth missing	3
.p=proxy	30
.r=rff	4
.u=unmar	321
.v=sp nr	154
0.rarely or none of the time < 1 day	750
1.some or a little of the time 1-2 days	280
2.occasionally or a moderate amount of 3	77
3.most or all of the time 5-7 days	36
Value-----	R0GOINGD
.d=dk	89
.m=oth missing	7
.p=proxy	58
.r=rff	12
0.rarely or none of the time < 1 day	1157
1.some or a little of the time 1-2 days	268
2.occasionally or a moderate amount of 3	55
3.most or all of the time 5-7 days	37
Value-----	S0GOINGD
.d=dk	67
.m=oth missing	3
.p=proxy	30
.r=rff	8
.u=unmar	321
.v=sp nr	154
0.rarely or none of the time < 1 day	861
1.some or a little of the time 1-2 days	176
2.occasionally or a moderate amount of 3	37
3.most or all of the time 5-7 days	26
Value-----	R0ENLIFED
.d=dk	52
.m=oth missing	7
.p=proxy	58
.r=rff	2
0.rarely or none of the time < 1 day	1114
1.some or a little of the time 1-2 days	333
2.occasionally or a moderate amount of 3	66
3.most or all of the time 5-7 days	51
Value-----	S0ENLIFED
.d=dk	40
.m=oth missing	3
.p=proxy	30
.r=rff	1
.u=unmar	321
.v=sp nr	154
0.rarely or none of the time < 1 day	839
1.some or a little of the time 1-2 days	222
2.occasionally or a moderate amount of 3	40
3.most or all of the time 5-7 days	33

How Constructed:

R0DEPRES, R0EFFORTD, R0SLEEP, R0WHAPPY, R0FLONED, R0GOINGD, R0SADD and R0ENLIFED are asked of the respondent's feelings much of the time over the week prior to the interview. R0DEPRES indicates whether the respondent was feeling depressed. R0EFFORTD indicates whether the respondent was feeling that everything was an effort. R0SLEEP indicates whether the respondent's sleep was restless. R0WHAPPY indicates whether the respondent felt happy. R0FLONED indicates whether the respondent felt lonely. R0GOINGD indicates whether the respondent felt he/she could not get going. R0SADD indicates whether the respondent felt sad. R0ENLIFED indicates whether the respondent felt he/she enjoyed life.

A code of 0 indicates that the respondent had the particular feeling rarely or none of the time < 1 day. A code of 1 indicates the respondent had the particular feeling some

or a little of the time 1-2 days. A code of 2 indicates that the respondent had the particular feeling occasionally or a moderate amount of 3-4 days. A code of 3 indicates that the respondent had the particular feeling most or all of the time 5-7 days. The questions were not asked if proxy answered the physical activities questions. If the questions were skipped because proxy, the answers are set to .p.

R0CESDD is the sum of R0DEPRESD, R0EFFORTD, R0SLEEPD, R0FLONED, R0SADD, R0GOINGD, (3-R0WHAPPYD), (3-R0ENLIFED). Thus the higher the score, the more negative the respondent's feelings were during the past week.

S0DEPRESD, S0EFFORTD, S0SLEEPD, S0WHAPPYD, S0FLONED, S0GOINGD, S0SADD and S0ENLIFED are taken from spouse's R0DEPRESD, R0EFFORTD, R0SLEEPD, R0WHAPPYD, R0FLONED, R0GOINGD, R0SADD and R0ENLIFED variables, respectively.

LASI Variables Used:

HT306	FELT DEPRESSED
HT307	FELT EVERYTHING WAS AN EFFORT
HT311	SLEEP WAS RESTLESS
HT312	FELT UNHAPPY
HT314	FELT LONELY
HT316	DID NOT ENJOY LIFE
HT318	FELT SAD
HT320	COULD NOT GET GOING

Doctor diagnosed health problems: Ever Have Condition
--

Wave	Variable	Label	Type
1	R0HIBPE	r0hibpe: w0 ever had high blood pressure	Categ
1	S0HIBPE	s0hibpe: w0 ever had high blood pressure	Categ
1	R0DIABE	r0diabe: w0 ever had diabetes	Categ
1	S0DIABE	s0diabe: w0 ever had diabetes	Categ
1	R0CANCRE	r0cancre: w0 ever had cancer	Categ
1	S0CANCRE	s0cancre: w0 ever had cancer	Categ
1	R0LUNGE	r0lunge: w0 ever had lung disease	Categ
1	S0LUNGE	s0lunge: w0 ever had lung disease	Categ
1	R0HEARTE	r0hearte: w0 ever had heart prob	Categ
1	S0HEARTE	s0hearte: w0 ever had heart prob	Categ
1	R0STROKE	r0stroke: w0 ever had stroke	Categ
1	S0STROKE	s0stroke: w0 ever had stroke	Categ
1	R0ARTHRE	r0arthre: w0 ever had arthritis	Categ
1	S0ARTHRE	s0arthre: w0 ever had arthritis	Categ
1	R0PSYCHE	r0psyche: w0 ever had psych prob	Categ
1	S0PSYCHE	s0psyche: w0 ever had psych prob	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0HIBPE	1672	0.17	0.38	0.00	1.00
S0HIBPE	1204	0.16	0.36	0.00	1.00
R0DIABE	1658	0.09	0.28	0.00	1.00
S0DIABE	1192	0.08	0.28	0.00	1.00
R0CANCRE	1678	0.00	0.05	0.00	1.00
S0CANCRE	1207	0.00	0.06	0.00	1.00
R0LUNGE	1658	0.04	0.20	0.00	1.00
S0LUNGE	1193	0.03	0.18	0.00	1.00
R0HEARTE	1665	0.03	0.18	0.00	1.00
S0HEARTE	1197	0.03	0.18	0.00	1.00
R0STROKE	1659	0.01	0.09	0.00	1.00

