

Consumer Expenditure Surveys Public-Use Microdata (PUMD) Overview

Scott Curtin

Division of Consumer Expenditure Surveys
CE Microdata User's Workshop
July 13-15, 2016



Follow Along!

<http://www.bls.gov/cex/pumd.htm>

Quick Links

- » [Introduction to PUMD](#)
- » [Data files](#)
- » [Documentation](#)
- » [Supplemental resources](#)
- » [Public-Use Microdata Annual Users' Workshop](#) provides users the opportunity to learn how to use the PUMD and to present how they used PUMD in their research.
- » The next workshop is July 12-15, 2016. Interested? Register [here](#).



PUMD Data Package

■ Data files:

▶ File types available include:

SAS (*.sas7bdat), STATA (*.dta), SPSS (*.sav), comma-delimited (*.csv)

■ Documentation:

▶ Documents covering specific years (changes with each new PUMD release)

▶ Documents covering all years (Static or evolving documentation)

■ Supplementary resources:

▶ Provides further information regarding PUMD methods and requirements

Data Files

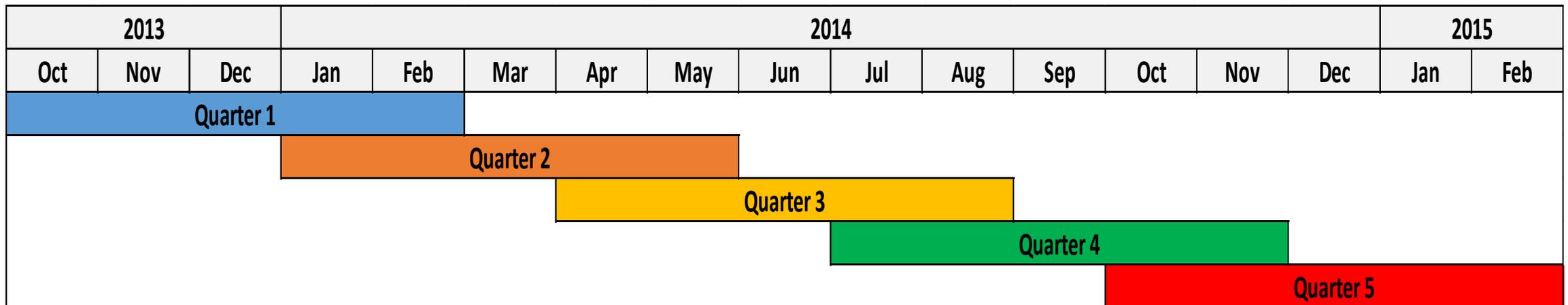


Data Files

- PUMD files contain respondent data collected in either the *Interview* or *Diary* survey.
- These data can be used to create and correlate information in a variety of ways:
 - ▶ Descriptive Statistics
 - How many CU's in the sample purchased a new car in 2014?
 - ▶ Inferential Statistics
 - What is the average annual expenditure on eggs by all CU's in the northeast region?

Interview Survey Data Files

- Each release contains 5 quarters of collected respondent data
- These 5 quarterly datasets contain respondent data that collectively spans across 17 months.



