

This document was generated: December 4, 2021 at 11:59 PM

SIPP Synthetic Beta v5.1

(102 variables)

Last update to metadata: 2015-10-26 13:25:57 (upload date)

Document Date: June 18, 2014

Codebook prepared by: Cornell NSF-Census Research Network

Data prepared by: United States Department of Commerce. Bureau of the Census.

Principal Investigator(s): United States Department of Commerce. Bureau of the Census. , Social Security Administration. , Internal Revenue Service. , and Cornell University. Labor Dynamics Institute.

Citation

Please cite this codebook as:

Comprehensive Extensible Data Documentation and Access Repository. Codebook for the SIPP Synthetic Beta 5.1 [Codebook file]. Cornell Institute for Social and Economic Research and Labor Dynamics Institute [distributor]. Cornell University, Ithaca, NY, 2014

Please cite this dataset as:

U.S. Census Bureau. SIPP Synthetic Beta: Version 5.1 [Computer file]. Washington DC; Cornell University, Synthetic Data Server [distributor], Ithaca, NY, 2013

Abstract

The SIPP Synthetic Beta (SSB) is a Census Bureau product that integrates person-level micro-data from a household survey with administrative tax and benefit data. These data link respondents from the Survey of Income and Program Participation (SIPP) to Social Security Administration (SSA)/Internal Revenue Service (IRS) Form W-2 records and SSA records of retirement and disability benefit receipt, and were produced by Census Bureau staff economists and statisticians in collaboration with researchers at Cornell University, the SSA and the IRS. The purpose of the SSB is to provide access to linked data that are usually not publicly available due to confidentiality concerns. To overcome these concerns, Census has synthesized, or modeled, all the variables in a way that changes the record of each individual in a manner designed to preserve the underlying covariate relationships between the variables. The only variables that were not altered by the synthesis process and still contain their original values are gender and a link to the first reported marital partner in the survey. Seven SIPP panels (1990, 1991, 1992, 1993, 1996, 2001, 2004) form the basis for the SSB, with a large subset of variables available across all the panels selected for inclusion and harmonization across the years. Administrative data were added and some editing was done to correct for logical inconsistencies in the IRS/SSA earnings and benefits data.

Datasets

ssb_v5_1_synthetic1_1.sas7bdat

<http://www.census.gov/programs-surveys/sipp/methodology/sipp-synthetic-beta-data-product.html> (SAS)

ssb_v5_1_synthetic1_1.dta

<http://www.census.gov/programs-surveys/sipp/methodology/sipp-synthetic-beta-data-product.html> (Stata)

Terms of Use

Access Levels

released

The data can only be used on the VirtualRDC Synthetic Data Server at Cornell University. While no SSB data downloads are permitted at this time, users do not have to operate behind the Census Bureau firewall to access this server.

restricted

No description given

Access Restrictions (Default)

The data can only be used on the VirtualRDC Synthetic Data Server at Cornell University. While no SSB data downloads are permitted at this time, users do not have to operate behind the Census Bureau firewall to access this server.

Access Requirements

Researchers interested in using the SSB can submit an application to the Census Bureau. The application form and instructions can be downloaded from

<http://www.census.gov/programs-surveys/sipp/methodology/sipp-synthetic-beta-data-product.html> . Applications will be judged solely of feasibility of the proposed project (i.e., that the necessary variables are available on the SSB). Once an application has been accepted, the new user will be given an account on a server where the data can be accessed and analyzed.

Additional information:

<http://www.census.gov/programs-surveys/sipp/methodology/sipp-synthetic-beta-data-product.html>

Access Permission Requirements

The SSB files have been cleared by the Census Bureau Disclosure Review Board, SSA, and IRS for use by individuals without Census Bureau Special Sworn Status and outside of Census Bureau facilities.

Citation Requirements

We request that researchers who publish results from analyses done using these data cite the SSB as their data source and acknowledge the use of the SDS server at Cornell and the support of Census staff in running any validation programs. These citations will help ensure continued funding for the SDS server and the creation of the Gold Standard File and the SSB.

Suggested acknowledgement:

This analysis was first performed using the SIPP Synthetic Beta (SSB) on the Synthetic Data Server housed at Cornell University which is funded by NSF Grant #SES-1042181. These data are public use and may be accessed by researchers outside secure Census facilities. For more information, visit http://www.census.gov/sipp/synth_data.html. Final results for this paper were obtained from a validation analysis conducted by Census Bureau staff using the SIPP

Completed Gold Standard Files and the programs written by this author and originally run on the SSB. The validation analysis does not imply endorsement by the Census Bureau of any methods, results, opinions, or views presented in this paper.

Disclaimer

The data synthesis process employed by Census to protect the linked data from the risk of disclosing the identity of individuals is relatively new and substantially changes both the survey and administrative data. The intent of the modeling done as part of the synthesis is to preserve relationships among variables that are of interest to researchers while ensuring that personally identifiable information is not revealed to the data user. It has not been feasible to ensure accuracy by comparing every relationship among SSB variables with the corresponding relationship in the underlying confidential micro-data. Hence, we strongly urge researchers not to publish results produced from the SSB without first requesting that Census validate these results with confidential data housed in a secure environment at the Census Bureau. Census will perform this validation free of charge to researchers, as resources permit and according to the protocol established by the three agencies involved and outlined below. Without validation of results, Census, SSA, and IRS make no guarantee of the validity of the SSB for any research purpose. See <http://www.census.gov/programs-surveys/sipp/methodology/sipp-synthetic-beta-data-product.html> for validation conditions.

Contact

For questions regarding this data collection, please contact: sehsd.synthetic.data.use.list@census.gov

Additional Information

Related Material

I. Using SSB:

The GSF and Completed Data implicates contain personally identifiable information protected by Titles 13, 26, and 42 and cannot be accessed without Census Bureau Special Sworn Status nor outside of Census Bureau facilities. The SSB files, however, have been cleared by the Census Bureau Disclosure Review Board, SSA, and IRS for use by individuals without Census Bureau Special Sworn Status and outside of Census Bureau facilities.

Researchers interested in using the SSB can submit an application to the Census Bureau. The application form and instructions can be downloaded from

<http://www.census.gov/programs-surveys/sipp/methodology/sipp-synthetic-beta-data-product.html> . Applications will be judged solely on feasibility of the proposed project (i.e., that the necessary variables are available on the SSB). Once an application has been accepted, the new user will be given an account on a server where the data can be accessed and analyzed. While no SSB data downloads are permitted at this time, users do not have to operate behind the Census Bureau firewall to access this server.

The SSB is designed to be analytically valid in that sense that point estimates should be unbiased and estimated variances should lead to inferences similar to those that would be drawn from an identical analysis on the Completed Data implicates. Initial tests of analytic validity of the SSB have been promising. All SSB users are invited to help further test the analytic validity of the SSB by submitting programs used to analyze the SSB to be run on the Completed Data and/or Gold Standard files. Users need only inform Census Bureau staff

of the location on the server of such programs and work with Census Bureau staff to ensure that the programs run without error. Census Bureau staff will run the programs on the confidential data and release to the user resulting output that are cleared for release by the Census Bureau Disclosure Review Board. In order to evaluate the effects of the data synthesis separate from the effect of imputing missing data, comparisons should be made between results from the SSB and the Completed Data. To evaluate the effects of missing data imputation, comparisons should be made between results from the Completed Data and the Gold Standard.

II. Codebook for v5.1 at

http://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/SSB_v5_1_Codebook.pdf

III. When analyzing the SSB, users should account for the multiple imputation aspect of the SSB by averaging statistics of interests across all sixteen implicates. Variance measures should be created following the appropriate multiple imputation formulae as described in the document Using the SIPP Synthetic Beta for Analysis .

