Behind the Numbers: Understanding the Survey of Consumer Finances

Sherman D. Hanna[1]
Kyoung Tae Kim[2]
Suzanne Lindamood[3]

Abstract

The Survey of Consumer Finances (SCF) is the most frequently used dataset for research in this journal, but many researchers and readers do not fully understand some of the dataset's complex details. This article provides insight into important issues that researchers and readers need to understand to accurately conduct and interpret SCF-based research. The issues addressed include the primary economic unit versus the household, identifying the respondent versus the head, limitations of variables in the survey, imputation and implicates, shadow variables, the public dataset versus the full dataset, weighting of analyses, and the use of replicate weights.

Keywords: Complex Survey Design, Multiple Imputation, Replicate Weights, Statistical Analysis, Survey of Consumer Finances, Survey Research

Importance of Research Using the SCF

The Survey of Consumer Finances (SCF) has been the survey dataset most frequently used for research published in this journal. In the first 20 volumes of the *Journal of Financial Counseling and Planning*, an SCF dataset was used in 64 of the 98 articles that used a national dataset (Ji, Hanna, Lawrence, & Miller, 2010). Likewise, Hanna et al. (2011) found that the SCF was the most frequently used national dataset in the first 18 volumes of *Financial Services Review*. One reason researchers frequently use the SCF is that it is nationally representative of households of all ages in the United States and contains more detailed household balance sheet information of U.S. households than other datasets (Bricker, Henriques, Krimmel, & Sabelhaus, 2016). The SCF also provides information about income sources, household characteristics, and a number of attitudinal and expectation questions. The Federal Reserve Board has sponsored the SCF as a cross-sectional survey every three years since 1983 and has twice sponsored panel surveys with re-interviews of households in 1983-1986-1989 and again in 2007-2009. Articles in the *Federal Reserve Bulletin* (e.g., Bricker et al., 2017) provide overviews of important patterns found in the SCF datasets regarding the financial status of U.S. households. Codebooks for each dataset (e.g., SCF Staff, 2017a) provide very detailed information about sampling and questions.

[1] Sherman D. Hanna, Professor, Human Sciences Department, Ohio State University, 1787 Neil Ave., Columbus, OH 43210. 614-292-4584. E-mail: hanna.1@osu.edu.
[2] Kyoung Tae Kim, Assistant Professor, Department of Consumer Sciences, 312 Adams Hall, Box 870158, University of Alabama, Tuscaloosa, AL 35487. E-mail: ktkim@ches.ua.edu.
[3] Suzanne Lindamood, JD, PhD, Retired, Ohio Legislative Service Commission, Columbus, OH. E-mail: Suzanne@lindamood.com.

Overview of Article

The main objective in this article is to explain some of the more challenging issues related to using SCF datasets. Each SCF dataset is complex, presenting changes from one dataset to the next, and therefore, it is important for researchers and those reading SCF-based research to understand some of these complexities. As Lindamood, Hanna, and Bi (2007) noted, many researchers have failed to document exactly what they did in terms of basic issues, such as sample weighting and identification of the respondent. Others made mistakes in variable selection by not understanding definitions and limitations. These problems are not limited to early users of the SCF as we noticed similar carelessness in more recent publications. In this article, we discuss some of the more frequently misunderstood issues covered in previous literature, especially the Lindamood et al. (2007) and the Shin and Hanna (2017) articles. We also cover articles and working papers the SCF staff has published, as well as detailed analyses of issues not covered in previous journal articles, such as the "shadow variables" and changes in key variables over time.

In this article, we first describe the survey design, which is important for understanding the sampling weights needed in the descriptive analyses. We discuss the issue with the primary economic unit, followed by a discussion of the survey process, including the timing of the surveys and the selection of respondents. We then provide an overview regarding the range of questions in the SCF. We discuss in detail issues related to imputation of data and the implicates, followed by a discussion of the shadow variables. We also discuss the public dataset and how it differs from the internal dataset used by the Federal Reserve Board in regards to published tables. Finally, we discuss the weighting of analyses, a topic covered in depth by Shin and Hanna (2017) and we provide practical examples of creating weight variables for different purposes.

Although much of the information in this article is based on information in the SCF codebooks, this is the first compilation of information and guidelines to be published in a journal. Such a compilation is important because the extensive length of the codebooks makes it challenging to find details and/or draw conclusions. For instance, the 2016 codebook (SCF Staff, 2017a) has 827 pages[4] and it is apparent that some researchers have not carefully read sections relevant to their analyses. Table 1 lists 12 articles published in the *Journal of Financial Counseling and Planning* since 2012 that used SCF datasets. Of those 12 articles, 10 identified whether weighting had been used in multivariate analyses, 7 mentioned whether Repeated-Imputation Inference (RII) had been used, 7 mentioned the internal versus public datasets, 6 had clear discussions about the issue of the head versus the respondent, 4 mentioned the issue of the replicate weights, 2 mentioned the shadow variables, and none mentioned the issue with the primary economic unit. These are all key issues that should be considered by authors, reviewers, readers, and researchers when using SCF datasets.

