Analysis of the Efficacy of the SIPP-EHC's Type2 Question Battery
Finalized March 13, 2013

Lindsay M. Monte and Ashley Edwards
U.S. Census Bureau
Social Economic and Housing Statistics Division[1]

(Originally prepared as a poster for 2012 APPAM Fall Conference, titled, "Income Estimation for Former Household Members in the Redesigned SIPP: A Methodological Analysis of Instrument Efficacy"[2]*)*

Background

The Survey of Income and Program Participation (SIPP) is a nationally representative, longitudinal household survey conducted by the U.S. Census Bureau. The SIPP provides monthly data on the family composition and economic well-being of American households, and has been conducted by the Census Bureau since 1984. Under the current SIPP design, survey respondents are followed over the course of a two to four year panel, and interviews occur every four months.

However, in 2006, Congress mandated a reengineering of the SIPP in an effort to make the survey more timely, while also reducing administrative costs and respondent burden (National Research Council, 2009). In complying with this directive, the Census Bureau has worked to redesign the SIPP, most notably by changing the personal interview schedule from every four months to an annual interview. To facilitate this change, the redesigned SIPP instrument utilizes an Event History Calendar, which encourages respondents to recall events in the reference year in relation to other events, and serves as a memory aid for respondents over the longer recall period. The redesigned survey, known as the Survey of Income and Program Participation – Event History Calendar, or the SIPP-EHC, has undergone a number of preliminary field tests, and is expected to replace the SIPP survey as the production survey instrument in calendar year 2014.

Research Question

One of the challenges the Census Bureau faces in transitioning to annual interviews in the SIPP-EHC is the question of how to capture data on individuals who were part of the household at some point in the reference year, but who left the household prior to the interview. The data about these persons are important in order to understand the dynamics of household composition and income changes over the course of a calendar year, which is a central goal of the SIPP.

Under the current SIPP interview schedule, short-term residents are more likely to be rostered at one of the three data collection points throughout the year.[3] In comparison, a single annual interview will only

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on methodological or operational issues are those of the authors and are not necessarily those of the U.S. Census Bureau. Any error or omissions are the sole responsibility of the authors. Email: lindsay.m.monte@census.gov and ashley.edwards@census.gov

[2] The poster is included as a graphic at the end of this write-up.

capture data on individuals if they are living in the household at the time of the interview, which can occur anywhere from December of the reference period (the year of interest) to several months after the end of the reference period. Therefore, in the SIPP-EHC, if an individual lived in a sampled household with survey respondents at any time during the reference period, but no longer lives at the sample address at the time of the interview (and was not an original sample person), they will not be interviewed under the new survey design. These individuals are referred to as "Type2" people in the SIPP-EHC. [4]

Despite their status as former members of SIPP-EHC households, these "Type2" individuals are important for recording household composition, calculating accurate monthly income, and compiling estimates of eligibility for, and receipt of, social welfare programs and private benefits such as health insurance. Therefore, in an effort to ensure that we are capturing at least some information about these former household members, researchers at the Census Bureau added a battery of questions to the SIPP-EHC instrument (see Figure 1 for a list of the questions asked). The goal of this Type2 question battery is to collect basic demographic information, such as age and sex, while also better understanding how these Type2 individuals participate in the household. The survey asks respondents to report basic information for the Type2 individuals with whom they had previously lived. For example, the Type 2 question battery establishes the relationships between Type2 persons and interviewed household members, and collects fundamental economic information for Type2 individuals via questions about employment status and annual income.

In this analysis, we test the efficacy of the SIPP-EHC with regard to the capture of information about these Type2 individuals, when compared to information about similar individuals in current production SIPP. Specifically, we test whether the information we collect in the Type2 question battery matches the demographic and economic characteristics of similar persons in the SIPP. (Due to time constraints, we do not address the question of whether the information collected about Type2 individuals, if proximate, is sufficient for its intended purposes, although this is a question that we believe merits further examination; see Conclusions and Implications for further discussion.)

In order to conduct this analysis, we compared the information collected through the battery of Type2 questions in the 2010 Wave 1 SIPP-EHC[5] field test to information gathered via full interviews in the 2008 SIPP Panel for a sample of individuals who would have been considered "Type2" individuals under the SIPP-EHC design.

---

[3] Of course, even with a four-month reference period, some short-stay individuals will inevitably be missed. That is, although the SIPP never asked about such persons, we cannot dismiss the likely presence of some individuals who move in with a respondent after one interview, stay for at least a month, but move out before the next interview. However, such individuals are neither observed nor referenced in the SIPP instrument.

[4] Type2s should not be confused with Movers. Movers, under both SIPP designs, are original sample persons – that is, individuals residing at the residence at the time of the first interview – who move from the original address. Movers are followed to their new address and interviewed there, along with anyone else who lives with them. Type2 persons in the SIPP-EHC design never enter the sample; instead, they are the individuals who lived with respondents during the course of the reference year, but who are not eligible to be interviewed.

[5] The 2010 SIPP-EHC references calendar year 2009.

In comparing data across the SIPP and SIPP-EHC surveys, we aspired to answer the following basic questions:

1. How well does the SIPP-EHC capture Type2 people when compared to SIPP data?
    a. Do we capture the number of Type2s in the SIPP-EHC that we would expect to, based on comparable samples in the SIPP?
    b. Are the Type2 people captured in the SIPP-EHC demographically and economically similar to similar individuals captured in the SIPP?
2. Are we accurately measuring annual income amounts for Type2 people in the SIPP-EHC?
    a. Are there ways that we could improve our battery of Type2 questions in the SIPP-EHC to improve income imputation, and better understand household financial well-being?