S0STROKE	1193	0.01	0.08	0.00	1.00
R0ARTHRE	1656	0.08	0.27	0.00	1.00
S0ARTHRE	1192	0.07	0.25	0.00	1.00
R0PSYCHE	1662	0.02	0.12	0.00	1.00
S0PSYCHE	1196	0.01	0.12	0.00	1.00

Categorical Variable Codes

Value-----	R0HIBPE
.d=dk	5
.m=oth missing	6
0.no	1381
1.yes	291

Value-----	S0HIBPE
.d=dk	1
.m=oth missing	3
.u=unmar	321
.v=sp nr	154
0.no	1016
1.yes	188

Value-----	R0DIABE
.d=dk	3
.m=oth missing	22
0.no	1512
1.yes	146

Value-----	S0DIABE
.d=dk	1
.m=oth missing	15
.u=unmar	321
.v=sp nr	154
0.no	1091
1.yes	101

Value-----	R0CANCRE
.d=dk	1
.m=oth missing	4
0.no	1673
1.yes	5

Value-----	S0CANCRE
.m=oth missing	1
.u=unmar	321
.v=sp nr	154
0.no	1203
1.yes	4

Value-----	R0LUNGE
.d=dk	1
.m=oth missing	24
0.no	1587
1.yes	71

Value-----	S0LUNGE
.m=oth missing	15
.u=unmar	321
.v=sp nr	154
0.no	1155
1.yes	38

Value-----	R0HEARTE
.d=dk	1
.m=oth missing	17
0.no	1609
1.yes	56

Value-----		S0HEARTE
.m=oth missing		11
.u=unmar		321
.v=sp nr		154
0.no		1159
1.yes		38

Value-----		R0STROKE
.d=dk		1
.m=oth missing		22
.r=rfr		1
0.no		1646
1.yes		13

Value-----		S0STROKE
.m=oth missing		14
.r=rfr		1
.u=unmar		321
.v=sp nr		154
0.no		1185
1.yes		8

Value-----		R0ARTHRE
.d=dk		1
.m=oth missing		26
0.no		1520
1.yes		136

Value-----		S0ARTHRE
.m=oth missing		16
.u=unmar		321
.v=sp nr		154
0.no		1109
1.yes		83

Value-----		R0PSYCHE
.d=dk		1
.m=oth missing		20
0.no		1637
1.yes		25

Value-----		S0PSYCHE
.m=oth missing		12
.u=unmar		321
.v=sp nr		154
0.no		1179
1.yes		17

How Constructed:

R0HIBPE, R0DIABE, R0CANCRE, R0LUNGE, R0HEARTE, R0STROKE, R0PSYCHE, and R0ARTHRE indicate the respondent's answer to the question regarding whether R0HIBPE, R0DIABE, R0CANCRE, R0LUNGE, R0HEARTE, R0STROKE, R0PSYCHE, and R0ARTHRE or not a doctor has told the respondent he/she had a specific condition. Don't know, not applicable, or refused values of R0HIBPE, R0DIABE, R0CANCRE, R0LUNGE, R0HEARTE, R0STROKE, R0PSYCHE, and R0ARTHRE are assigned special missing codes .d, .n, .r, respectively.

R0HIBPE indicates whether the respondent reported having high blood pressure or hypertension.

R0DIABE indicates whether the respondent reported having diabetes or high blood sugar.

R0CANCRE indicates whether the respondent reported having cancer or a malignant tumor (excluding minor skin cancers).

R0LUNGE indicates whether the respondent reported having chronic lung disease such as chronic bronchitis or emphysema.

R0HEARTE indicates whether the respondent reported having angina, a heart attack (including myocardial infarction or coronary thrombosis), congestive heart failure, a

heart murmur, an abnormal heart rhythm, or any other heart trouble. R0HEARTE indicates whether the respondent reported any of these conditions.

R0STROKE indicates whether the respondent reported having a stroke (cerebrovascular disease).

R0PSYCHE indicates whether the respondent reported having any emotional, nervous, or psychiatric problems.

R0ARTHRE indicates whether the respondent reported having arthritis (including osteoarthritis or rheumatism).

Respondents identify conditions by selecting them from a card containing a list of conditions (a so-called show card).

S0HIBPE, S0DIABE, S0CANCRE, S0LUNGE, S0HEARTE, S0STROKE, S0PSYCHE, and S0ARTHRE are taken from the spouse's R0HIBPE, R0DIABE, R0CANCRE, R0LUNGE, R0HEARTE, R0STROKE, R0PSYCHE, and R0ARTHRE variables.