IV. Protocol for Validation of Results:

Census will validate results obtained from the SSB on the internal, confidential version of these data (Completed Gold Standard Files). Users who wish to obtain validated results should follow the protocol outlined here. The restricted access site will provide SAS and Stata analysis software and a computing environment similar to the one used to analyze the confidential Completed Gold Standard data on Census Bureau internal computers. Researchers should follow the Census Bureau programming requirements described in SSB Validation Request Guidelines to ensure that the programs will successfully transfer to internal Census computers for validation. Researchers should plan to share their results and programs from the synthetic data analysis with Census, ORES/SSA and SOI/IRS. After programs have successfully run without error on the synthetic data, researchers may request that Census run these programs on the Completed Gold Standard Files. Only programs successfully run without error on the SDS will be eligible to be run on the confidential data by Census staff. Any programs that produce errors on the Completed Gold Standard Files will be returned to users for correction. Once an analysis has been repeated on the Completed Gold Standard File, the results will be reviewed by Census staff for disclosure concerns. Researchers should familiarize themselves with standard Census disclosure rules for outside projects (See the RDC Researcher Handbook here) and should fill out the appropriate memo documenting the requested output (see RDC Disclosure Request Memo). Data products and output approved by Census staff will be released to the users, ORES/SSA, and SOI/IRS. The validation process can be accomplished in as little as one week for simple results that are generated by clean code and have no disclosure issues. However if the code does not run properly, the sample sizes are too small, or the researcher does not accurately fill out the disclosure memo, the process can take much longer. Census makes no guarantee on the length of time between submission of programs and the release of results from the confidential data. For more information about the validation process, including advice on how to make the process go smoothly and quickly, please see SSB Validation Request Guidelines .

This documentation derived from:

Test

Variable Groups - SIPP Synthetic Beta v5.1

Aged Spouse Benefit

Benefits Variables

Demographic Variables

Detailed Earnings Record Variables

Disability Benefit

Disability Variables

Economic Variables

Education Variables

Fertility Variables

Geographic Variables

Health Insurance Variables

IRS/SSA Variables

Identifiers

Income Variables

Labor Force Variables

Lifespan Variables

MBR/PHUS Variables

Marital History Variables

Other Benefit

Retirement Benefit

SIPP Arrays

Summary Earnings Record Variables

Supplemental Security Record Variables

Widowed Spouse Benefit

Variable Name

cur_endmar_flag

Label

Flag: Linked marriage ended

Concept

Type

numeric

Files

F1 F2

Values (3 total)

0	Linked marriage did not end
1	Linked marriage ended
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

Marital History Variables

Variable Name

cur_endmar_reas

Label

Flag: Reason linked marriage ended

Concept

Type

numeric

Files

F1 F2

Values (3 total)

0	Linked marriage ended in divorce
1	Linked marriage ended because of a death
Sysmiss	There is no linked marriage OR there is a linked marriage but it did not end during panel

Value Ranges

Value Range

Range: [0 , 1]

Groups

Marital History Variables

Variable Name	panel
Label	SIPP Panel Year
Concept	
Type	numeric
Files	F1 F2

Full Description

indicates panel of source record

Values (7 total)

1990
1991
1992
1993
1996
2001
2004

Value Ranges

Value Range

Range: [1990 , 2004]

Groups

Identifiers

Variable Name sipp_panel_end_date

Label SIPP Panel End Date

Concept

Type numeric

Files F1 F2

Full Description

The date the SIPP panel ended.

Value Ranges

Value Range

Range: [11809 , 17501]

Groups

Identifiers

Variable Name

cur_startmar

Label

SAS Date linked marriage began

Concept

Type

numeric

Files

F1 F2

Full Description

Date the linked marriage started

Value Ranges

Value Range

Range: [-9010.4448454821 , 17471]

Groups

Marital History Variables

Variable Name

cur_endmar

Label

SAS Date linked marriage ended

Concept

Type

numeric

Files

F1 F2

Full Description

Date the linked marriage ended

Value Ranges

Value Range

Range: [10958 , 16071]

Groups

Marital History Variables

Variable Name	totfam_kids
Label	Total Number of Children in Family
Concept	
Type	numeric
Files	F1 F2

Full Description

Number of children under the age of 18 that live in a family in the interview month in which marital status is first observed (for those without spouses during the course of the SIPP) or in which the respondent's spouse is assigned. This number is the same for all family members and does not indicate that the children are related to a particular individual (fnkids for 1990-1993 panels, rfnkids for 1996-2004 panels).

Value Ranges

Value Range

Range: [0 , 20]

Groups

Fertility Variables

Variable Name	mh1
Label	Flag: Marital History Event 1
Concept	
Type	numeric
Files	F1 F2

Full Description

First marital history event flag. Indicates whether respondent entered into a first marriage.

Values (3 total)

0	Never married
1	First marriage occurred
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

Marital History Variables

Variable Name	mh2
Label	Flag: Marital History Event 2
Concept	
Type	numeric
Files	F1 F2

Full Description

Second marital history event flag. Indicates whether respondent's first marriage ended in widowhood or divorce.

Values (4 total)

0	First Marriage did not end over course of survey
1	First marriage ended in widowhood
2	First marriage ended in divorce/separation
Sysmiss	

Value Ranges

Value Range

Range: [0 , 2]

Groups

Marital History Variables

Variable Name	mh3
Label	Flag: Marital History Event 3
Concept	
Type	numeric
Files	F1 F2

Full Description

Third marital history event flag. Indicates whether respondent entered into a second marriage.

Values (3 total)

0	No second marriage
1	Second marriage occurred
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

Marital History Variables

Variable Name	mh4
Label	Flag: Marital History Event 4
Concept	
Type	numeric
Files	F1 F2

Full Description

Fourth marital history event flag. Indicates whether respondent's second marriage ended in widowhood or divorce.

Values (4 total)

0	Second marriage did not end over course of survey
1	Second marriage ended in widowhood
2	Second marriage ended in divorce/separation
Sysmiss	

Value Ranges

Value Range

Range: [0 , 2]

Groups

Marital History Variables

Variable Name

mh5

Label

Flag: Marital History Event 5

Concept

Type

numeric

Files

F1 F2

Full Description

Fifth marital history event flag. Indicates whether respondent entered into a third marriage.

Values (3 total)

0	No third marriage
1	Third marriage occurred
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

Marital History Variables

Variable Name

mh6

Label

Flag: Marital History Event 6

Concept

Type

numeric

Files

F1 F2

Full Description

Sixth marital history event flag. Indicates whether respondent's third marriage ended in widowhood or divorce.

Values (4 total)

- | | |
|---|--|
| 0 | Third marriage did not end over course of survey |
| 1 | Third marriage ended in widowhood |
| 2 | Third marriage ended in divorce/separation |

Sysmiss

Value Ranges

Value Range

Range: [0 , 2]

Groups

Marital History Variables

Variable Name

mh7

Label

Flag: Marital History Event 7

Concept

Type

numeric

Files

F1 F2

Full Description

Seventh marital history event flag. Indicates whether respondent entered into a fourth marriage.

Values (3 total)

0	No fourth marriage
1	Fourth marriage occurred
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

Marital History Variables

Variable Name	mh8
Label	Flag: Marital History Event 8
Concept	
Type	numeric
Files	F1 F2

Full Description

Eighth marital history event flag. Indicates whether respondent's fourth marriage ended in widowhood or divorce.