Methods[6]

The 2010 SIPP-EHC field test collected proxy data on 714 Type2 individuals. Our comparison sample in the 2008 SIPP consisted of 455 individuals who would have been considered Type2 persons under the SIPP-EHC survey design, referred to here as "*Pseudo-Type2"* persons. Both SIPP-EHC and SIPP samples consisted of individuals who met the following conditions:

- did not live at the sampled address at the time of the Q1 2010 interview
- lived at some point between January and December 2009 with someone who remained residing at the sampled address as of the Q1 2010 interview.

In order to make non-biased comparisons across the SIPP-EHC and SIPP samples, the SIPP sample was subset to match the 2009 calendar year, as well as the sampling characteristics and geography of the SIPP-EHC sample.[7] Pseudo-Type2 individuals in the 2008 SIPP were captured in at least one of the three SIPP interviews in calendar year 2009, and, as opposed to the limited proxy Type2 data captured in the SIPP-EHC, Psuedo-Type2 individuals answered the full survey instrument (via self or proxy report), providing a more complete picture of the demographic and economic characteristics of these transitional household members.

Our analysis was conducted in two parts. First, we compared rates, as well as demographic and economic characteristics, across Type2 and Psuedo-Type2 individuals by survey instrument. This was done to assess whether the SIPP-EHC instrument is capturing the same number and type of individuals as those captured through the 2008 SIPP. Chi-square and t-tests were utilized to assess the significance of differences.

[6] All data are subject to error arising from a variety of sources, including sampling error, non-sampling error, model error, and any other sources of error. For further information on SIPP statistical standards and accuracy, go to http://www.census.gov/sipp/source.html.

[7] The SIPP and SIPP-EHC samples were drawn from the same sampling frame. However, the SIPP-EHC was limited to low-income strata in only a set number of states. In order to make the SIPP and SIPP-EHC samples geographically and economically comparable, the SIPP sample was therefore limited to households drawn from low-income strata in the same states used for the SIPP-EHC sample. Additionally, the 2010 SIPP-EHC, administered in the early months of 2010, referenced the calendar year of 2009. In order to have the two datasets reference the same time period, the SIPP data was limited to the calendar year of 2009, which generally corresponded to Waves 2, 3, 4, and 5 of the 2008 panel.

Second, we tested whether the income data reported for Type2s in the SIPP-EHC were consistent with income reported by similar Pseudo Type2s. Initially, we compared means using reported data. However, given high rates of missing data, we also generated income amounts for individuals age 15 and older using repeated specifications of a Bayesian Bootstrap Multiple Imputation model, and compared the means to those reported by the Pseudo Type2 sample. Our repeated specifications included imputing income for all Type2s, imputing only for those with missing data, imputing only for those with income responses, and imputing missing responses while retaining reported income.

For each of these specifications, we ran two independent imputation models on the SIPP-EHC Type2 data. In our first model, we used only information available via survey questions in both samples: sex, age, education and employment during the reference year. In the second model, we attempt to replicate standard income imputation processes used in the 2008 SIPP. That is, income imputation models used by the Census Bureau generally include not only sex, age, education, and work, but also measures of race and marital status, neither of which were asked about in the Type2 question battery in the SIPP-EHC. Therefore, we derived this additional demographic information for our SIPP-EHC sample, creating binary (Black/Non-Black) race and marital status (Married/Not married) variables.

To derive race, we first checked to see whether anyone in the sampled household identified themselves as a blood relative of the Type-2 person (i.e., sibling, parent, child); if so, we assigned that relative's race to the Type2 person (N=243). However, if there were no blood relative in the household, we made an assumption of racial homogamy (Lamanna & Riedmann, 2008), and assigned the Type2 person the modal race of the household (N=462). In the handful of cases in which the household was evenly split between Black and non-Black residents, we assigned the Type2 person as Black[8] (N=9).

For marital status, we first checked to see whether anyone in the household identified the Type2 person as a spouse (N=54), and if so, assigned that Type2 person as married. We further made the logical assumption that if two opposite sex Type-2 persons of proximate age moved into and out of the household at the same time that they were likely spouses, and so assigned both Type2s as married (Strong, DeVault, & Cohen, 2010; N=71). In all other cases, marital status was assigned as not married.

We ran four implicates of each model and averaged the variances to generate to a single mean value for imputed annual income. We then compared the average of each set of implicates to the mean incomes reported for the sample of Pseudo Type2s in the production SIPP data.

Results[9]

Table 1 shows that the SIPP-EHC captures fewer Type2s overall than does the SIPP, and our data further suggest that under-reporting may be particularly evident in the early months of the reference year (see both Table 2 and Figure 2). However, Table 2 shows that the SIPP-EHC captures more short stay Type2s, defined as those residing with the respondents between one to four months during the reference

[8] This likely dampens the significance of race, as we undoubtedly assign some non-black individuals as black under this rule. However, given the small sample size, we believe the likely effect to be minimal.

[9] All comparative statements in this report have undergone statistical testing, and, unless otherwise noted, are statistically significant at the 95 percent confidence level.

period. This suggests that our speculation that the SIPP missed short-stay household members is likely accurate.

In looking at the descriptive characteristics of Type2s across survey instruments, Table 1 also shows that Type2s in the SIPP-EHC are much more likely to be non-relatives of respondents than are Pseudo-Type2s in the SIPP. They are additionally distinct in a number of other ways - including by race, sex, age, and education – suggesting that the SIPP-EHC is not necessarily capturing the same individuals as the SIPP.