[4] Each codebook is posted as a text file. For references to page numbers, we imported the text file into Word and used a 9 point Courier New Font.

Table 1.
Selected Articles Published after 2010 in the Journal of Financial Counseling and Planning that Used SCF Datasets.

Authors (year)	SCF Datasets Used
Fisher & Anong (2012)	2007 SCF
Letkiewicz & Hanna (2013)	1992-2007 SCF
Hong & Hanna (2014)	1992-2007 SCF
Shin & Hanna (2015)	2010 SCF
Bieker & Yuh (2015)	2010 SCF
Lei & Yao (2016)	2013 SCF
Huang, Xu & Chiang (2016)	2007 SCF
Fisher (2016)	2013 SCF
Schooley & Worden (2016)	2007-2009 SCF panel
Chalise & Anong (2017)	2007-2009 SCF panel
Kim, Wilmarth & Henager (2017)	2007-2013 SCF
Hanna, Zhang & Kim (2017)	2013 SCF

Note. Table created by authors based on search at afcpe.org for "Survey of Consumer Finances" in articles published since 2010.

Survey Design and Research Considerations

The SCF is composed of two samples: (1) a multi-stage area-probability sample (AP) and (2) a "list" sample (Bricker et al., 2017; SCF Staff, 2017a). The AP sample is selected by a standard multi-stage geographic area probability design while the list sample is drawn from a list of taxpayers provided by the Internal Revenue Service (IRS) from the statistical records of individual income tax filings. The SCF list sample is designed to disproportionately select households that are likely to be relatively wealthy as one objective of the SCF is to estimate household wealth in the U.S. If random population sampling was used, it would not result in the number of wealthy households needed to enable a robust analysis. Sampling and interviewing procedures for the 1992 to 2013 cross-sectional SCF datasets are similar to the procedures for the 2016 SCF. In addition, in 2009 the SCF re-interviewed 89% of eligible participants from the 2007 interviews, resulting in the first SCF panel dataset (Bricker, Bucks, Kennickell, Mach, & Moore, 2011) since the 1983-1986-1989 SCF panel dataset (Kennickell, 1996). The 2007-2009 panel dataset provides additional insight that would not be accessible using the cross-sectional samples, such as the impact of the Great Recession on households (e.g., Kim & Hanna, 2016). While valuable, it is not likely that the SCF will have another panel dataset in the foreseeable future.

The Primary Economic Unit

Many federal surveys, such as the U.S. Census, collect information for households. In the SCF, however, the survey unit is not the household, it is the "Primary Economic Unit" (PEU), which may be a household but also could be a subset of a household. A PEU is "an economically dominant single individual or couple … and all other individuals in the household who are financially interdependent with that individual or couple" (SCF Staff, 2017a, p. 4). The codebook gives this example: "… in the case of a household composed of a married couple who own their home, a minor child, a dependent adult child, and a financially independent parent of one member in the couple, the PEU would be the couple and the two children" (SCF Staff, 2017a, p. 4). In the 2016 dataset, 13% of the PEUs were in a household that also contained one or more members not in the PEU. For convenience, most researchers refer to households rather than to PEUs, and even in publications by SCF staff (e.g., Bricker et al., 2017) the terms "households" or "families" are used. Researchers should realize that incomes and other variables are for the PEU, not the household.

The Survey Process and Selection of the Respondent

Researchers should be aware of the timing of the interviews to understand what respondents knew at the time of the interview and any considerations in the greater economy that might impact answers. Interviewing typically starts in May of the survey year, shortly after most people have finished filing their income taxes. Almost all of the interviews are completed by the end of the calendar year (Kennickell, 2008). The SCF strives to have the most accurate responses possible from each household, therefore, an initial interview with an economically dominant adult member of the PEU is conducted to gain basic information about the household and to determine which household member will be the survey respondent. For households (PEUs) consisting of two adults sharing resources, an attempt is made to have the more financially knowledgeable spouse or partner to be the respondent and answer all questions (Lindamood, et al., 2007).