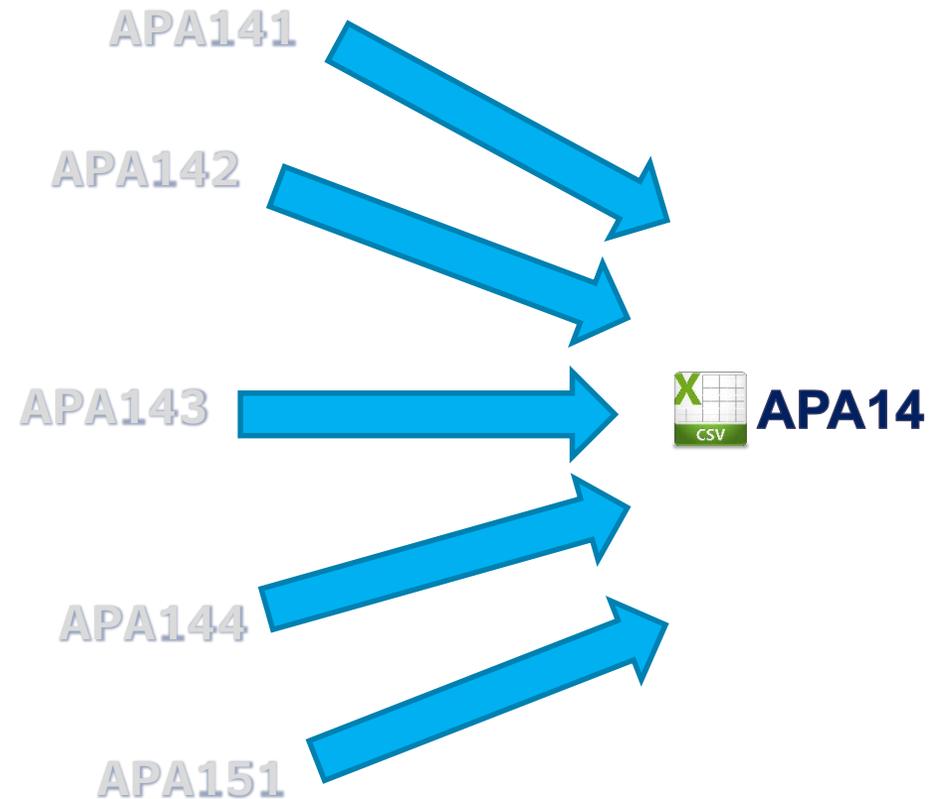
Interview Survey Data Files

- Quarterly datasets (5 files each)
 - FMLI – CU characteristics, incl. summary expenditures
 - MEMI – Member characteristics
 - MTBI – Monthly expenditures
 - ITBI - Monthly Income
 - ITII – Monthly Income (Imputation iterations included)
 - NTAXI - Federal and state tax information
- Concatenated 5 quarter dataset (1 file each)
 - EXPN Files (APA, APB) – structured to closely resemble the way questions are ordered in the Computer Assisted Personal Interview (CAPI) Survey questionnaire.
- Concatenated 9 quarter dataset (1 file each)
 - FPAR – CU-level paradata – interview outcome
 - MCHI – Contact history

Quarterly



Concatenated



Diary Data Files

- Each release contains 4 quarters of collected respondent data
- These 4 quarterly datasets will contain respondent data that spans across 12 months.

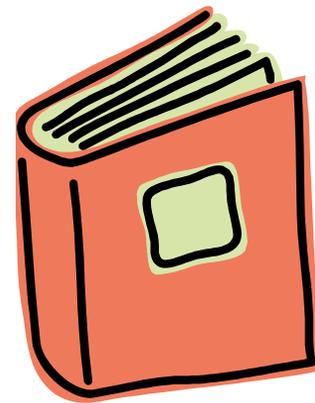
2014											
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Quarter 1			Quarter 2			Quarter 3			Quarter 4		



Diary Survey Data Files

- Quarterly datasets (4 files each)
 - FMLD – CU characteristics, income, weights, and summary level expenditures
 - MEMD – CU member characteristics and income
 - EXPD – Detailed weekly expenditures categorized by UCC
 - DTBD – Annual Income
 - DTID – Annual Income (Imputation iterations included)
- No paradata for Diary
- No EXPN Files (APA, APB)

Documentation



Documentation Covering All Years

- **Source selection file (XLS)**

Identifies which survey data item (category) comes from when combining the two CE surveys for 1996 forward.

- **Response rates (PDF)**

Provides information on the response rates for the Interview and Diary survey for 2000 forward.

- **Description of income imputation**

Provides information on the methods BLS uses to estimate income since 2004.