However, despite these differences in the demographic characteristics of Type2 and Pseudo Type2s captured across surveys instruments, we do not find significant differences in mean reported income across surveys (see Table 3). There are some differences in income distributions across surveys, with Type2s being less likely than Psuedo-Type2s to have reported incomes between $25,000 and $50,000, while also being more likely than Psuedo-Type2s to have reported annual incomes greater than $50,000 (see Table 3). Nonetheless, when taken in the aggregate, mean reported income is not significantly different for the two samples (see Table 3).

This comparison relies on only reported income, however; for approximately 40% of the SIPP-EHC Type2s, we did not obtain a useable income response (see Figure 3). Moreover, we find some significant demographic differences between Type2s for whom we have reports of income and those for whom we do not (see Figure 4).

Therefore, we employ two different multiple imputation specifications to arrive at income estimates for all members of the Type2 sample. Nonetheless, regardless of how we consider non-response in our imputations, we do not find any significant differences in mean imputed income; either when compared to the SIPP mean or the mean of SIPP-EHC Type2 reports (see Model 1, Table 4). We further find that the addition of derived demographic information does not change our income imputations (see Model 2, Table 4)

Conclusions and Implications

The goal of this research was to assess the efficacy and accuracy of the Type2 question battery included in the re-engineered SIPP instrument, the SIPP-EHC. Given concerns that a single annual interview would not capture data from former household members no longer present at time of interview, the Type2 question battery was added in order to gain a more complete compositional and financial portrait of subjects' households. In this analysis, we have compared the data collected via this question battery in the 2010 SIPP-EHC to full survey data from a matched sample of respondents from the 2008 SIPP to see whether the addition of this small block of questions has been successful in garnering the Census Bureau data proximate to that captured by the SIPP.

These results suggest that, for the Type2s captured in the SIPP-EHC, the current question battery may be sufficient; that is, these results do not suggest the need to add questions about race or marital status to the Type2 question battery in the SIPP-EHC. Moreover, the follow-up income questions in instances of a "don't know" response, or a refusal to answer the open-ended income question (see questions 7a-7c,

Figure 1), may not be necessary as the demographic information currently collected appears to be sufficient to impute income in these instances.

However, additional efforts may be needed to capture Type2s whose co-residence was early in the calendar year (see Table 2 and Figure 2), and additional research is needed to understand differential non-response. That is, we do not find evidence for income differences between the samples at a macro level, post-imputation, but the SIPP-EHC records fewer Type2s than would be expected based on SIPP data, and the income information for those Type2s reported does not appear to be missing at random. The apparent underreporting of Type2s should also be examined, and more work will be needed to tease out whether differential non-response for demographically different Type2s affects our results.

Additionally, the natural question that flows from this analysis remains unanswered; we have not assessed whether the available Type2 information is sufficient for its intended purposes. Although we have found income estimates to be proximate, the Type2 question battery does not include any questions about how resources are shared between the Type2 person and the respondents with whom they lived.[10] Moreover, income data collected about Type2 individuals is only reported at an annual level. Thus, these data provides limited insight into how household income and poverty status change, month by month, as these Type2 individuals move into and out of sampled households. Particularly in light of the high number of non-relative Type2s reported in the SIPP-EHC, assumptions about how resources are shared seem problematic (see Table 1). If the intention is to use the Type2 information to round out our understanding of the finances of SIPP-EHC respondents, we feel that additional consideration should be given to the question of whether the information currently being gathered will be sufficient.

[10] Such as whether respondents shared meals with the Type2 person or if that Type2 person contributed to the rent or mortgage.