LASI Variables Used:

HT002	HEALTH PROFESSIONAL TOLD YOU THAT YOU HAVE HBP OR HYPERT
HT006	ANY HEALTH PROFESSIONAL TOLD YOU THAT YOU HAVE DIABETES
HT012	DOCTOR TOLD YOU THAT YOU HAVE CANCER
HT019	DOCTOR TOLD YOU THAT YOU HAVE CHRONIC LUNG DISEASE
HT023	DOCTOR TOLD YOU THAT YOU HAVE HEART PROBLEMS
HT032	DOCTOR TOLD YOU THAT YOU HAD STROKE
HT043	DOCTOR TOLD YOU THAT YOU HAVE ARTHRITIS
HT050	DOCTOR TOLD YOU THAT YOU HAVE PSYCHIATRIC PROBLEMS

Health behaviors: Physical Activity or Exercise

Wave	Variable	Label	Type
1	R0VGACTX_L	r0vgactx_l: w0 frequent vigorous physical activity or exerci	Categ
1	S0VGACTX_L	s0vgactx_l: w0 frequent vigorous physical activity or exerci	Categ
1	R0MDACTX_L	r0mdactx_l: w0 frequent moderate physical activity or exerci	Categ
1	S0MDACTX_L	s0mdactx_l: w0 frequent moderate physical activity or exerci	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0VGACTX_L	1677	3.88	1.66	1.00	5.00
S0VGACTX_L	1206	3.76	1.70	1.00	5.00
R0MDACTX_L	1676	3.28	1.88	1.00	5.00
S0MDACTX_L	1206	3.27	1.86	1.00	5.00

Categorical Variable Codes

Value-----	R0VGACTX_L
.m=oth missing	5
.r=rf	1
1.every day	333
2.more than once a week	119
3.once a week	71
4.1-3 times per month	45
5.hardly ever or never	1109

Value-----	S0VGACTX_L
.m=oth missing	2
.u=unmar	321
.v=sp nr	154
1.every day	260
2.more than once a week	103
3.once a week	54
4.1-3 times per month	38
5.hardly ever or never	751

Value-----	R0MDACTX_L
.d=dk	1
.m=oth missing	5
.r=rf	1
1.every day	605
2.more than once a week	93
3.once a week	76
4.1-3 times per month	33
5.hardly ever or never	869

Value-----	S0MDACTX_L
.m=oth missing	2
.u=unmar	321
.v=sp nr	154
1.every day	428
2.more than once a week	75
3.once a week	61
4.1-3 times per month	26
5.hardly ever or never	616

How Constructed:

R0VGACTX_L and R0MDACTX_L indicate frequency of vigorous or moderately energetic physical activity, respectively.

A code of 1 indicates the respondent reported taking part in the given level of physical activity every day, a code of 2 indicates more than once a week, a code of 3 indicates once a week, a code of 4 indicates one to three times a month, and code of 5 indicates hardly ever or never.

Don't know, not applicable, or refused values of R0VGACTX_L or R0MDACTX_L are assigned special missing codes .d, .n, .r, respectively.

R0VGACTX_L and R0MDACTX_L are taken from spouse's R0VGACTX_L and R0MDACTX_L variables.

LASI Variables Used:

HT218	HOW OFTEN PHYSICAL ACTIVITY
HT219	HOW OFTEN DO YOU TAKE PART IN MODERATELY ENERGETIC ACTIV

BMI

Wave	Variable	Label	Type
1	R0BMI	r0bmi: w0 body mass index=kg/m2	Cont
1	S0BMI	s0bmi: w0 body mass index=kg/m2	Cont
1	R0HEIGHT	r0height: w0 height in meters	Cont
1	S0HEIGHT	s0height: w0 height in meters	Cont
1	R0WEIGHT	r0weight: weight in kilograms	Cont
1	S0WEIGHT	s0weight: w0 weight in kilograms	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0BMI	1499	22.31	5.12	11.65	65.79
S0BMI	1088	22.40	4.95	11.65	55.75
R0HEIGHT	1526	1.58	0.10	1.02	1.91
S0HEIGHT	1104	1.59	0.10	1.02	1.91
R0WEIGHT	1503	55.79	14.06	28.00	160.50
S0WEIGHT	1091	56.80	13.68	28.00	105.00

How Constructed:

R0BMI, R0HEIGHT and R0WEIGHT are the respondent's body mass index, height and weight, and, respectively.

BMI is weight divided by the square of height. Height is given in meters, weight in kilograms.

LASI only provides height and weight information in the Biomarker data.

The raw biomarker data are available upon request.

S0BMI, S0HEIGHT and S0WEIGHT are taken from spouse's R0BMI, R0HEIGHT and R0WEIGHT variables.