Values (4 total)

0	Fourth marriage did not end over course of survey
1	Fourth marriage ended in widowhood
2	Fourth marriage ended in divorce/separation
Sysmiss	

Value Ranges

Value Range

Range: [0 , 2]

Groups

Marital History Variables

Variable Name

mh_date1

Label

Date of Marital History Event 1

Concept

Type

numeric

Files

F1 F2

Full Description

SAS date value

Value Ranges

Value Range

Range: [-12368.9569092027 , 17471]

Groups

Marital History Variables

Variable Name

mh_date2

Label

Date of marital history event 2

Concept

Type

numeric

Files

F1 F2

Full Description

SAS date value

Value Ranges

Value Range

Range: [-10318.8063402205 , 16071]

Groups

Marital History Variables

Variable Name

mh_date3

Label

Date of marital history event 3

Concept

Type

numeric

Files

F1 F2

Full Description

SAS date value

Value Ranges

Value Range

Range: [-10266.2469875312 , 17501]

Groups

Marital History Variables

Variable Name

mh_date4

Label

Date of marital history event 4

Concept

Type

numeric

Files

F1 F2

Full Description

SAS date value

Value Ranges

Value Range

Range: [-7991.34877841126 , 17501]

Groups

Marital History Variables

Variable Name

mh_date5

Label

Date of marital history event 5

Concept

Type

numeric

Files

F1 F2

Full Description

SAS date value

Value Ranges

Value Range

Range: [-5511.43090506919 , 17471]

Groups

Marital History Variables

Variable Name

mh_date6

Label

Date of marital history event 6

Concept

Type

numeric

Files

F1 F2

Full Description

SAS date value

Value Ranges

Value Range

Range: [-1416.79843489124 , 17471]

Groups

Marital History Variables

Variable Name

mh_date7

Label

Date of marital history event 7

Concept

Type

numeric

Files

F1 F2

Full Description

SAS date value

Value Ranges

Value Range

Range: [11323 , 15582.7638571104]

Groups

Marital History Variables

Variable Name

mh_date8

Label

Date of marital history event 8

Concept

Type

numeric

Files

F1 F2

Full Description

SAS date value

Value Ranges

Value Range

Range: [11436.2674992709 , 16010]

Groups

Marital History Variables

Variable Name	personid
Label	Unique person identifier
Concept	
Type	numeric
Files	F1 F2

Full Description

Across the Gold Standard and Completed Data files, personid uniquely identifies SIPP respondents. In the SSB, personid uniquely identifies records within a particular implicate. In order to strengthen confidentiality protection, personid in the SSB does not link records across implicates or to the Gold Standard and Completed Data files.

Value Ranges

Value Range

Range: [1 , 1000]

Groups

Identifiers

Variable Name	spouse_personid
Label	Personid of spouse
Concept	
Type	numeric
Files	F1 F2

Full Description

Personid of linked spouse. Across the Gold Standard and Completed Data files, spouse_personid uniquely identifies spouses of SIPP respondents. In the SSB, spouse_personid uniquely identifies records within a particular implicate. In order to strengthen confidentiality protection, spouse_personid in the SSB does not link records across implicates or to the Gold Standard and Completed Data files. Linked spouse is defined as the first person to whom the SIPP respondent was married during the time period covered by the SIPP panel. Individuals could enter the panel already married and then each would be linked to the other. Individuals could also get married during the course of the panel. If this was the first observed marriage for each member of the couple, they were linked together. Individuals could also get divorced during the course of the panel and then remarry. In many cases, this later marriage caused a new individual to join the panel. This new SIPP respondent would only be linked to his or her spouse if the spouse (and original SIPP sample member) had not already been observed married to someone else. If the original SIPP sample member had been previously linked by marriage to another SIPP sample member, this original link was maintained in spouse_personid. However the marital history reflects the ending of this marriage and the occurrence of the next marriage for the original SIPP sample member. Likewise, the new SIPP sample member who joins through marriage will have that marriage date recorded in his or her marital history but will have a blank spouse_personid. In summary, this variable captures only one marriage partner and does not provide a history of marriage partners even if this history is (partially) observed in the SIPP. The link between SIPP respondents and their spouses has not been perturbed in any way in the SSB. The same individuals will be linked as married partners in the Gold Standard, the Completed Data, and the SSB.

Value Ranges

Value Range

Range: [1167 , 444444]

Groups

Identifiers

Variable Name	male
Label	Male
Concept	
Type	numeric
Files	F1 F2

Full Description

In the 1990-2004 Census-internal SIPP panels, a value for sex is included on each wave file. Thus, there are actually as many sex variables as there are waves of the survey and some changes occur across waves as a result of data collection error. Sex is selected from the array of variables sex1-sex{max number of waves} in which the wave corresponds either to the month in which marital status is first observed (for those without spouses during the course of the SIPP) or to the month in which the respondent's spouse is assigned instead of from a fixed point in the survey. Thus when a spouse is never assigned, an individual's gender comes from the first wave where they report being not married. For individuals who are assigned a spouse, gender comes from the first wave where they reveal their spouse. Finally, an indicator variable for males was created from the categorical sex variable for analytic convenience. This variable is unsynthesized on the SSB and is never missing so there are no imputed values in the Completed Data.

Values (2 total)

0	Female
1	Male

Value Ranges

Value Range

Range: [0 , 1]

Groups

Demographic Variables

Variable Name	birthdate
Label	Date of Birth
Concept	
Type	numeric
Files	F1 F2

Full Description

This variable was taken from a hierarchy of SSA sources instead of the respondent-provided value in the SIPP. Date of birth was selected from the first non-missing value in the following files: (i) SSA's Master Benefits Record (MBR) file, (ii) the Census Bureau's Person Characteristic File (PCF) whose main input is the SSA Numident file, and (iii) SSA's Supplemental Security Record (SSR) file. Thus, this variable is administrative and sometimes differs from the birth date reported in the SIPP survey itself. When missing due to the lack of a validated SSN for the SIPP respondent, date of birth was imputed using date of birth from the Census-internal version of the SIPP. We chose the administrative source for two reasons. First, the administrative birth date was more often consistent with the other administrative data (benefits and earnings). For example, when age was calculated using the administrative birth date, there were fewer individuals who appeared to retire before age 62. Second, the differences between the administrative birth date and the birth date reported in the survey helped to increase the difficulty of re-identifying a record in the original SIPP public use data from a record in the synthetic data, thus improving the confidentiality protections. This variable is coded as a SAS date variable. This format gives the number of days between the date of birth and January 1, 1960. An individual born on January 1, 1959 would have `birthdate=-365` and an individual born on January 1, 1961 would have `birthdate=365`.

Value Ranges

Value Range

Range: [-24204.5838012695 , 10589.4281616211]

Groups

Demographic Variables
Lifespan Variables

Variable Name	race
Label	Race
Concept	
Type	numeric
Files	F1 F2

Full Description

In the 1990-2004 Census-internal SIPP panels, a value for race is included on each wave file. Thus, there are actually as many race variables as there are waves of the survey and some changes occur across waves as a result of data collection error. Race is chosen by creating an array of variables race1-race{max number of waves} and choosing the first non-missing value. Thus race comes from the first wave in which the individual was interviewed instead of from a fixed point in the survey.