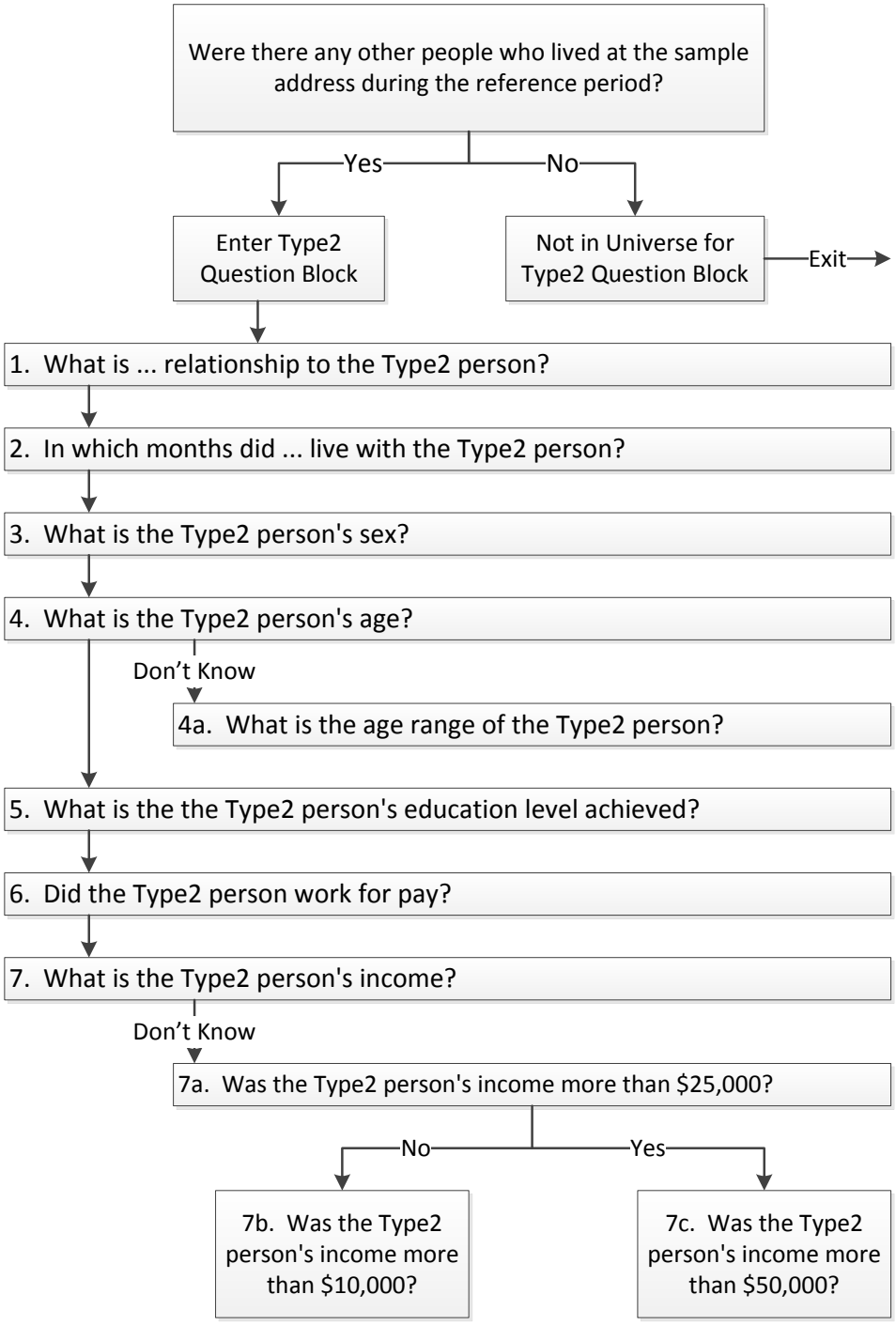
Works Cited

Lamanna, M.A., and Riedmann, A. (2008). Marriages and Families: Making Choices in a Diverse Society. Wadsworth Publishing.

National Research Council. (2009). Reengineering the Survey of Income and Program Participation. Panel on the Census Bureau's Reengineered Survey of Income and Program Participation., Constance F. Citro and John Karl Scholz, Editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academy Press.

Strong, B., DeVault, C. , and Cohen, T.F. (2010). The Marriage and Family Experience: Intimate Relationships in a Changing Society. Wadsworth: Belmont, CA.

Figure 1. 2010 SIPP-EHC Type2 Question Battery



Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2008 Panel and 2010 SIPP-EHC.

Table 1. T-Test Results, Demographic Analysis

Type2, Demographic Analysis

	SIPP	SIPP-EHC	Diff (SIPP-SIPPEHC)	
Total	n=8,377	n=15,452		
Presence of Type2s	0.0543	0.0462	0.0081	*
	(0.0025)	(0.0017)	(0.0029)	
Type2, Sex	n=455	n=714		
Male	0.5275	0.4538	0.0737	*
	(0.0234)	(0.0186)	(0.0299)	
Type2, Race	n=455	n=695		
White	0.5560	0.5813	-0.0253	
	(0.0233)	(0.0187)	(0.0299)	
Black	0.3736	0.3137	0.06	*
	(0.0227)	(0.0176)	(0.0285)	
Asian	0.0462	0.0705	-0.0243	
	(0.0099)	(0.0097)	(0.0144)	
Other	0.0242	0.0345	-0.0104	
	(0.0072)	(0.0069)	(0.0104)	
Type2, Age	n=455	n=714		
LT 5	0.0879	0.0630	0.0249	
	(0.0133)	(0.0091)	(0.0156)	
5-14	0.1099	0.0868	0.0231	
	(0.0147)	(0.0105)	(0.0177)	
15-19	0.0769	0.1120	-0.0351	*
	(0.0125)	(0.0118)	(0.0179)	
20-29	0.2923	0.3165	-0.0242	
	(0.0213)	(0.0174)	(0.0277)	
30-49	0.2374	0.2493	-0.0119	
	(0.0200)	(0.0162)	(0.0258)	
50-69	0.1099	0.1289	-0.0190	
	(0.0147)	(0.0125)	(0.0196)	
70+	0.0857	0.0434	0.0423	*
	(0.0131)	(0.00763)	(0.0142)	

continued, next page

Type2, Demographic Analysis (continued)

Type2, Education	n=366	n=652		
LT High School	0.1175	0.3696	-0.2521	*
	(0.0169)	(0.0189)	(0.0282)	
HS Diploma	0.7213	0.4693	0.2520	*
	(0.0235)	(0.0196)	(0.0315)	
College	0.1612	0.1610	0.0002	
	(0.0192)	(0.0144)	(0.0240)	
Type2, Relationship Status	n=455	n=714		
Spouse	0.2022	0.0896	0.1126	*
	(0.0188)	(0.0107)	(0.0201)	
Child	0.3121	0.2885	0.0236	
	(0.0217)	(0.0170)	(0.0274)	
Parent	0.3538	0.2563	0.0975	*
	(0.0224)	(0.0164)	(0.0272)	
Sibling	0.2066	0.2787	-0.0721	*
	(0.0190)	(0.0168)	(0.0259)	
Other relative	0.3560	0.3627	-0.00670	
	(0.0225)	(0.0180)	(0.0288)	
Non-relative	0.0835	0.3291	-0.2456	*
	(0.0130)	(0.0176)	(0.0244)	

* $p<0.05$

Standard errors shown in parentheses

Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2008 Panel and 2010 SIPP-EHC.