Health behaviors: Drinking

Wave	Variable	Label	Type
1	R0DRINK	r0drinkb: w0 ever drinks any alcohol before	Categ
1	S0DRINK	s0drink: w0 ever drinks any alcohol before	Categ
1	R0DRINKN	r0drinkn: w0 # drinks/day when drinks	Cont
1	S0DRINKN	s0drinkn: w0 # drinks/day when drinks	Cont
1	R0DRINKD	r0drinkd: w0 r # days/week drinks	Cont
1	S0DRINKD	s0drinkd: w0 r # days/week drinks	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0DRINK	1675	0.13	0.34	0.00	1.00
S0DRINK	1206	0.15	0.36	0.00	1.00
R0DRINKN	121	6.67	7.29	0.00	30.00
S0DRINKN	97	6.85	7.44	0.00	30.00
R0DRINKD	116	3.14	2.07	0.00	7.00
S0DRINKD	92	3.25	1.99	0.00	7.00

Categorical Variable Codes

Value-----	R0DRINK
.d=dk	1
.m=oth missing	5
.r=rf	2
0.no	1454
1.yes	221

Value-----	S0DRINK
.m=oth missing	2
.u=unmar	321
.v=sp nr	154
0.no	1026
1.yes	180

How Constructed:

R0DRINK indicates whether the respondent has had an alcoholic drink during the last 12 months. A code of 0 indicates that the respondent reports not having had an alcoholic drink during the last 12 months. A code of 1 indicates that the respondent reports having had an alcoholic drink during the last 12 months.

R0DRINKN indicates the number drinks the respondent reported having an alcoholic drink in the last month. If a respondent reports they did not have any drink in the past week this R0DRINKN is assigned a value of 0.

R0DRINKD indicates the number of days the respondent reported having on the day when they drank the most during the previous week. The question is only asked if the

respondent reported having an alcoholic drink currently. If a respondent reports they did not have any drink in the past week R0DRINKN is assigned a value of 0.

S0DRINK, S0DRINKN and S0DRINKD are taken from spouse's R0DRINK, R0DRINKN and R0DRINKD, respectively.

LASI Variables Used:

HT214	CONSUMPTION OF ALCOHOLIC BEVERAGES
HT216	NUMBER OF DRINKS IN LAST MONTH
HT217	NUMBER OF DAYS PER WEEK DRINKING

Health behaviors: Smoking (cigarettes)

Wave	Variable	Label	Type
1	R0SMOKEV	r0smokev: w0 ever smoke	Categ
1	S0SMOKEV	s0smokev: w0 ever smoke	Categ
1	R0SMOKEN	r0smoken: w0 smoke now	Categ
1	S0SMOKEN	s0smoken: w0 smoke now	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0SMOKEV	1677	0.18	0.39	0.00	1.00
S0SMOKEV	1205	0.19	0.39	0.00	1.00
R0SMOKEN	1676	0.14	0.35	0.00	1.00
S0SMOKEN	1205	0.15	0.35	0.00	1.00

Categorical Variable Codes

Value-----	R0SMOKEV
.d=dk	1
.m=oth missing	5
0.no	1373
1.yes	304

Value-----	S0SMOKEV
.d=dk	1
.m=oth missing	2
.u=unmar	321
.v=sp nr	154
0.no	975
1.yes	230

Value-----	R0SMOKEN
.m=oth missing	7
0.no	1443
1.yes	233

Value-----	S0SMOKEN
.m=oth missing	3
.u=unmar	321
.v=sp nr	154
0.no	1028
1.yes	177

How Constructed:

R0SMOKEV indicates whether the respondent reports ever smoked tobacco such as cigarette, bidi, cigar, hookah or used smokeless tobacco, such as chewing tobacco, qutka, pan massla. A code of 0 indicates that the respondent reports never having smoked. A code of 1 indicates that the respondent reports having ever smoked. Don't know, not applicable, or refused values to R0SMOKEV are assigned special missing codes .d, .n, .r, respectively.

R0SMOKEN indicates whether the respondent reports smoking, chewing or sniffing tobacco at all nowadays. This question is only asked if the respondent reports ever smoking. If the respondent reported never smoking R0SMOKEN is assigned a value of "no". A code

of 0 indicates that the respondent reports not smoking nowadays. A code of 1 indicates that the respondent reports smoking nowadays. Don't know, not applicable, or refused values of R0SMOKEN are assigned special missing codes .d, .n, .r, respectively.

S0SMOKEV and S0SMOKEN are taken from spouse's R0SMOKEN and R0SMOKEN variables.

LASI Variables Used:

HT205	SMOKED TOBACCO OR SMOKELESS TOBACCO
HT206S1	CURRENTLY SMOKE OR QUIT
HT206S1	CURRENTLY SMOKE OR QUIT
HT206S3	CURRENTLY SMOKE OR QUIT

Section D: Income

Individual Earnings

Wave	Variable	Label	Type
1	R0IEARN	r0iearn: w0 income: individual earning	Cont
1	S0IEARN	s0iearn: w0 income: individual earning	Cont
1	R0IFEARN	r0ifearn: w0 incflag: individual earning	Categ
1	S0IFEARN	s0ifearn: w0 incflag: individual earning	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0IEARN	1683	1785.15	12659.57	0.00	240000.00
S0IEARN	1208	2087.87	14266.73	0.00	240000.00
R0IFEARN	1683	0.01	0.08	0.00	1.00
S0IFEARN	1208	0.01	0.08	0.00	1.00

Categorical Variable Codes

Value-----	R0IFEARN
0.no	1671
1.yes	12
Value-----	S0IFEARN
.u=unmar	321
.v=sp nr	154
0.no	1200
1.yes	8

How Constructed:

R0IEARN is the sum of respondent's primary job wage/salary income, self-employment income, and 2nd job earnings from last 12 months.

The amount is expressed as India Rupee. The amount is imputed if missing.