Documentation Covering Specific Years

- Users' manual
- Data dictionaries
- Hierarchical groupings
- Sample code
- Resources
- Errata

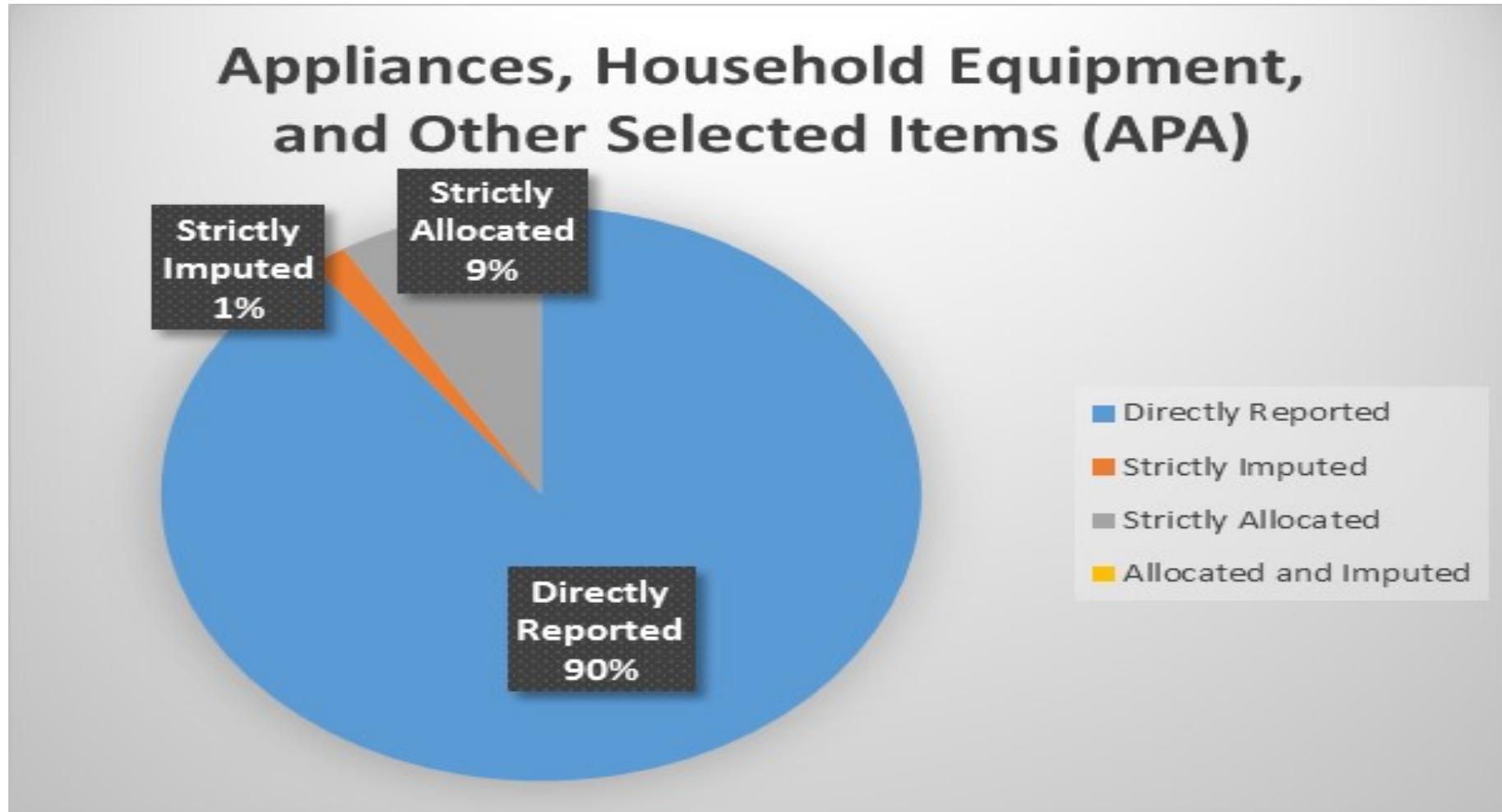


Users' Manual

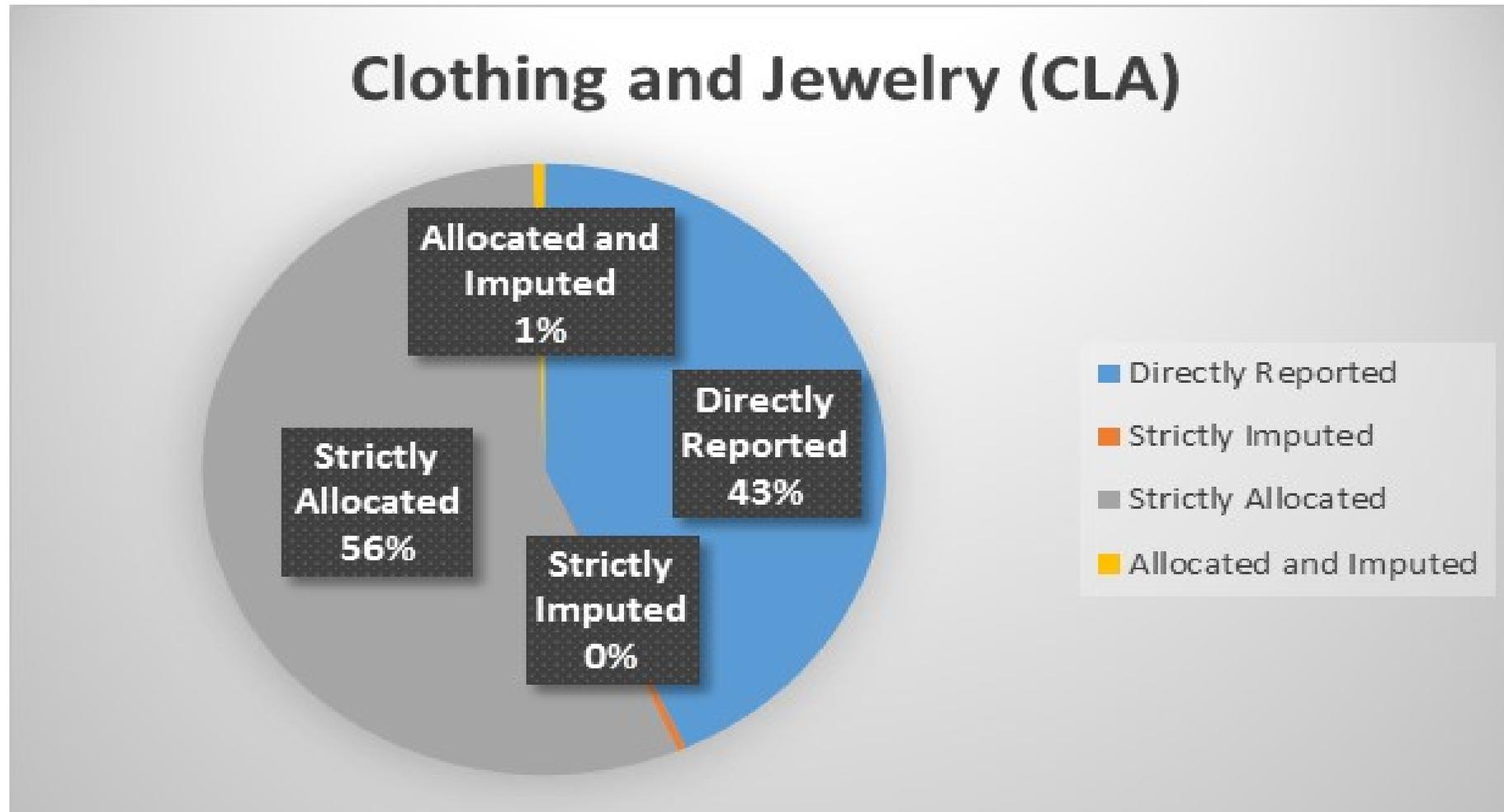
- Describes the content and methods for the data collected within the Interview and Diary Survey
- Includes definitions of our flags
 - Valid? Allocated? Imputed? Topcoded? Suppressed? A little of each?
- Estimation procedures/formulas
 - How to produce a collection year estimate
 - How to produce a calendar year estimate
- Information on changes to the data.
 - Variable changes
 - Questionnaire/Decennial Census sample redesign