Table 2. T-Test Results, Length of Stay Analysis

Type 2, Length of Stay Analysis

	SIPP	SIPP-EHC	Diff (SIPP-SIPPEHC)	
Type2, Length of Stay	n=455	n=714		
Mean Duration, in months	6.7868	6.1232	0.6636	*
	(0.1661)	(0.1345)	(0.2144)	
Type2, Months Present	n=455	n=714		
January	0.8505	0.6583	0.1923	*
	(0.0167)	(0.0178)	(0.026)	
February	0.8154	0.6527	0.1627	*
	(0.0182)	(0.0178)	(0.0267)	
March	0.7780	0.6415	0.1366	*
	(0.0195)	(0.0180)	(0.0274)	
April	0.7451	0.6317	0.1134	*
	(0.0205)	(0.0181)	(0.0279)	
May	0.6989	0.5980	0.1009	*
	(0.0215)	(0.0184)	(0.0287)	
June	0.6220	0.5910	0.0309	
	(0.0228)	(0.0184)	(0.0294)	
July	0.5297	0.5588	-0.0292	
	(0.0234)	(0.0186)	(0.0299)	
August	0.4659	0.5098	-0.0439	
	(0.0234)	(0.0187)	(0.0300)	
September	0.4000	0.4230	-0.0230	
	(0.0230)	(0.0185)	(0.0296)	
October	0.3648	0.3473	0.0175	
	(0.0226)	(0.0178)	(0.0287)	
November	0.3055	0.2885	0.0170	
	(0.0216)	(0.0170)	(0.0274)	
December	0.2110	0.2227	-0.0117	
	(0.0191)	(0.0156)	(0.0248)	
Type2, Duration Frequency	n=455	n=714		
1-4 months	0.2901	0.4006	-0.1105	*
	(0.0213)	(0.0184)	(0.0286)	
5-8 months	0.3736	0.3207	0.0529	
	(0.0227)	(0.0175)	(0.0284)	
9-12 months	0.3363	0.2787	0.0576	*
	(0.0222)	(0.0168)	(0.0275)	

* $p<0.05$

Standard errors shown in parentheses

Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2008 Panel and 2010 SIPP-EHC.

Table 3. T-Test Results, Economic Analysis

Type2, Economic Analysis

	SIPP	SIPP-EHC	Diff (SIPP-SIPPEHC)	
Type2, Annual Income	n=367	n=372		
Mean Income	15694.1	18615.5	-2921.3	
	860.3	1658.3	1875.6	
Type2, Income Distribution	n=367	n=372		
LT $10K	0.4605	0.4946	-0.0341	
	(0.0261)	(0.0260)	(0.0368)	
$10K - $25K	0.3134	0.2688	0.0445	
	(0.0242)	(0.0230)	(0.0334)	
$25K - $50K	0.1935	0.1371	0.0564	*
	(0.0206)	(0.0179)	(0.0273)	
$50K+	0.0327	0.0995	-0.0668	*
	(0.0093)	(0.0155)	(0.0182)	