R0IFEARN indicates whether the amount is imputed or not.

LASI Variables Used:

WE002	WORK FOR AT LEAST ON HOUR LAST WEEK
WE206	AT LEAST ONE HOUR LAST WEEK IN NON AGRICULTURAL WAGE OR S
WE207	AGRICULTURAL WORKER BUT ON LEAVE
WE329	WORK FOR SOMEONE ELSE OR SELF EMPLOYED
WE363	CURRENT PAY AFTER TAXES AND OTHER DEDUCTIONS
WE364	CURRENT PAY AFTER TAXES PER UNIT
WE365	HOW MUCH EARNED FROM WORKING ON JOB IN PAST 12 MO AFTER
WE366S1	FRINGE BENEFITS PROVIDED BY COMPANY
WE366S2	FRINGE BENEFITS PROVIDED BY COMPANY
WE366S3	FRINGE BENEFITS PROVIDED BY COMPANY
WE366S4	FRINGE BENEFITS PROVIDED BY COMPANY
WE366S6	FRINGE BENEFITS PROVIDED BY COMPANY
WE366S5	FRINGE BENEFITS PROVIDED BY COMPANY

WE366S7	FRINGE BENEFITS PROVIDED BY COMPANY
WE377S8	FRINGE BENEFITS PROVIDED BY COMPANY
WE424	HOW MUCH OR PROFITS CONSTITUTE OWN INCOME
WE437	CURRENTLY ANY JOBS IN ADDITION TO MAIN JOB
WE440	AVERAGE PRE TAX MONTHLY INCOME OR WAGE FROM SIDE JOB

Income from Employer Pension

Wave	Variable	Label	Type
1	R0IPEN	r0ipen: w0 income: income from pension	Cont
1	S0IPEN	s0ipen: w0 income: income from pension	Cont
1	R0IFPEN	r0ifpen: w0 incflag: income from pension	Categ
1	S0IFPEN	s0ifpen: w0 incflag: income from pension	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0IPEN	1683	300.54	1631.75	0.00	24000.00
S0IPEN	1208	270.08	1677.08	0.00	24000.00
R0IFPEN	1683	0.00	0.07	0.00	1.00
S0IFPEN	1208	0.00	0.06	0.00	1.00

Categorical Variable Codes

Value-----	R0IFPEN
0.no	1675
1.yes	8
Value-----	S0IFPEN
.u=unmar	321
.v=sp nr	154
0.no	1203
1.yes	5

How Constructed:

R0IPEN is the sum of respondent's income from all pensions source. These different pension sources are

- 1) current pension amount from the officially retired organized sector,
- 2) current pension amount from previous employer for those who are no longer working,
- 3) current pension amount from commercial pension,
- 4) current pension amount from government: widowers, the disabled, agricultural workers, freedom fighters and the elderly may receive pension payments from the government,
- 5) survival pension: only widowers are asked if they receive a survival pension from the employment of the late spouse

The amount does not include the lump sum amount.

The amount is expressed as India Rupee. The amount is imputed if missing.

R0IFPEN indicates whether the amount is imputed or not.

We used several variables from pension section to derive these variables but the variables from pension section are not available for public.

LASI Variables Used:

WE001	ENGAGED IN AGRICULTURAL WORK IN PAST YEAR
WE002	WORK FOR AT LEAST ON HOUR LAST WEEK
WE003	EVER WORKED FOR PAY FOR MORE THAN A FEW MONTHS
WE003	EVER WORKED FOR PAY FOR MORE THAN A FEW MONTHS

Family Capital Income

Wave	Variable	Label	Type
1	H0ICAP	h0icap: w0 income: hh capital income	Cont
1	H0IFCAP	h0ifcap: w0 incflag: hh capital income	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0ICAP	1683	21936.17	125374.06	0.00	2500000.00
H0IFCAP	1683	0.03	0.18	0.00	1.00

Categorical Variable Codes

Value	H0IFCAP
0.no	1626
1.yes	57

How Constructed:

H0ICAP is the sum of household net income from agricultural activities and income from all assets, such as, business or agricultural income, gross rent, dividend and interest income, trust funds or other asset income.

The amount is expressed as India Rupee. The amount is imputed if missing.

H0IFCAP indicates whether the amount is imputed or not.

LASI Variables Used:

AD017	RENT OUT HOUSING UNITS
AD018	TOTAL RENTAL INCOME LAST 12 MONTHS
AD102A	HOW MANY CARS
AD102B	HOW MANY TRUCKS
AD102C	HOW MANY OTHER AUTOMOBILES
AD102D	HOW MANY BICYCLES
AD102E	HOW MANY MOTORCYCLES
AD104A	RECEIVE INCOME FROM RENTING CAR
AD104B	RECEIVE INCOME FROM RENTING TRUCK
AD104C	RECEIVE INCOME FROM RENTING OTHER AUTOS
AD104D	RECEIVE INCOME FROM RENTING BICYCLE
AD104E	RECEIVE INCOME FROM RENTING MOTORCYCLE
AD104F	RECEIVE INCOME FROM RENTING SCOOTER
AD105A	RENTAL INCOME FOR CAR
AD105B	RENTAL INCOME FOR TRUCK
AD105C	RENTAL INCOME FOR OTHER AUTO
AD105D	RENTAL INCOME FOR OTHER BICYCLE
AD105E	RENTAL INCOME FOR MOTORCYCLE
AD105F	RENTAL INCOME FOR SCOOTER
AD120	RECEIVE INTEREST OR DIVIDENDS PAST YR
AD121	RETURNS ON THESE FINANCIAL INVESTMENTS PAST YR
AD122	ANYONE OWE YOU MONEY
AD124	RECEIVE INTEREST FROM PERSONAL LOANS IN PAST YR
AG001	HOUSEHOLD HAVE CULTIVATED LAND
AG006	SPEND ON IRRIGATION IN THE PAST YEAR
AG008	RENT OUT ANY LAND IN PAST 12 MONTHS