It is important to understand that the respondent and the head in couple households are not necessarily the same person and therefore, the two terms are not interchangeable in SCF datasets. The respondent is the person identified by the person contacted in the initial interview, as the more financially knowledgeable person, while the household head is a title assigned by the SCF. For mixed-sex couple households, the SCF always assigns the title of "household head" to the male partner. In addition, for same-sex couple households, the SCF assigns the title of "household head" to the older person in the partnership. Understanding whether the respondent is the same person as the head is especially important when analyzing attitudes in relation to demographic characteristics such as age. While the respondent is asked to answer the attitude questions in terms of both partners, such as risk tolerance, it is plausible that many answers reflect the respondent's own attitudes rather than the couple's joint attitudes (e.g., Hanna & Lindamood, 2010). Our analysis of the 2016 SCF data indicates that in 45% of couple households, the respondent is not the household head. Therefore, it is inappropriate to analyze the relationship between the head's characteristics (e.g., age and/or education) and attitudes and expectations, as those are provided only by the respondent. Another variable, which only applies to the respondent in the public versions of SCF datasets, is racial/ethnic self-identification.

Therefore, for couples where spouses/partners have different racial/ethnic identifications, the status of the respondent will not be the same as the status of the head. Based on U.S. Census surveys, in 2010, 8.4% of all married couples had different racial/ethnic identifications, and 15.1% of new marriages in 2010 had spouses with different racial/ethnic identifications (Wang, 2012). In 2010, 9.5% of all couple households had spouse/partners with different racial/ethnic identifications (Lofquist, Lugaila, O'Connell, & Feliz, 2012). Hanna and Lindamood (2008, p. 50) stated: "For brevity, in some parts of our discussion we refer to household racial/ethnic status (e.g., White households) rather than the more accurate term 'households with a White respondent'."

Questions in the SCF

Overview

The amount of information for each PEU in the SCF (referred to as "household" in this article for convenience) varies according to the complexity of the household's financial situation. This complexity can be visualized as being somewhat similar to a household's federal income tax return. For renter households with no investments or businesses, a federal income tax return could be a few pages long. Similarly, that household might have relatively little information in the SCF beyond basic household characteristics, income, and answers to attitude questions. Contrastingly, a household with many investments and/or businesses could have a federal income tax return of 100 pages or more and information that contains hundreds of variables in the SCF.

Table 2 shows 15 categories of questions in the SCF. The survey starts with screening questions and does not ask related questions that do not apply to the remainder of the survey. For example, if a respondent indicates that the household does not have a business, questions related to business ownership/management are not asked. The Federal Reserve Board generally attempts to maintain consistency in questions over the years in order to compare responses over time, but in each survey there are a number of variables could be dropped, added, or modified. For instance, over 150 variables in the 2013 SCF were removed or replaced by other variables in the 2016 SCF (SCF Staff, 2017b). For comparisons over time, it is important to check codebooks for each survey year to see if variables of interest have changed or if the definition of a seemingly similar variable has changed.

Financial Situation, Practices, and Credit

The survey includes a number of questions related to economic expectations, credit attitudes, and experiences with credit, such as being turned down for credit. There are also questions related to the types of financial institutions used, including payday lenders (e.g., Lee & Kim, 2018). Detailed questions about the principal residence include whether the property is used for a business or farm, structure type, as well as purchase price or rent and how likely the household is to move in the next year. For homeowners, there are many details about mortgage terms as well as any home equity lines of credit. The 2016 SCF contains more details compared to previous surveys on the household's education loans, including whether the household has started repayments or has had any loan forgiveness. There are many questions about other types of loans, and whether the household has ever filed for bankruptcy. Questions introduced in the 2016 SCF include whether the property has ever had foreclosure proceedings.

Table 2.
Categories of Questions in the Survey of Consumer Finances

Section	Topic
	Household Listing
Section A	Economic Expectations, Credit Attitudes and Financial Institutions
Section B	Payment Methods and Credit Cards
Section D	Principal Residence and Lines of Credit
Section E	Real Estate and Loans to Others
Section F	Businesses
Section G	Vehicles
Section H	Education Loans
Section I	Other Loans
Section J	Attitudes About Saving and Investing
Section N	Financial Assets
Section R	Work and Pensions
Section T	Income, Taxes, Income Expectations, and Support
Section X	Inheritances and Charity
Section Y	Demographics, Health

Note. Created by authors from "Summary listing of questions asked in the 2016 SCF", https://www.federalreserve.gov/econres/files/scfoutline.2016.pdf, downloaded September 27, 2017.