Values (3 total)

1	White
2	Black
3	Other

Value Ranges

Value Range

Range: [1 , 3]

Groups

Demographic Variables

Variable Name	hispanic
Label	Hispanic
Concept	
Type	numeric
Files	F1 F2

Full Description

In the 1990-1993 SIPP panels, a value for ethnicity is included on each wave file. Thus, there are actually as many ethnicity variables as there are waves of the survey and some changes occur across waves as a result of data collection error. Ethnicity is chosen by creating an array of variables `ethncty1-ethncty{max number of waves}` and choosing the first non-missing value. Thus, ethnicity comes from the first wave in which the individual was interviewed instead of from a fixed point in the survey. Respondents are coded as Hispanic if they have an ethnicity code between 14 and 20. In the 1996-2004 panels, the longitudinally-edited version contains only one value for ethnicity across all waves (`eorigin`) and this value is used. Respondents are coded as Hispanic if they have an ethnicity code between 20 and 28 in 1996 and 2001, or if they have an ethnicity code of 1 in 2004.

Values (2 total)

0	Non-hispanic
1	Hispanic

Value Ranges

Value Range

Range: [0 , 1]

Groups

Demographic Variables

Variable Name	flag_deathdate_exist
Label	Flag: Existence of Date of Death
Concept	
Type	numeric
Files	F1 F2

Full Description

Flag to indicate that this respondent died after being interviewed and no later than 2006.

Values (2 total)

0	Death date does not exist, respondent did not die during this interval
1	Death date exists, respondent died during this interval

Value Ranges

Value Range

Range: [0 , 1]

Groups

Demographic Variables

Lifespan Variables

Variable Name	deathdate
Label	Date of Death
Concept	
Type	numeric
Files	F1 F2

Full Description

Date of death from administrative data. This variable also is obtained using a hierarchy of administrative sources: (i) SSA's MBR file, (ii) the Census PCF with death information coming from the SSA Numident and Master Death Files, and (iii) SSA's SSR file. This variable is coded as a SAS date variable. This format gives the number of days between the date of birth and January 1, 1960. An individual born on January 1, 1959 would have birthdate=-365 and an individual born on January 1, 1961 would have birthdate=365.

Value Ranges

Value Range

Range: [10958 , 17897]

Groups

Demographic Variables
Lifespan Variables

Variable Name

obs_first_sipp_mar_num

Label

Ordinal Number of First Obs Marriage

Concept

Type

numeric

Files

F1 F2

Full Description

Tells which of the marriages described in the marital history (mh) arrays is the linked marriage.

Values (5 total)

1	Linked marriage is first reported marriage described by mh1, mh2, mh_date1, and mh_date2
2	Linked marriage is first reported marriage described by mh3, mh4, mh_date3, and mh_date4
3	Linked marriage is first reported marriage described by mh5, mh6, mh_date5, and mh_date6
4	Linked marriage is first reported marriage described by mh7, mh8, mh_date7, and mh_date8
Sysmiss	When there is no linked marriage

Value Ranges

Value Range

Range: [1 , 4]

Groups

Marital History Variables

Variable Name	flag_mar4t
Label	Flag: 4 or More Marriages
Concept	
Type	numeric
Files	F1 F2

Full Description

Flag for existence of a marriage for which date is unknown because it was not collected in the SIPP. The marital history topical module asks about a person's first and second marriages and then his or her most recent marriage. If any other marriages occurred after the second but before the most recent, no information about this marriage is collected. However, individuals are categorized as having 1, 2, 3, or more than 3 marriages. We create flag_mar4t to identify individuals who reported more than 3 marriages at the time of the topical module.

Values (3 total)

0	No additional marriage occurred with unknown date
1	An additional marriage occurred but with unknown date
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

Marital History Variables

Variable Name	own_kids_ever
Label	Number of Children Ever Born
Concept	
Type	numeric
Files	F1 F2

Full Description

Number of children ever born. This is taken from the wave 2 Fertility history topical module (TM8752 and TM8754 for 1990-1993 panels; tfrchl and tmomchl for 1996-2004 panels).

Value Ranges

Value Range

Range: [0 , 20]

Groups

Demographic Variables
Fertility Variables

Variable Name	first_birth_year
Label	Year of Birth of First Child
Concept	
Type	numeric
Files	F1 F2

Full Description

This is taken from the wave 2 Fertility history topical module (TM8762 and TM8794 for 1990-1993 panels; tibrthyr for 1996-2004 panels).

Values (1 total)

Sysmiss - Structurally missing (own_kids_ever=0)

Value Ranges

Value Range

Range: [1929 , 2005]

Groups

Demographic Variables
Fertility Variables

Variable Name	last_birth_year
Label	Year of Birth of Last Child
Concept	
Type	numeric
Files	F1 F2

Full Description

This is taken from the wave 2 Fertility history topical module (TM8768 and TM8782 for 1990-1993 panels; tlbirtyr for 1996-2004 panels).

Values (1 total)

Sysmiss - Structurally missing (own_kids_ever=0)

Value Ranges

Value Range

Range: [1929 , 2005]

Groups

Demographic Variables
Fertility Variables

Variable Name	educ_5cat
Label	Education Category (5)
Concept	
Type	numeric
Files	F1 F2

Full Description

Highest level of education attained at the time of the education history topical module. This variable was created from information gathered in the topical module on education history and represents the highest level of education achieved up to the point of the administration of the topical module questions. For individuals who did not answer the topical module education history questions, we imputed values for educ_5cat.

Values (5 total)

1	No high school degree
2	High school degree
3	Some college
4	College degree
5	Graduate degree

Value Ranges

Value Range

Range: [1 , 5]

Groups

Education Variables

Variable Name

current_enroll_hs

Label

Flag currently enrolled in high school

Concept

Type

numeric

Files

F1 F2

Full Description

This variable indicates whether an individual is still enrolled in high school. It can be used to distinguish the difference between people with educ_5cat=1 who have not finished their education and those who have dropped out of high school.

Values (3 total)

0	Not currently enrolled in high school
1	Currently enrolled in high school
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

Education Variables

Variable Name	year_end_hs
Label	Year Ended HS (or less) Education
Concept	
Type	numeric
Files	F1 F2

Full Description

The wave 2 Education and Training History topical module provides knowledge of the year that high school was last attended (variables TM8404 and TM8412 for 1984 and 1990-1993 panels; variables tltschl and thsyr for 1996-2004 panels).

Value Ranges

Value Range

Range: [1914 , 2004]

Groups

Education Variables

Variable Name	year_beg_posths
Label	Year Began Post-HS Education
Concept	
Type	numeric
Files	F1 F2

Full Description

The wave 2 Education and Training History topical module provides knowledge of the year that post-high school education began (variable TM8420 for 1990-1993 panels; variable tcollstr for 1996-2004 panels).

Values (1 total)

Sysmiss - Structurally missing (educ_5cat=1 or educ_5cat=2)

Value Ranges

Value Range

Range: [1922 , 2004]

Groups

Education Variables

Variable Name

current_enroll_coll

Label

Flag currently enrolled in college

Concept

Type

numeric

Files

F1 F2

Full Description

Indicates whether an individual is enrolled in college at time of SIPP education history topical module and has not finished his/her education. This variable can be used to differentiate between individuals who completed some college and stopped school and those who have finished some college but not yet stopped attending school.

Values (3 total)

0 | Not currently enrolled in college

1 | Currently enrolled in college

Sysmiss

Value Ranges

Value Range

Range: [0 , 1]

Groups

Education Variables

Variable Name	year_end_posths
Label	Year Ended Post-HS Education
Concept	
Type	numeric
Files	F1 F2

Full Description

The wave 2 Education and Training History topical module provides knowledge of the year that post-high school education ended (variables TM8426 and TM8440 for 1990-1993 panels; variables tlastcol, tvocyr, tassocyr, tbachyr, and tadvncyr for 1996-2004 panels).