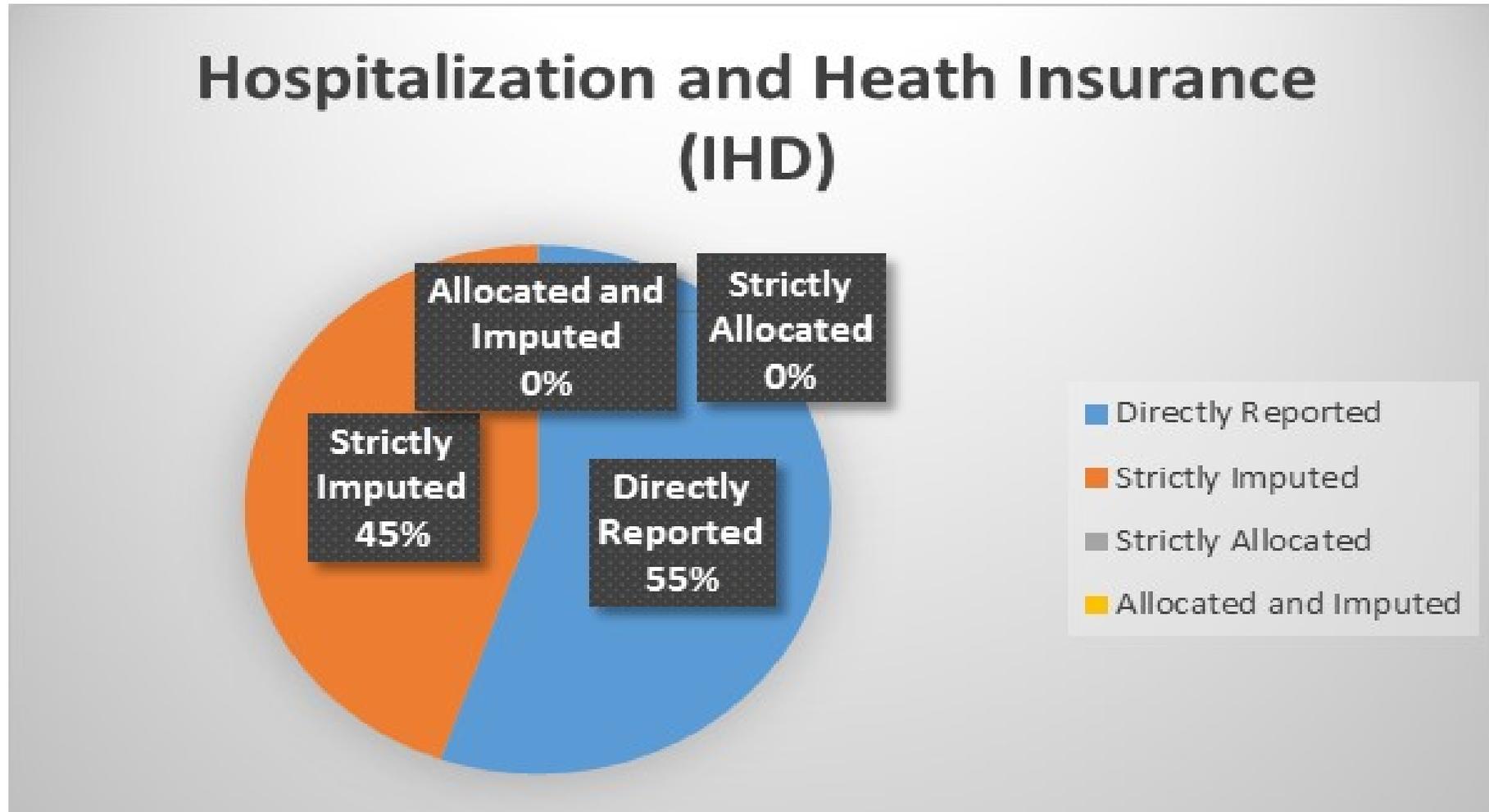
Percentage of records reported directly, imputed, allocated, or imputed and allocated by record type, Interview Survey, 2014 Q1-2015 Q1