AG016	HOUSEHOLD OWN FARMING ASSETS
AG019A	TRACTOR RENTAL INCOME PAST YR
AG019B	PLOUGHING IMPLEMENT RENTAL INCOME
AG019C	RENTAL INCOME FROM CART
AG019D	RENTAL INCOME FROM THRESHER
AG019E	RENTAL INCOME FROM TROLLEY
AG019F	RENTAL INCOME FROM FOLDER CUTTING MACHINES
AG019G	RENTAL INCOME FROM GENERATORS
AG023	VALUE OF CROPS FORESTRY AND FISHING PAST YR
AG024	COST OF CROPS FORESTRY AND FISHING PAST 12 MO
AG025	HOUSEHOLD HAD ANY LIVESTOCK PAST 5 YRS
AG031_INCOME	RENTAL INCOME FROM RENTING OUT LIVESTOCK
AG031_RENT	INCOME FROM RENTING OUT LIVESTOCK
AG032	SELL LIVESTOCK PRODUCTS
AG033	VALUE OF LIVESTOCK PRODUCTS SOLD
AG034	COST OF PRODUCING LIVESTOCK PAST YR
AG116S1	WHICH FARMING ASSETS OWN
AG116S2	WHICH FARMING ASSETS OWN
AG116S3	WHICH FARMING ASSETS OWN
AG116S4	WHICH FARMING ASSETS OWN
AG116S5	WHICH FARMING ASSETS OWN
AG116S6	WHICH FARMING ASSETS OWN

Family Government Transfer Income

Wave	Variable	Label	Type
1	H0IGXFR	h0igxfr: w0 income: hh income from gov transfer w/ ration	Cont
1	H0IGXFRB	h0igxfrb: w0 income: hh income from gov transfer wo/ration	Cont
1	H0IFGXFR	h0ifgxfr: w0 incflag: hh income from gov transfer w/ ration	Categ
1	H0IFGXFRB	h0ifgxfrb: w0 incflag: hh income from gov transfer wo/ ratio	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0IGXFR	1683	9116.99	65015.55	0.00	864000.00
H0IGXFRB	1683	583.96	6082.63	0.00	93000.00
H0IFGXFR	1683	0.02	0.16	0.00	1.00
H0IFGXFRB	1683	0.00	0.05	0.00	1.00

Categorical Variable Codes

Value	H0IFGXFRB
0.no	1678
1.yes	5

How Constructed:

H0IGXFR is the income from government transfers including ration.

H0IGXFRB is the income from government transfers excluding ration.

The amount is expressed as India Rupee. The amount is imputed if missing.

H0IFGXFR indicates whether the amount is imputed or not.

LASI Variables Used:

IN107	HOUSEHOLD HAVE RATION CARD IN LAST 12 MONTHS
IN109_SETS1	RECEIVED GOVERNMENT SUBSIDIES OR TRANSFERS PAST 12 MO
IN109_SETS2	RECEIVED GOVERNMENT SUBSIDIES OR TRANSFERS PAST 12 MO
IN109_SETS3	RECEIVED GOVERNMENT SUBSIDIES OR TRANSFERS PAST 12 MO
IN109_SETS4	RECEIVED GOVERNMENT SUBSIDIES OR TRANSFERS PAST 12 MO
IN109_SETS5	RECEIVED GOVERNMENT SUBSIDIES OR TRANSFERS PAST 12 MO
IN109_SETS6	RECEIVED GOVERNMENT SUBSIDIES OR TRANSFERS PAST 12 MO
IN109_SETS7	RECEIVED GOVERNMENT SUBSIDIES OR TRANSFERS PAST 12 MO
IN109_SETS8	RECEIVED GOVERNMENT SUBSIDIES OR TRANSFERS PAST 12 MO
IN109_SETS9	RECEIVED GOVERNMENT SUBSIDIES OR TRANSFERS PAST 12 MO

Family Income from Remittances

Wave	Variable	Label	Type
1	H0IREMI	h0iremi: w0 income: hh income from remittance	Cont
1	H0IFREMI	h0ifremi: w0 incflag: hh income from remittance	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0IREMI	1683	8533.04	64807.10	0.00	864000.00
H0IFREMI	1683	0.02	0.15	0.00	1.00

Categorical Variable Codes

Value	H0IFREMI
0.no	1646
1.yes	37

How Constructed:

H0IREMI is the income from remittances. Remittances from up to 3 sources are summed to obtain the summary measures. This includes both regular and irregular remittances.

The amount is expressed as India Rupee. The amount is imputed if missing.

H0IFREMI indicates whether the amount is imputed or not.