Values (1 total)

Sysmiss - Structurally missing (educ_5cat=1 or educ_5cat=2)

Value Ranges

Value Range

Range: [1926 , 2004]

Groups

Education Variables

Variable Name

year_bach

Label

Year of Bachelor Degree

Concept

Type

numeric

Files

F1 F2

Full Description

year bachelor's degree was finished

Value Ranges

Value Range

Range: [1926 , 2005]

Groups

Education Variables

Variable Name	field_bach
Label	Field of Bachelor Degree
Concept	
Type	numeric
Files	F1 F2

Full Description

Field in which bachelor's degree was obtained. Taken from topical history module on education history. Categories for 1996-2004 panels and for 1990-1993 panels

Values (22 total)

0	Unknown
1	Agriculture/Forestry
2	Art/Architecture (1996-2004 panels); Biology (1990-1993 panels)
3	Business/Management
4	Communications (1996-2004 panels); Economics (1990-1993 panels)
5	Computer and Information Sciences (1996-2004 panels); Education (1990-1993 panels)
6	Education (1996-2004 panels); Engineering (1990-1993 panels)
7	Engineering (1996-2004 panels); English/Journalism (1990-1993 panels)
8	English/Literature (1996-2004 panels); Home Economics (1990-1993 panels)
9	Foreign Language (1996-2004 panels); Law (1990-1993 panels)
10	Health Sciences (1996-2004 panels); Liberal Arts/Humanities (including arts, architecture, music, languages, philosophy) (1990-1993 panels)
11	Liberal Arts/Humanities (1996-2004 panels); Mathematics/Statistics (1990-1993 panels)
12	Mathematics/Statistics (1996-2004 panels); Medicine (1990-1993 panels)

13	Nature Sciences (Biological and Physical) (1996-2004 panels); Nursing, Pharmacy, Health Technologies (1990-1993)
14	Philosophy/Religion/Theology (1996-2004 panels); Physical or Earth Sciences (1990-1993 panels)
15	Pre-Professional (1996-2004 panels); Police Science or Law Enforcement (1990-1993 panels)
16	Psychology
17	Social Sciences/History (1996-2004 panels); Religion/Theology (1990-1993 panels)
18	Other (1996-2004 panels); Social Sciences (1990-1993 panels)
19	Vocational or Technical Studies (1990-1993 panels)
20	Other (1990-1993 panels)

Sysmiss

Value Ranges

Value Range

Range: [0 , 20]

Groups

Education Variables

Variable Name foreign_born

Label Foreign Born

Concept

Type numeric

Files F1 F2

Full Description

Immigrant Status, born in country other than U.S. Taken from wave 2 topical module (TM8730, TM8734, TM8709 1990-1993 panels; eprstate, ebrstate and rcitiznt 1996 panel; eprstate, ebrstate and tcitiznt 2001 panel; eprstate, ebrstate, citiz, and ebornus 2004 panel)

Values (2 total)

0	Born in U.S.
1	Born in country other than U.S.

Value Ranges

Value Range

Range: [0 , 1]

Groups

Demographic Variables

Variable Name	time_arrive_usa
Label	Decade of Arrival to US (Foreign Born)
Concept	
Type	numeric
Files	F1 F2

Full Description

Decade arrive in U.S. (answered when SIPP respondent was foreign_born) The year of arrival to the U.S. is from the Census-internal SIPP files (TM8736 1990-1993 panels; rmoveus 1996 panel; rmoveus 2001-2004 panels) . =Structurally missing, out of scope for question (foreign_born=0)

Values (11 total)

1	Before 1959
2	1960-1964
3	1965-1969
4	1970-1974
5	1975-1979
6	1980-1981
7	1982-1984
8	1985-1993
9	1994-1999
10	2000-2004
Sysmiss	Structurally missing, out of scope for question (foreign_born=0)

Value Ranges

Value Range

Range: [1 , 10]

Groups

Demographic Variables

Variable Name	own_home
Label	Own a Home
Concept	
Type	numeric
Files	F1 F2

Full Description

Indicates whether the individual owned a home in the SIPP panel.

Values (2 total)

0	Do not own a home
1	Own a home

Value Ranges

Value Range

Range: [0 , 1]

Groups

Economic Variables

Variable Name	homeequity
Label	Home Equity
Concept	
Type	numeric
Files	F1 F2

Full Description

Self-reported home equity value

Value Ranges

Value Range

Range: [-194713.396972656 , 652816.126953125]

Groups

Economic Variables

Variable Name	nonhouswealth
Label	Non-Housing Financial Wealth
Concept	
Type	numeric
Files	F1 F2

Full Description

Non-housing wealth = total wealth minus home equity

Value Ranges

Value Range

Range: [-29979.16015625 , 2370592.9140625]

Groups

Economic Variables

Variable Name	totnetworth
Label	Total Net Worth
Concept	
Type	numeric
Files	F1 F2

Full Description

Total net worth

Value Ranges

Value Range

Range: [-200331.076019287 , 2700581.59277344]

Groups

Economic Variables

Variable Name	ind_exist
Label	Flag: Industry Assigned
Concept	
Type	numeric
Files	F1 F2

Full Description

Does person have valid industry from a job held during survey?

Values (2 total)

0	No, last worked 1984 or earlier, or no valid industry reported
1	Yes

Value Ranges

Value Range

Range: [0 , 1]

Groups

Economic Variables

Variable Name

occ_exist

Label

Flag: Occupation Assigned

Concept

Type

numeric

Files

F1 F2

Full Description

Does person have valid occupation from a job held during survey?

Values (2 total)

0 | No, last worked 1984 or earlier, or no valid industry reported

1 | Yes

Value Ranges

Value Range

Range: [0 , 1]

Groups

Economic Variables

Variable Name	ind_4cat
Label	Industry Category (4)
Concept	
Type	numeric
Files	F1 F2

Full Description

Industry is a characteristic of an individual's job and hence varies over time. There are industry values reported for (potentially) two jobs in each wave of the survey. Industry is chosen by summing earnings associated with the array of variables ws1ind1-ws1ind{max number of waves} and ws2ind1-ws2ind{max number of waves} in the 1990-1993 panels, and ejbind1_1-ejbind1_{max number of waves} and ejbind2_1-ejbind2_{max number of waves} in the 1996-2004 panels and choosing the industry associated with the greatest total earnings. Thus industry is the industry from which greatest earnings are derived in the survey.

Values (5 total)

1	Manufacturing
2	Wholesale/retail trade
3	FIRE, services, public administration, military
4	Agriculture, mining, construction, transportation, communications, and public utilities
Sysmiss	

Value Ranges

Value Range

Range: [1 , 4]

Groups

Economic Variables

Variable Name	occ_3cat
Label	Occupation Category (3)
Concept	
Type	numeric
Files	F1 F2

Full Description

Occupation is a characteristic of an individual's job and hence varies over time. There are occupation values reported for (potentially) two jobs in each wave of the survey. Occupation is chosen by summing earnings associated with the array of variables ws1occ1-ws1occ{max number of waves} and ws2occ1-ws2occ{max number of waves} in the 1990-1993 panels, and tjbocc1_1-tjbocc1_{max number of waves} and tjbocc2_1-tjbocc2_{max number of waves} in the 1996-2004 panels and choosing the occupation associated with the greatest total earnings. Thus occupation is the occupation from which greatest earnings are derived in the survey.

Values (4 total)

1	Managerial and professional specialty occupations
2	Technical, sales, and administrative support occupations
3	Other
Sysmiss	

Value Ranges

Value Range

Range: [1 , 3]

Groups

Economic Variables

Variable Name pension_in_scope_empl

Label In-Scope for Pension (Level II)

Concept

Type numeric

Files F1 F2

Full Description

Individual must have been employed at time of pension topical module in order to answer the pension questions.