Percentage of records reported directly, imputed, allocated, or imputed and allocated by record type, Interview Survey, 2014 Q1-2015 Q1



Percentage of records reported directly, imputed, allocated, or imputed and allocated by record type, Interview Survey, 2014 Q1-2015 Q1



Data Dictionaries

Click on Tools to convert files to PDF.

Interview Data Dictionary

FMLI – Consumer Unit (CU) Characteristics And Income File *CU and Interview Identifiers*

VARIABLE NAME	DESCRIPTION	FLAG	FORMAT	NOTE
NEWID	<p>CU identification number. Digits 1-7 (CU sequence number, 1 through 9999999) uniquely identify the CU. Digit 8 is the interview number, 2 through 5.</p> <p>It is possible for a CU to skip an interview. For example, a CU could have a 2nd, 3rd and 5th interview but no 4th interview.</p> <p>Values of NEWID contain a leading zero. Therefore it will appear the NEWIDs are 7 numbers long, when they are in fact 8 numbers.</p> <p>BLS derived</p>		NUM(8)	
QINTRVMO	<p>Interview month</p> <p>Control Card 37</p>		CHAR(2)	
QINTRVYR	<p>Interview year</p> <p>Control Card 37</p>		CHAR(4)	
HH_CU_Q	<p>Count of CUs in household</p> <p>BLS derived</p>	HH_CU_Q_	NUM(2)	
HHID	<p>Identifier of household with more than one CU. Household with only one CU will be set to missing.</p> <p>BLS derived</p>	HHID_	NUM(3)	



Data Dictionaries

- List the variables, definitions, data formats, and any changes.
- Changes:
 - ▶ N(Yyyq) – indicates new variable
 - ▶ D(Yyyq) – indicates deleted variable
 - ▶ C(Yyyq) – indicates a change within the variable



Data Dictionaries

- Access database
 - ▶ Search by variable name
 - ▶ Search by RECTYPE
 - ▶ Search by keyword in description



Hierarchical Groupings

- Establishes a relation between the published tables' series and their contributing expenditure classification codes (UCC)
- Contains information on table aggregation, source selection, and UCC labels
- Three separate stub files are available
 - Diary Stub: DStubYYYY.txt
 - Interview Stub: IStubYYYY.txt
 - Integrated Stub: IntStubYYYY.txt