LASI Variables Used:

IN117	HOUSEHOLD RECEIVE ANY REMITTANCES FROM HOUSEHOLD MEMBERS
IN123_1_	DONOR SENT PAST 12 MONTHS SPECIFY TIME PERIOD
IN123_2_	DONOR SENT PAST 12 MONTHS SPECIFY TIME PERIOD
IN123_3_	DONOR SENT PAST 12 MONTHS SPECIFY TIME PERIOD

Family Total Private Transfers

Wave	Variable	Label	Type
1	H0IPTRAN	h0iptran: w0 income: hh income from private transfer	Cont
1	H0IFPTRAN	h0ifptran: w0 incflg: hh income from private transfer	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0IPTRAN	1683	10018.58	69571.08	0.00	864000.00
H0IFPTRAN	1683	0.02	0.15	0.00	1.00

Categorical Variable Codes

Value	H0IFPTRAN
0.no	1645
1.yes	38

How Constructed:

H0IPTRAN is the income from private transfers. Respondents are asked to report all other transfers from private sources, e.g., gifts, donations, and inheritances.

The question was asked for monthly amounts and # of months received. The yearly value was calculated.

The amount is expressed as India Rupee. The amount is imputed if missing.

H0IFPTRAN indicates whether the amount is imputed or not.

LASI Variables Used:

IN126	OTHER PRIVATE TRANSFERS LAST 12 MONTHS
IN127	VALUE OF GIFTS LAST 12 MONTHS

Total Household Income

Wave	Variable	Label	Type
1	H0ITOT	h0itot: w0 income: total hh income	Cont
1	H0IFTOT	h0iftot: w0 incflag: total hh income	Categ

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0ITOT	1683	110956.11	222165.74	0.00	2740000.00
H0IFTOT	1683	0.10	0.30	0.00	1.00

Categorical Variable Codes

Value	H0IFTOT
0.no	1514
1.yes	169

How Constructed:

H0ITOT is the sum of all income in the household, that is, the sum of HwIEMP, HwISLEMP, HwICAP, HwIGXFR, H0IREMI, H0IPTRAN, and other income for the last 12 months.

The amount is expressed as India Rupee. The amount is imputed if missing.

H0IFTOT indicates whether the amount is imputed or not.

Differences with the RAND HRS

Total family income in LASI is expressed in Rupee, whereas the equivalent measure in RAND HRS is in current dollars. Therefore, conversion into a common currency is necessary before comparison of these data. Components included in LASI and RAND HRS are different for this variable, as described above, representing different institutional arrangements in each country. In any case, we kept the concepts included as comparable as possible.

Section H: Family Structure

Number of People Living in the Household

Wave	Variable	Label	Type
1	H0HHRES	h0hhres: w0 # of people living in the household	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0HHRES	1682	5.23	2.83	1.00	20.00

How Constructed:

H0HHRES indicates the total number of people usually live in the household. This is taken from the self-reported answer from the family respondent.

LASI Variables Used:

CV002	NUMBER OF PEOPLE IN THE HOUSEHOLD
-------	-----------------------------------

Number of Children

Wave	Variable	Label	Type
1	R0CHILD	r0child: w0 # of living children	Cont
1	S0CHILD	s0child: w0 # of living children	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
R0CHILD	1679	2.99	1.83	0.00	15.00
S0CHILD	1207	3.04	1.77	0.00	15.00

How Constructed:

R0CHILD indicates the number of living children that the respondent has.

S0CHILD is taken from the spouse's R0CHILD variable.

LASI Variables Used:

FS201	HOW MANY LIVING CHILDREN
-------	--------------------------

Section K: Consumption

Total Household Food Consumption

Wave	Variable	Label	Type
1	H0CFOOD	h0cfood: total food consumption(inc home grown)	Cont
1	H0CFOODB	h0cfoodb: total food consumption(ex home grown)	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0CFOOD	1683	74064.72	61097.31	0.00	711800.00
H0CFOODB	1683	56075.58	45647.28	0.00	661800.00

How Constructed:

H0CFOOD is the total household food expenditures for last 12 months. This measure is comprised of three types of food expenditures:

- 1) Purchased food
- 2) Home grown or in-kind food transfer
- 3) Meals eaten out

H0CFOODB is the food expenditures excluding the 2)Home grown.

These questions were asked on a daily, weekly, or monthly basis. All purchase amounts are adjusted to a 12-month value for aggregation. There are some adjustments made for extreme values and discrepancies.

The amount is expressed as India Rupee. The amount is imputed if missing.