Values (2 total)

0 Pension was not in scope

1 Pension was in scope

Value Ranges

Value Range

Range: [0 , 1]

Groups

Economic Variables

Variable Name	db_pension
Label	Defined Benefit Pension Plan
Concept	
Type	numeric
Files	F1 F2

Full Description

Defined Benefit Pension Plan

Values (3 total)

0	No defined benefit pension plan
1	Had defined benefit pension plan
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

Economic Variables

Variable Name

dc_pension

Label

Defined Contribution Pension Plan

Concept

Type

numeric

Files

F1 F2

Full Description

Defined Contribution Pension Plan

Values (3 total)

0 No defined contribution pension plan

1 Had defined contribution pension plan

Sysmiss

Value Ranges

Value Range

Range: [0 , 1]

Groups

Economic Variables

Variable Name	sum_disab
Label	Disability (Sum of Core and TM)
Concept	
Type	numeric
Files	F1 F2

Full Description

Health limits kind or amount of work For the 1996-2004 panels, information on work-limiting disability comes from core question (edisabl) during wave 2 for people ages 15-69, when respondents were asked whether health limited the type or amount of work. This indicator is supplemented with details from the Functional Limitations and Disability topical module (wave 5 for 1996-2004 panels, variable ejobdif) that covers people ages 16-67. For the 1990-1993 panels, disability information comes from the core question (disab) during wave 2 for people ages 15-69, when respondents were asked whether health limited the type or amount of work. This information is supplemented with details from the Functional Limitations and Disability topical module (waves 3, 3, 6, 3 for 1990-1993 panels, variables TM8914, TM8918, and TM8920) that covers people ages 16-67. In order to make the sum_disab consistent for all panels, both responses were taken from the core and topical modules, when available, with any positive indication of health limiting the kind or amount of work flagging the positive response. .=Structurally missing, out of scope for question (sum_disab_in_scope=0)

Values (3 total)

0	No (sum_disab_in_scope=1)
1	Yes (sum_disab_in_scope=1)
Sysmiss	Structurally missing, out of scope for question (sum_disab_in_scope=0)

Value Ranges

Value Range

Range: [0 , 1]

Groups

Education Variables

Disability Variables

Variable Name

sum_disab_nowork

Label

Disability Prevents Work (Sum of Core and TM)

Concept

Type

numeric

Files

F1 F2

Full Description

Health prevents work For the 1996-2004 panels, information on work-preventing disability comes from core question edisprev during wave 2 for people ages 15-69, when respondents were asked whether health prevented work. This indicator is supplemented with details from the Functional Limitations and Disability topical module (wave 5 for 1996-2004 panels, variable ejobcant) that covers people ages 16-67. For the 1990-1993 panels, the core questionnaire does not ask respondents whether health prevents work. This information is solely obtained from the Functional Limitations and Disability topical module (waves 3, 3, 6, 3 for 1990-1993 panels, variables TM8922 and TM8924) that covers people ages 16-67. When available, the core and topical modules were used in conjunction to construct summary measures of disability, with any positive indication of health preventing work flagging the positive response. .=Structurally missing, out of scope for question (sum_disab_in_scope=0 or { sum_disab_in_scope=1, sum_disab=0}) 0=No (sum_disab=1) 1=Yes (sum_disab=1)

Values (3 total)

0	No (sum_disab=1)
1	Yes (sum_disab=1)
Sysmiss	Structurally missing, out of scope for question (sum_disab_in_scope=0 or { sum_disab_in_scope=1, sum_disab=0})

Value Ranges

Value Range

Range: [0 , 1]

Groups

Education Variables

Disability Variables

Variable Name	totearn_ser_YYYY
Label	SER: Total Earnings
Concept	
Type	numeric
Files	F1 F2

Full Description

Annual earnings taxed by FICA; these variables include earnings only up to the FICA taxable maximum, i.e., these earnings measures are capped. YYYY=1951-2006

Value Ranges

Value Range

Range: [0 , 94192]

Groups

Summary Earnings Record Variables

Variable Name

wqc_yrtot_YYYY

Label

SER: Annual Total Covered Quarters of Work

Concept

Type

numeric

Files

F1 F2

Full Description

Indicates the total number of quarters of FICA-covered work in year YYYY, where YYYY=1951-2006.

Values (5 total)

0

1

2

3

4

Value Ranges

Value Range

Range: [0 , 4]

Groups

Summary Earnings Record Variables

Variable Name	nondefer_der_fica_YYYY
Label	DER: Non-Deferred FICA
Concept	
Type	numeric
Files	F1 F2

Full Description

Non-deferred earnings paid to an individual from jobs covered by FICA tax; summed across all employers in the DER to give a person-level total for each year YYYY, where YYYY=1978-2006.

Value Ranges

Value Range

Range: [0 , 3372544]

Groups

Detailed Earnings Record Variables

Variable Name	nondefer_der_nonfica_YYYY
Label	DER: Non-Deferred Non-FICA
Concept	
Type	numeric
Files	F1 F2

Full Description

Non-deferred earnings paid to an individual from jobs NOT covered by FICA tax; summed across all employers in the DER to give a person-level total for each year YYYY, where YYYY=1978-2006.

Value Ranges

Value Range

Range: [0 , 524288]

Groups

Detailed Earnings Record Variables

Variable Name defer_der_fica_YYYY

Label DER: Deferred FICA

Concept

Type numeric

Files F1 F2

Full Description

deferred earnings from jobs covered by FICA tax; summed across all employers in the DER to give a person-level total for each year. While the variable exists on the Gold Standard for the years 1978-1986, it is always missing in this time period. The year 1987 is the first year with positive deferred wages. On the synthetic and completed gold standard files, we only keep 1990-2006 because so few people had deferred wages between 1987 and 1989 that we could not reliably synthesize these variables.

Value Ranges

Value Range

Range: [0 , 32112]

Groups

Detailed Earnings Record Variables

Variable Name	defer_der_nonfica_YYYY
Label	DER: Deferred Non-FICA
Concept	
Type	numeric
Files	F1 F2

Full Description

deferred earnings from jobs NOT covered by FICA tax; summed across all employers in the DER to give a person-level total for each year. While the variable exists on the Gold Standard for the years 1978-1986, it is always missing in this time period. The year 1987 is the first year with positive deferred wages. On the synthetic and completed gold standard files, we only keep 1990-2006 because so few people had deferred wages between 1987 and 1989 that we could not reliably synthesize these variables.

Value Ranges

Value Range

Range: [0 , 13902]

Groups

Detailed Earnings Record Variables

Variable Name	flag_in_mbr
Label	Flag: In MBR
Concept	
Type	numeric
Files	F1 F2

Full Description

This flag indicates that a person matched to the SSA Master Beneficiary File (MBR). The person's SSN showed up in the MBR because they received benefits of some kind.

Values (2 total)

0	Respondent was not matched to MBR
1	Respondent was matched to MBR

Value Ranges

Value Range

Range: [0 , 1]

Groups

MBR/PHUS Variables

Variable Name	mbr_retire
Label	MBR: receive retire benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

This variable indicates that a person received retirement benefits at some point during the time period covered by the MBR extract (1964-2007). These benefits were the result of the individual's own earnings history. This variable is only in scope if flag_in_mbr=1.

Values (3 total)

0	Does not receive monthly retire benefit
1	Receives monthly retire benefit
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

MBR/PHUS Variables
Retirement Benefit

Variable Name

mbr_retire_stddate

Label

MBR: startdate of benefit

Concept

Type

numeric

Files

F1 F2

Full Description

Date when the person first began receiving own retirement benefits, conditional on having ever received this type of benefit.