Hierarchical Groupings

IntStub2014 - Notepad				
File	Edit	Format	View	Help
* *	Integrated stub parameter file	HEADINTG	H	1 CUCHARS
1 1	Number of consumer units (in thousands)	CONSUMIT	S	1 CUCHARS
* *	Lower limit	QUINTLIM	S	1 CUCHARS
* *	Percent distribution of consumer units	CUDISTRB	S	1 CUCHARS
* *	Consumer unit characteristics:	TITLECU	T	1 CUCHARS
1 2	Income before taxes	INCBFTAX	G	1 CUCHARS
1 3	Meals as pay	800700	I	1 CUCHARS
1 3	Rent as pay	800710	I	1 CUCHARS
1 3	Income before taxes	980000	I	1 CUCHARS
1 2	Income after taxes	INCAFTAX	G	1 CUCHARS
1 3	Meals as pay	800700	I	1 CUCHARS
1 3	Rent as pay	800710	I	1 CUCHARS
1 3	Income after taxes	980071	I	1 CUCHARS
1 1	Age of reference person	980020	I	1 CUCHARS
* *	Average number in consumer unit:	TITLEACU	T	1 CUCHARS
1 1	People	980010	I	1 CUCHARS
1 1	Children under 18	980050	I	1 CUCHARS
1 1	Adults 65 and older	980060	I	1 CUCHARS
1 1	Earners	980030	I	1 CUCHARS
1 1	Vehicles	VEHICLES	G	1 CUCHARS
1 4	Vehicles (owned)	980040	I	1 CUCHARS
1 4	Vehicles (leased)	980360	I	1 CUCHARS
* *	Percent distribution:	TITLEPD	T	1 CUCHARS
* *	Reference person:	TITLESRP	T	1 CUCHARS
1 1	Men	980210	I	1 CUCHARS
1 1	Women	980220	I	1 CUCHARS
* *	Housing tenure:	TITLEHT	T	1 CUCHARS
1 1	Homeowner	HOMEOWN	G	1 CUCHARS
1 4	with mortgage	980230	I	1 CUCHARS
1 4	without mortgage	980240	I	1 CUCHARS
1 1	Renter	980260	I	1 CUCHARS
* *	Race of reference person:	TITLERRP	T	1 CUCHARS
1 1	Black or African-American	980270	I	1 CUCHARS
1 1	white, Asian, and all other races	WHTNDOTH	G	1 CUCHARS
1 4	white	980281	I	1 CUCHARS
1 4	Asian	980282	I	1 CUCHARS
1 4	All other races	980283	I	1 CUCHARS
* *	Hispanic or Latino origin of reference person:	TITLEHOP	T	1 CUCHARS
1 1	Hispanic or Latino	980285	I	1 CUCHARS
1 1	Not Hispanic or Latino	980286	I	1 CUCHARS
* *	Education of reference person:	TITLEEDU	T	1 CUCHARS
1 1	Elementary (1-8)	980290	I	1 CUCHARS
1 1	High school (9-12)	980300	I	1 CUCHARS
1 1	College	980310	I	1 CUCHARS
1 1	Never attended and other	980320	I	1 CUCHARS
1 1	At least one vehicle owned or leased	980350	I	1 CUCHARS
1 1	At least one vehicle owned	980330	I	1 CUCHARS
1 1	At least one vehicle leased	980340	I	1 CUCHARS
1 1	Average annual expenditures	TOTALEXP	G	1 EXPEND
1 2	Food	FOODTOTL	G	1 EXPEND
1 3	Food at home	FOODHOME	G	1 FOOD



Sample Programs

■ SAS :

- ▶ Means and SE (Interview, Diary, and Integrated)
 - Goal is to replicate publication totals
- ▶ CE Macros (including documentation)

■ R:

- ▶ Means and SE (Interview, Diary, and Integrated)
 - Goal is to replicate publication totals

Resources

▶ Vehicle File (CAPIVEH)

- Crosswalk for vehicle codes found in Interview survey EXPN files VEH and LSD.

ACU	ACURA
ALF	ALFA ROMEO
AST	ASTON MARTIN
AUD	AUDI
AUS	AUSTIN-HEALEY
BEN	BENTLEY
BMW	BMW
BUI	BUICK
CAD	CADILLAC
CHE	CHEVROLET
CHR	CHRYSLER
DAW	DAEWOO
DOD	DODGE
EAG	EAGLE
FER	FERRARI
FIA	FIAT
FOR	FORD
GEO	GEO
GMC	GMC
HON	HONDA
HUM	HUMMER

Errata

- Provides information on errors and their correction to the files of a particular year.



Supplementary resources



Supplementary resources

- [CE Glossary](#)
- [CE FAQ](#)
- [User's guide to income imputation \(PDF\)](#)
- [Protection of respondent confidentiality](#)



Scott Curtin
CE PUMD Section Chief
Division of Consumer Expenditure Surveys
202-691-6233
Curtin.Scott@bls.gov