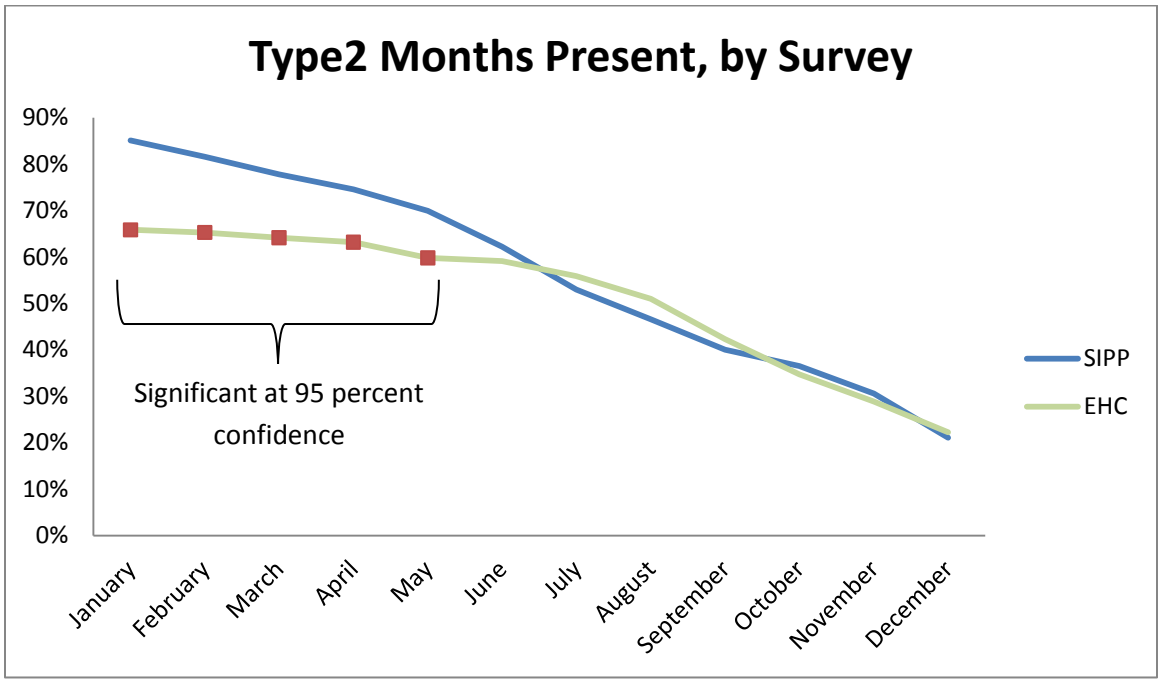
* p<0.05

Standard errors shown in parentheses

Of individuals aged 15 and older

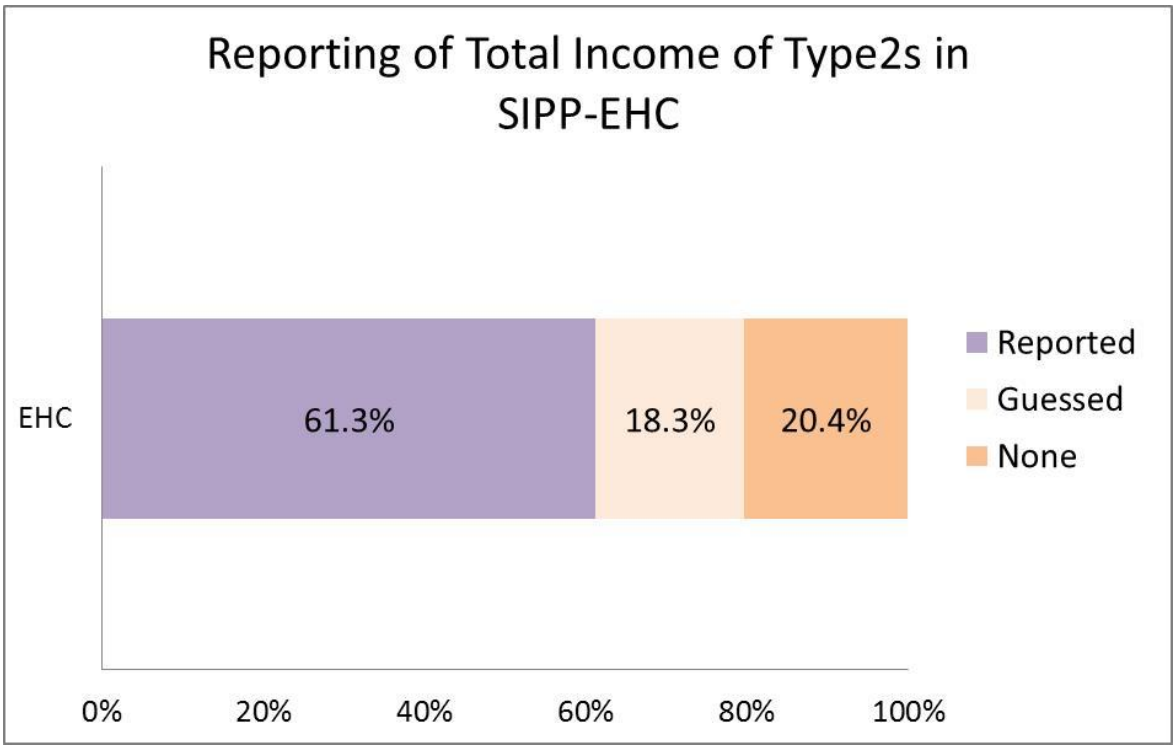
Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2008 Panel and 2010 SIPP-EHC.

Figure 2. Type2 Months Present, by Survey



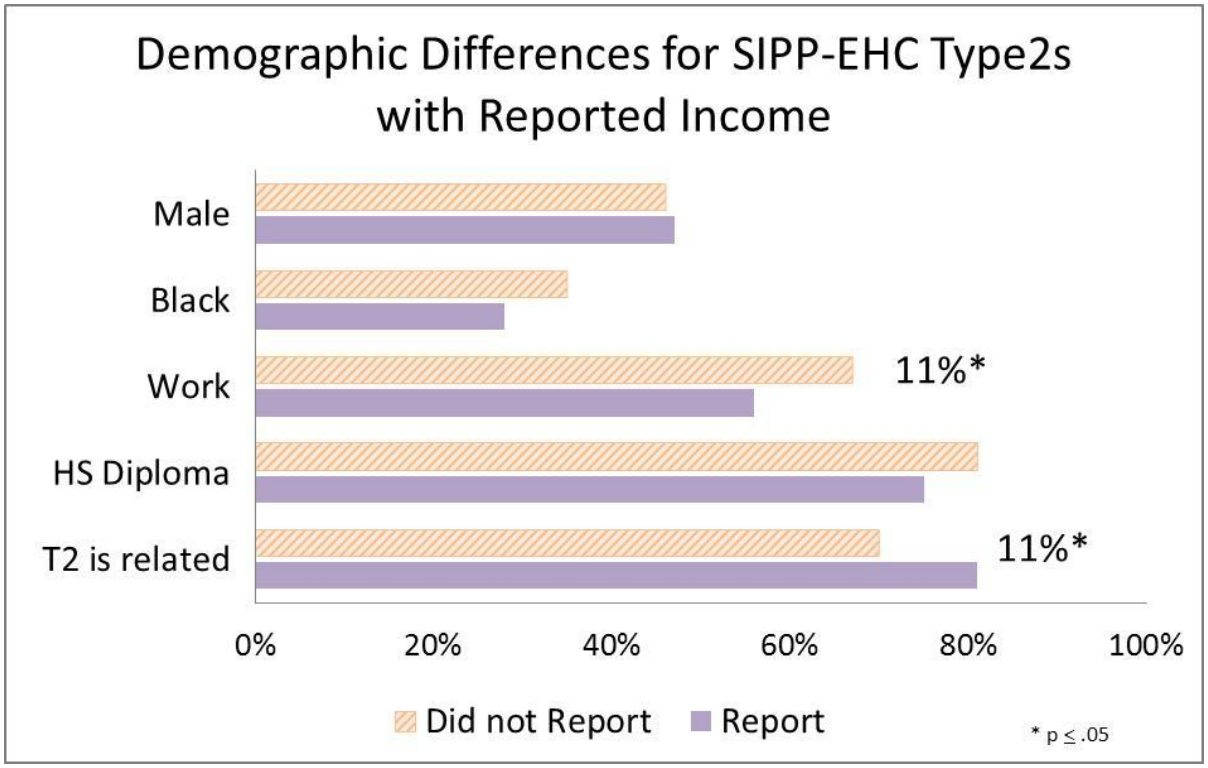
Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2008 Panel and 2010 SIPP-EHC.