LASI Variables Used:

CO002A	STAPLE FOODS PURCHASING FREQUENCY
CO002B	RUPEES SPENT ON PURCHASED STAPLE FOODS LAST 30 DAYS
CO002C	RUPEES SPENT ON HOME GROWN STAPLE FOODS IN LAST 30 DAYS
CO003A	PURCHASING FREQUENCY OF GRAINS
CO003B	RUPEES SPENT ON PURCHASED GRAINS
CO003C	RUPEES SPENT ON HOME GROWN PULSES
CO004A	PURCHASING FREQUENCY OF MILK PRODUCTS
CO004B	RUPEES SPENT ON PURCHASED MILK PRODUCTS
CO004C	RUPEES SPENT ON HOME GROWN MILK PRODUCTS
CO005A	PURCHASING FREQUENCY OF VEGETABLES
CO005B	RUPEES SPENT ON PURCHASED VEGETABLES
CO005C	RUPEES SPENT ON HOME GROWN VEGETABLES
CO006A	PURCHASING FREQUENCY OF FRUITS
CO006B	RUPEES SPENT ON PURCHASED FRUITS
CO006C	RUPEES SPENT ON HOME GROWN FRUITS
CO007A	PURCHASING FREQUENCY OF SPICES AND OILS
CO007B	RUPEES SPENT ON PURCHASED SPICES AND OILS
CO007C	RUPEES SPENT ON HOME GROWN SPICES AND OILS
CO008A	PURCHASING FREQUENCY OF SUGAR AND SUGAR PRODUCTS
CO008B	RUPEES SPENT ON PURCHASED SUGAR AND SUGAR PRODUCTS
CO008C	RUPEES SPENT ON HOME GROWN SUGAR AND SUGAR PRODUCTS
CO009A	PURCHASING FREQUENCY OF EGG, FISH, MEAT
CO009B	RUPEES SPENT ON PURCHASED EGG, FISH, MEAT
CO009C	RUPEES SPENT ON HOME GROWN EGG, FISH, MEAT

CO010A	PURCHASING FREQUENCY OF NON ALCOHOLIC DRINKS
CO010B	RUPEES SPENT ON PURCHASED NON ALCOHOLIC DRINKS
CO010C	RUPEES SPENT ON HOME GROWN NON ALCOHOLIC DRINKS
CO011A	PURCHASING FREQUENCY OF ALL OTHER FOOD ITEMS
CO011B	RUPEES SPENT ON PURCHASED ALL OTHER FOOD ITEMS
CO011C	RUPEES SPENT ON HOME GROWN ALL OTHER FOOD ITEMS
CO012	SPENT ON DINING OUT LAST 30 DAYS

Total Household Consumption

Wave	Variable	Label	Type
1	H0CTOT	h0ctot: total hh expenditure in last yr(inc home grown)	Cont
1	H0CTOTB	h0ctotb: total hh expenditure in last yr(ex home grown)	Cont

Descriptive Statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
H0CTOT	1683	158445.65	206324.25	0.00	2803428.00
H0CTOTB	1683	140706.88	197587.28	0.00	2735028.00

How Constructed:

H0CTOT is the total household expenditures for last 12 months. This measure is comprised of food and non-food expenditures. The food expenditure is described above. For the non-food expenditures are included:

1) Regular household fee, such as communication fees, utilities, payments for household servants, home repair, and vehicle repair.

2) Taxes

3) Loans

4) Insurance

5) Fuel

6) Durables

7) Education

8) Entertainment

9) Clothing

10) Medical expense

11) Transportation

12) Alcohol and tobacco

13) Holiday celebrations

14) Remittances

15) Rental expenses

16) Other expenditure

H0CTOTB is the food expenditures excluding the 2)Home grown.

All sub-items are converted to 12-month value before being totaled up.

The amount is expressed as India Rupee. The amount is imputed if missing.

LASI Variables Used:

CO015	SPENT ON COMMUNICATION FEES IN LAST MONTH
CO016	RUPEES SPENT ON UTILITIES LAST MONTH
CO017	RUPEES SPENT ON FUEL LAST MONTH
CO018	RUPEES SPENT ON SERVANTS LAST MONTH
CO019	SPENT ON TRANSPORTATION LAST MONTH
CO020	SPENT ON HOUSEHOLD ITEMS LAST MONTH
CO021	SPENT ON PERSONAL CARE LAST MONTH
CO022	SPENT ON ENTERTAINMENT LAST MONTH
CO023	SPENT ON CLOTHING LAST MONTH
CO024A	FREQUENCY OF FOOT WEAR AND UMBRELLA PURCHASE
CO024B	FOOT WEAR AND UMBRELLA LAST 12 MONTHS
CO025A	FREQUENCY OF LONG DISTANCE TRAVEL PURCHASE
CO025B	SPENT ON LONG DISTANCE TRAVEL LAST 12 MONTHS
CO030A	RUPEES SPENT EDUCATION PURCHASING FREQUENCY
CO030B	RUPEES SPENT ON EDUCATION LAST 12 MONTHS
CO032A	FREQUENCY OF HEALTH CARE SPENDING
CO032B	RUPEES SPENT ON HEALTH CARE LAST 12 MONTHS
CO033A	FREQUENCY OF INSURANCE PREMIUM SPENDING
CO033B	RUPEES SPENT ON INSURANCE PREMIUM LAST 12 MONTHS
CO034A	FREQUENCY OF HOME MAINTENANCE AND REPAIR EXPENSES
CO034B	HOME MAINTENANCE AND REPAIR EXPENSES LAST 12 MONTHS
CO035A	FREQUENCY OF VEHICLE SERVICE CHARGES
CO035B	VEHICLE SERVICE CHARGES LAST 12 MONTHS
CO036A	FREQUENCY OF TAXES
CO036B	TAX CHARGES LAST 12 MONTHS
CO038A	FREQUENCY OF LOAN REPAYMENTS
CO038B	LOAN REPAYMENTS LAST 12 MONTHS