Value Ranges

Value Range

Range: [2435 , 17532]

Groups

MBR/PHUS Variables
Retirement Benefit

Variable Name	mbr_retire_totamt
Label	MBR: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total monthly amount of benefits received at beginning of own retirement benefit entitlement. In most cases this amount is from the same month as in MBR_retire_benefit_stdate. However, if data for that month were missing in the MBR extract, we searched through the monthly benefit array to find the first positive value. This amount can be a combination of payments due to multiple entitlement reasons (i.e. dual entitlement).

Value Ranges

Value Range

Range: [17.8085849001324 , 1641.12913888622]

Groups

MBR/PHUS Variables
Retirement Benefit

Variable Name	phus_retire_stddate
Label	PHUS: startdate of benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Date retirement benefits began being paid, as recorded in the PHUS. This date must be greater than or equal to the MBR retirement benefit start date. It also must be 1984 or later because PHUS data began in 1984.

Value Ranges

Value Range

Range: [8887 , 17501]

Groups

MBR/PHUS Variables
Retirement Benefit

Variable Name	phus_retire_totamt
Label	PHUS: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total amount of benefits as recorded in the PHUS in the first month of receiving own retirement benefits. This amount can be a sum of benefits received for different reasons (i.e. dual entitlement).

Value Ranges

Value Range

Range: [60.0592421000391 , 40436.8216200353]

Groups

MBR/PHUS Variables
Retirement Benefit

Variable Name	mbr_disab
Label	MBR: receive disab benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Indicates that individual received disability benefits at some point over the time period covered by the MBR. This variable is only in scope if flag_in_mbr=1.

Values (3 total)

0	Does not receive monthly disability benefit
1	Receives monthly disability benefit
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

MBR/PHUS Variables
Disability Benefit

Variable Name	mbr_disab_stddate
Label	MBR: startdate of benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Date at which individual began receiving own disability benefits. This date must be before individual reaches the full retirement age (FRA). FRA depends on the year the person reaches age 62. Any individual who turned 62 before 2000 had FRA=65 years old. Beginning in 2000, any individual turning 62 had full retirement age of 65 years $2 \times (\text{year_age_62} - 1999)$ months.

Value Ranges

Value Range

Range: [3319 , 17532]

Groups

MBR/PHUS Variables
Disability Benefit

Variable Name	mbr_disab_totamt
Label	MBR: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total monthly amount of benefits received at beginning of disability benefit entitlement. In most cases this amount is from the same month as in MBR_disab_benefit_stddate. However, if data for that month were missing in the MBR extract, we searched through the monthly benefit array to find the first positive value. This amount can be a combination of payments due to multiple entitlement reason (i.e. dual entitlement).

Value Ranges

Value Range

Range: [44.3054638169562 , 1841.771233258]

Groups

MBR/PHUS Variables
Disability Benefit

Variable Name	phus_disab_stddate
Label	PHUS: startdate of benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Date disability benefits began being paid, as recorded in the PHUS. This date must be greater than or equal to the MBR disability benefit start date. It also must be 1984 or later because PHUS data began in 1984.

Value Ranges

Value Range

Range: [8979 , 17501]

Groups

MBR/PHUS Variables
Disability Benefit

Variable Name	phus_disab_totamt
Label	PHUS: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total amount of benefits as recorded in the PHUS in the first month of receiving own disability benefits. This amount can be a sum of benefits received for different reasons (i.e. dual entitlement).

Value Ranges

Value Range

Range: [46.8844155356545 , 53547.281947684]

Groups

MBR/PHUS Variables
Disability Benefit

Variable Name	mbr_agedsp
Label	MBR: receive agedsp benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

This indicator reports receipt of aged spouse Social Security benefit. This variable is only in scope if flag_in_mbr=1.

Values (3 total)

0	Does not receive monthly agedsp benefit
1	Receives monthly agedsp benefit
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

MBR/PHUS Variables
Aged Spouse Benefit

Variable Name	mbr_agedsp_stddate
Label	MBR: startdate of benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Date when the person first began receiving aged spouse benefits, conditional on having ever received this type of benefit.

Value Ranges

Value Range

Range: [882 , 17532]

Groups

MBR/PHUS Variables
Aged Spouse Benefit

Variable Name	mbr_agedsp_totamt
Label	MBR: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total monthly amount of benefits received at beginning of aged spouse benefit entitlement. In most cases this amount is from the same month as in MBR_agedsp_benefit_stdate. However, if data for that month were missing in the MBR extract, we searched through the monthly benefit array to find the first positive value. This amount can be a combination of payments due to multiple entitlement reasons (i.e. dual entitlement).

Value Ranges

Value Range

Range: [39.5497995976014 , 1048.55813207881]

Groups

MBR/PHUS Variables
Aged Spouse Benefit

Variable Name	phus_agedsp_stddate
Label	PHUS: startdate of benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Date aged spouse benefits began being paid, as recorded in the PHUS. This date must be greater than or equal to the MBR aged spouse benefit start date. It also must be 1984 or later because PHUS data began in 1984.

Value Ranges

Value Range

Range: [8797 , 17348]

Groups

MBR/PHUS Variables
Aged Spouse Benefit

Variable Name	phus_agedsp_totamt
Label	PHUS: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total amount of benefits as recorded in the PHUS in the first month of receiving aged spouse benefits. This amount can be a sum of benefits received for different reasons (i.e. dual entitlement).

Value Ranges

Value Range

Range: [44.1471197253255 , 3491.84037259951]

Groups

MBR/PHUS Variables
Aged Spouse Benefit

Variable Name	mbr_widowsp
Label	MBR: receive widowsp benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

This variable indicates receipt of widowed spouse Social Security benefits at some point in the time period covered by the MBR (1962-2007). This variable is only in scope if flag_in_mbr=1.

Values (3 total)

0	Does not receive monthly widowsp benefit
1	Receives monthly widowsp benefit
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

MBR/PHUS Variables
Widowed Spouse Benefit

Variable Name	mbr_widowsp_stddate
Label	MBR: startdate of benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Date when the person first began receiving widowed spouse benefits, conditional on having ever received this type of benefit.

Value Ranges

Value Range

Range: [1735 , 17532]

Groups

MBR/PHUS Variables
Widowed Spouse Benefit

Variable Name	mbr_widowsp_totamt
Label	MBR: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total monthly amount of benefits received at beginning of widowed spouse benefit entitlement. In most cases this amount is from the same month as in MBR_agedsp_benefit_stddate. However, if data for that month were missing in the MBR extract, we searched through the monthly benefit array to find the first positive value. This amount can be a combination of payments due to multiple entitlement reasons (i.e. dual entitlement).

Value Ranges

Value Range

Range: [48.5188403635409 , 1773.36026566281]

Groups

MBR/PHUS Variables
Widowed Spouse Benefit

Variable Name	phus_widowsp_stddate
Label	PHUS: startdate of benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Date widowed spouse benefits began being paid, as recorded in the PHUS. This date must be greater than or equal to the MBR widowed spouse benefit start date. It also must be 1984 or later because PHUS data began in 1984.

Value Ranges

Value Range

Range: [8826 , 17501]

Groups

MBR/PHUS Variables

Widowed Spouse Benefit

Variable Name	phus_widowsp_totamt
Label	PHUS: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total amount of benefits as recorded in the PHUS in the first month of receiving widowed spouse benefits. This amount can be a sum of benefits received for different reasons (i.e. dual entitlement).