Figure 3. Reporting of Total Income of Type2s in SIPP-EHC



Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2008 Panel and 2010 SIPP-EHC.

Figure 4. Demographic Differences for SIPP-EHC Type2s with Reported Income



Note: Age and Marital Status were also compared, but the differences were not significant at the 95 percent confidence level.
Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2008 Panel and 2010 SIPP-EHC.

Table 4. Imputation Results

Model 1	Imputed		EHC		EHC Dif.	95% sig.	SIPP		SIPP Dif.	95% sig.
	Mean	Std. Err.	Mean	Std. Err.			Mean	Std. Err.		
All Type2 imputed	16,153	780	18,616	1,658	2,463	ns	15,694	860	-459	ns
Only missing Type2 imputed	17,738	1,269			878	ns			-2,043	ns
Only non-missing Type2 imputed	15,152	977			3,464	ns			543	ns
Missing T2 imputed, non-missing T2 retained	18,276	1,131			340	ns			-2,581	ns
Model 2										
All Type2 imputed	14,662	656	18,616	1,658	3,954	ns	15,694	860	1,032	ns
Only missing Type2 imputed	15,085	1,069			3,531	ns			609	ns
Only non-missing Type2 imputed	14,395	831			4,221	ns			1,300	ns
Missing T2 imputed, non-missing T2 retained	17,249	1,099			1,367	ns			-1,555	ns

Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2008 Panel and 2010 SIPP-EHC.

Income Estimation for Former Household Members in the Redesigned SIPP

A Methodological Analysis of Instrument Efficacy

Lindsay Monte & Ashley Edwards
Social, Economic, and Housing Statistics Division
United States Census Bureau

Background

The Census Bureau's Survey of Income and Program Participation (SIPP) is a longitudinal survey, collecting monthly data on family composition and economic well-being. In 2006, Congress mandated a reengineering of the SIPP (NRC, 2009). As part of this redesign, the interview schedule moved from every four months to a single annual interview. The redesigned survey utilizes an Event History Calendar as a memory aid, and is known as the Survey of Income and Program Participation – Event History Calendar, or the SIPP-EHC.

One of the challenges of the SIPP-EHC is capturing data on individuals who were part of the household at some point in the reference year, but left the household before the interview. Data on these individuals, referred to as *"Type2"* people, are important in order to understand how household composition and income change over the year.

Research Questions

1. How well does the SIPP-EHC capture Type2 people when compared to SIPP data?
   a. Are the Type2 people captured in the SIPP-EHC demographically and economically similar to similar individuals captured in the SIPP?
2. Are we accurately measuring annual income amounts for Type2 people in the SIPP-EHC?
   a. Are there ways that we could improve our battery of Type2 questions in the SIPP-EHC to improve income imputation, and better understand household financial well-being?

Data

We compare data on Type2 persons collected from the 2010 SIPP-EHC field test (n=714) to data collected through the 2008 SIPP panel.

Our comparison sample in the 2008 SIPP consists of individuals who would have been considered Type2 persons under the SIPP-EHC survey design, referred to as *"Pseudo-Type2"* persons (n=455).

Both samples consist of individuals who do not live at the sampled address at the time of the Q1 2010 interview, but who, at some point between January and December 2009, lived with someone who resides at the sampled address as of the Q1 2010 interview.[1]

Methods

Bayesian Bootstrap Multiple Imputation model to assess income

- Two imputation models:
  1. Using data available from the Type2 question battery
  2. Using question battery data, plus additional demographics[2]
- Four implicates of each model
- Results are the means across these implicates

Conclusions/Implications

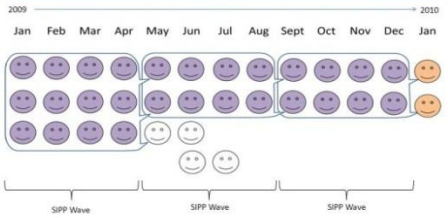
The SIPP-EHC captures fewer Type2s overall than does the SIPP (Fig.1), although the SIPP-EHC captures more short stay Type2s (Fig.9). Type2s are much more likely to be non-relatives of respondents than are Pseudo-Type2s (Fig.4), and are also distinct in a number of other ways (Figs.2,3,5,8), suggesting that the SIPP-EHC is not necessarily capturing the same individuals as the SIPP.

Despite these differences, we do not find significant differences in mean reported income between samples (Fig. 11), although Type2s are more likely to be reported as having annual incomes greater than $50,000. (Fig.6). However, for approximately 40% of the SIPP-EHC Type2s, we did not obtain a useable income response (Fig.12). Moreover, we find some significant demographic differences between Type2s for whom we have reports of income and those for whom we do not (Fig.13).

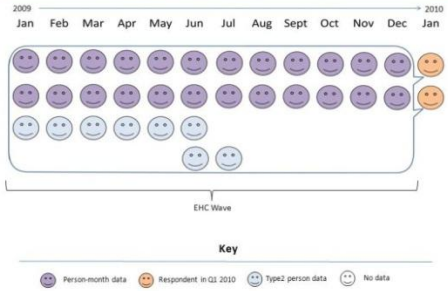
Nonetheless, regardless of how we consider non-response in our imputations, we do not find any significant differences in mean imputed income, either when compared to the SIPP mean or the mean of SIPP-EHC Type2 reports (see Model 1, Table 2). We further find that the addition of derived demographic information does not change our income imputations (see Model 2, Table 2).

This suggests that, for the Type2s captured, the current question battery may be sufficient, and follow-up income questions may not be necessary (Fig.10). However, additional efforts may be needed to capture Type2s whose co-residence was early in the year (Fig.7), and additional research is needed to understand differential non-response by demographic categories (Fig.13).

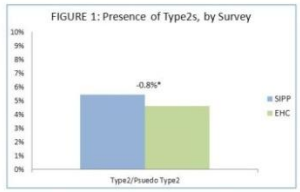
Pseudo-Type2s in the SIPP

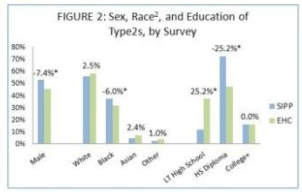


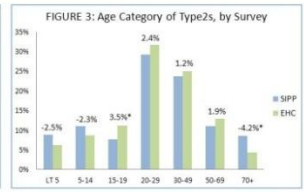
Type2s in the SIPP-EHC

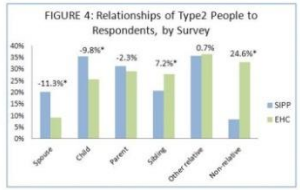


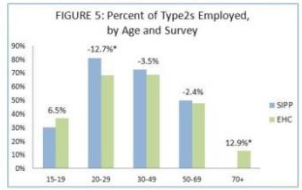
Who are our "Type2" people and how do they compare across surveys?

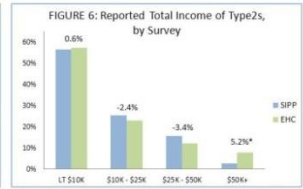


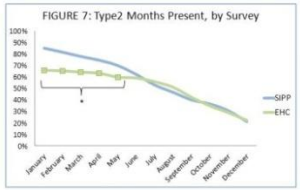


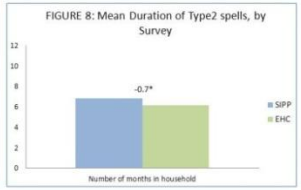












Economic Analysis

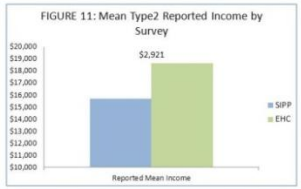
FIGURE 10: Income Questions from the SIPP-EHC Type2 Question Battery

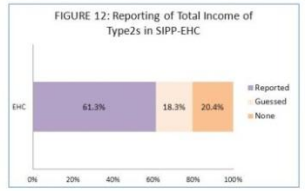
Asked about Type2 individuals age 15 and older:

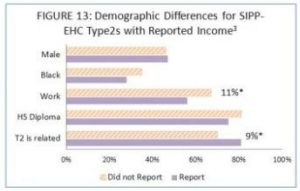
Including all income sources, what would you guess was (Type2 person's) annual income?

- IF VALID DOLLAR AMOUNT: Reported income
- IF DON'T KNOW / REFUSE:
  - *Was it more than $25,000?*
  - *Was it more than $50,000?*
  - *Was it less than $10,000?*

  } Guessed income
- IF DON'T KNOW / REFUSE: Non-response







* Defined as statistically significant at the 95 percent confidence level

DATA SOURCE: 2008 SIPP Panel and 2010 SIPP-EHC. http://www.census.gov/sipp/source.html

NOTES:
1. These data collected from the SIPP 2008 panel have been subset to match the 2009 calendar year, sampling characteristics, and geography of the 2010 SIPP-EHC field test. For this reason, the data presented here are not intended to be nationally representative. Differences in samples composition as well as the opportunity for attrition in the 2008 SIPP panel may introduce selection bias into our analysis.
2. Because race and marital status are not part of the Type2 questions, we logically impute them from available household data.
3. Age and marital status were also compared but the differences are not significant.

WORKS CITED:
National Research Council. Reengineering the Survey of Income and Program Participation. 2009. http://www.nap.edu/catalog.php?record_id=12715

Model Results

Table 1. Imputation Model Design

Independent Variables	SIPP	EHC	Model 1	Model 2	Outcome Variable: Variable	Outcome Variable: Definition
Sex	Available	Available	X	X	Annual person income from all sources	Defined as the sum of personal earned income, transfer income, and other income
Education	Available	Available	X	X		
Age	Available	Available	X	X		
Employment/Work	Available	Available	X	X		
Marital Status[2]	Available	Logically derived		X		
Race[2]	Available	Logically derived		X		

Table 2. Imputation Model Results

Model 1	Imputed: Mean	Imputed: Std. Err.	EHC: Mean	EHC: Std. Err.	EHC Dif.	95% Sig.	SIPP: Mean	SIPP: Std. Err.	SIPP Dif.	95% Sig.
All Type2 imputed	16,153	780	18,616	1,658	2,463	ns	15,694	860	-459	ns
Only missing Type2 imputed	17,738	1,269			878	ns			-2,043	ns
Only non-missing Type2 imputed	15,152	977			3,464	ns			543	ns
Missing T2 imputed, non-missing T2 retained	18,276	1,131			340	ns			-2,581	ns
Model 2										
All Type2 imputed	14,662	656	18,616	1,658	3,954	ns	15,694	860	1,032	ns
Only missing Type2 imputed	15,085	1,069			3,531	ns			609	ns
Only non-missing Type2 imputed	14,395	831			4,221	ns			1,300	ns
Missing T2 imputed, non-missing T2 retained	17,249	1,099			1,367	ns			-1,555	ns

The views expressed in this research, including those related to statistical, methodological, technical, or operational issues, are solely those of the authors and do not necessarily reflect the official positions or policies of the Census Bureau. The authors accept responsibility for all errors. This poster is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

The estimates in this poster are based on responses from a sample of the population and may differ from actual values because of sampling variability or other factors.