Value Ranges

Value Range

Range: [151.851675741228 , 2559.92913841332]

Groups

MBR/PHUS Variables

Widowed Spouse Benefit

Variable Name	mbr_other
Label	MBR: receive other benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

This variable indicates receipt of Social Security benefits due to one of four reasons: young widow caring for minor children, young spouse caring for minor children, disabled widow, and adult disabled in childhood.

Values (3 total)

0	Does not receive other monthly benefits
1	Receives other monthly benefits
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

MBR/PHUS Variables
Other Benefit

Variable Name	mbr_other_stddate
Label	MBR: startdate of benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Date when the person first began receiving other benefits, conditional on having ever received this type of benefit.

Value Ranges

Value Range

Range: [578 , 23954]

Groups

MBR/PHUS Variables
Other Benefit

Variable Name	mbr_other_totamt
Label	MBR: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total monthly amount of benefits received at beginning of other benefit entitlement. In most cases this amount is from the same month as in MBR_other_benefit_stdate. However, if data for that month were missing in the MBR extract, we searched through the monthly benefit array to find the first positive value. This amount can be a combination of payments due to multiple entitlement reasons (i.e. dual entitlement).

Value Ranges

Value Range

Range: [19.633348689397 , 1424.15337749385]

Groups

MBR/PHUS Variables

Other Benefit

Variable Name	phus_other_stddate
Label	PHUS: startdate of benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Date other benefit began being paid, as recorded in the PHUS. This date must be greater than or equal to the MBR other benefit start date. It also must be 1984 or later because PHUS data began in 1984.

Value Ranges

Value Range

Range: [9101 , 16557]

Groups

MBR/PHUS Variables
Other Benefit

Variable Name	phus_other_totamt
Label	PHUS: total monthly benefit
Concept	
Type	numeric
Files	F1 F2

Full Description

Total amount of benefits as recorded in the PHUS in the first month of receiving other benefits. This amount can be a sum of benefits received for different reasons (i.e. dual entitlement).

Value Ranges

Value Range

Range: [37.1732907512646 , 3241.65236921623]

Groups

MBR/PHUS Variables
Other Benefit

Variable Name	flag_in_ssr
Label	Flag: in SSR
Concept	
Type	numeric
Files	F1 F2

Full Description

This flag indicates that a person's SSN was found in the SSA Supplemental Security Records (SSR). This database tracks people who received SSI.

Values (2 total)

0	SSN not found in SSA Supplemental Security Records (SSR)
1	SSN found in SSR

Value Ranges

Value Range

Range: [0 , 1]

Groups

Supplemental Security Record Variables

Variable Name

ssr_ssi_amt_initial

Label

SSR: SSI Amount - Initial (\$2000)

Concept

Type

numeric

Files

F1 F2

Full Description

amount of monthly SSI payment at time of initial receipt

Value Ranges

Value Range

Range: [0.787527717649937 , 1281.06729125977]

Groups

Supplemental Security Record Variables

Variable Name

ssr_ssi_date_initial_entitle

Label

SAS Date - SSR: SSI Date of Initial Entitlement

Concept

Type

numeric

Files

F1 F2

Full Description

Date of initial entitlement to SSI benefits

Value Ranges

Value Range

Range: [5114 , 19602]

Groups

Supplemental Security Record Variables

Variable Name

hicov_YYYYM

Label

Health Insurance Coverage

Concept

Type

numeric

Files

F1 F2

Values (3 total)

0 | Respondent did not have health insurance coverage during this month

1 | Respondent had health insurance coverage during this month

Sysmiss

Value Ranges

Value Range

Range: [0 , 1]

Groups

SIPP Arrays

Health Insurance Variables

Variable Name

hiemp_YYYYM

Label

Health Insurance Coverage from Employer

Concept

Type

numeric

Files

F1 F2

Values (3 total)

0	Respondent did not have employer-provided health insurance
1	Respondent had employer-provided health insurance
Sysmiss	

Value Ranges

Value Range

Range: [0 , 1]

Groups

SIPP Arrays

Health Insurance Variables

Variable Name	totinc_YYYYM
Label	Total Personal Income
Concept	
Type	numeric
Files	F1 F2

Full Description

monthly personal income summed from all sources

Values (1 total)

Sysmiss

Groups

SIPP Arrays
Income Variables

Variable Name

tothours_YYYYM

Label

Total Hours Worked at All Jobs

Concept

Type

numeric

Files

F1 F2

Full Description

Total number of hours worked at all jobs in a given month

Values (1 total)

Sysmiss

Groups

SIPP Arrays

Labor Force Variables

Variable Name wksjob_YYYYM

Label Weeks at a Job

Concept

Type numeric

Files F1 F2

Full Description

Total number of weeks worked at a job in a given month

Values (1 total)

Sysmiss

Groups

SIPP Arrays

Labor Force Variables

Variable Name wkswp_YYYYM

Label Weeks With Pay

Concept

Type numeric

Files F1 F2

Full Description

Total number of weeks worked with pay in a month. Weeks worked with pay = weeks worked - weeks worked without pay;

Values (1 total)

Sysmiss

Groups

SIPP Arrays

Labor Force Variables

Variable Name	totearn_YYYYM
Label	Total SIPP Earnings
Concept	
Type	numeric
Files	F1 F2

Full Description

This variable is taken from the recoded public-use variable totearn - total person monthly earnings from all sources.

Values (1 total)

Sysmiss

Groups

SIPP Arrays

Income Variables

Variable Name

state

Label

State of Residence: FIPS code (modified)

Concept

Type

Files

Full Description

State of residence. FIPS State Code for state of residence first recorded in the SIPP. For married couples, we take the state value for both partners at the same point in the survey when we first observed the marriage. For individuals who never have an observed marriage during the panel, we take their first ever reported state value. *All panels prior to 2004 group some states together and give only one code for the group. For these panels, the individual FIPS code will not appear for states contained in a group.

Values (56 total)

1	Alabama
2	Alaska *see description
4	Arizona
5	Arkansas
6	California
8	Colorado
9	Connecticut
10	Delaware
11	DC
12	Florida
13	Georgia
15	Hawaii

- 16 Idaho *see description
- 17 Illinois
- 18 Indiana
- 19 Iowa *see description
- 20 Kansas
- 21 Kentucky
- 22 Louisiana
- 23 Maine *see description
- 24 Maryland
- 25 Massachusetts
- 26 Michigan
- 27 Minnesota
- 28 Mississippi *see description
- 29 Missouri
- 30 Montana *see description
- 31 Nebraska
- 32 Nevada
- 33 New Hampshire
- 34 New Jersey
- 35 New Mexico *see description
- 36 New York
- 37 North Carolina
- 38 North Dakota *see description
- 39 Ohio
- 40 Oklahoma
- 41 Oregon
- 42 Pennsylvania

44	Rhode Island
45	South Carolina
46	South Dakota *see description
47	Tennessee
48	Texas
49	Utah
50	Vermont *see description
51	Virginia
53	Washington
54	West Virginia *see description
55	Wisconsin
56	Wyoming *see description
61	(1990, 1991, 1992, 1993, 1996, 2001 panels) Maine, Vermont *see description
62	(1990, 1991, 1992, 1993 panels) Iowa, North Dakota, South Dakota (1996, 2001 panels) North Dakota, South Dakota, Wyoming *see description
63	(1990, 1991, 1992, 1993 panels) Alaska, Idaho, Montana, Wyoming *see description
90	(1984 panel only) Idaho, New Mexico, South Dakota, Wyoming *see description
91	(1984 panel only) Mississippi, West Virginia *see description

Groups

Geographic Variables