Saving and Investments

The SCF datasets include a number of questions about saving and investment attitudes, including reasons for saving (e.g., Lee & Hanna, 2015), the financial planning horizon (e.g., Hong & Hanna, 2014), risk tolerance (e.g., Yao, Hanna, & Lindamood, 2004; Yao, Gutter, & Hanna, 2005), and perceived retirement adequacy (e.g., Kim & Hanna, 2015). The SCF contains a number of sufficiently detailed questions related to financial assets designed to ascertain whether households owned various types of high return investments, as well as the proportion of high return investments across all investments (e.g., Shin & Hanna, 2015). For households managing a business, there are many questions, including years in the business and size of the business (Ji & Hanna, 2012).

Spending

One limitation of the SCF is that it does not provide variables allowing a researcher to identify total spending. There are variables that allow the identification of financial obligations, including rent, vehicle lease payments, and loan payments (Hanna, Yuh, & Chatterjee, 2012). There are also variables for respondent estimates of spending on food. Citing the lack of specific total spending information, Hong, Hanna, and Kim (2013) estimated total household spending by

comparing amounts that each household spent for financial obligations and food to the total household expenditures the U.S. Consumer Expenditure Survey reports by income categories. A more accurate and direct identification of total spending using SCF data is desired, however. The SCF also includes a question about spending relative to income, and this variable has been used in many articles as the basis for an indicator of whether the household overspent or saved during the previous year (Yuh & Hanna, 2010; Heckman & Hanna, 2015; Rha, Montalto, & Hanna, 2006).

Examples of Other Questions

There are also many questions related to employment compensation, pensions, and government benefits. Additional questions about household characteristics include perceived health status of the respondent and partner, whether either person smokes, and a new question in 2016 asking for the highest education degree earned by the parents of the head and the spouse/partner. The 2016 SCF added three financial literacy questions and one perceived financial knowledge question that will allow researchers to test the effect of financial literacy on financial behavior.

Imputation and Implicates

Imputation

Household surveys are limited by data that is missing, inconsistent, or unreliable, impacting the quality of any analysis, which depends on the quality of the data. Although each variable in the SCF dataset has a value, there are many cases in which the SCF staff, not the survey respondent, provided that value. The SCF deals with missing or inconsistent data in a variety of ways, depending on the specific problem. Sometimes it is possible for the SCF to infer the data from other information in the interview, the interviewer's observation, or by assuming a value from a response to a range of data when an actual number is needed. The SCF uses another traditional way of dealing with missing information in survey research, which is imputation – using known information such as age and education and information patterns to provide reasonable estimates of the missing values. It is not possible for a researcher to distinguish which variables contained answers that were not provided by the respondent, have been altered to protect household identity, or were selected for imputation of critical values. As stated in the 2016 codebook:

> "There is no key in this codebook or in the data set that would allow users to identify with certainty either which data items have been smoothed or otherwise altered, or which cases were selected for imputation of critical values that is, the shadow variables in this data set may not always reflect the true original status of every variable. Although this blurring of the data will have some effect on analysis, that effect should be negligible in most cases (SCF Staff, 2017a, p. 45)."

Implicates

To provide values for missing data or to mask the identity of a respondent, the SCF uses a sophisticated system of imputing information, providing an estimate for the variability of the imputed value along with a range of possible responses. The SCF actually calculates five different estimates for a missing value and thus, for each household, provides five complete datasets. Each of these datasets is called an implicate. It is important for researchers to note that

for the same household, data in one implicate may differ from data in another implicate. This may be due to different values being calculated to reflect variance in the possible values for a missing response, or a purposeful "miscoding" of data to mask the identity of a household, as Lindamood et al. (2007) discussed in detail.

The issue with using appropriate methods for the statistical tests of datasets with imputed data has been discussed for many years, most notably in publications by Rubin regarding multiple imputation (e.g., Rubin, 1996). For accurate statistical tests using the SCF, it is essential that researchers properly combine the five implicates for each household, as failure to do so results in a dataset five times the actual sample size, which leads to an overestimation of significance levels. At the same time, while using one implicate produces the correct sample size, using only one underestimates the variance and in general is more likely to result in statistically significant effects. Montalto and Sung (1996) and Montalto and Yuh (1998) analyzed the effect of using the RII procedure for analyses of SCF data. While use of the RII is important, Lindamood et al. (2007) observed that a number of journal articles using multivariate analyses of SCF datasets did not mention the RII approach, and in some cases, performed analyses on only one implicate. Lindamood et al. (2007) conducted a comparison of a logistic regression model using the RII technique and one using averaging of implicates. They found that use of RII resulted in fewer effects having statistical significance. Lindamood et al. recommended that the RII method be used for any statistical tests containing SCF datasets. The SCF codebooks provide guidance on using RII methods (SCF Staff, 2017a).

Shadow Variables

A researcher should be aware of the nature of the variables in an analysis and whether the value was calculated by the SCF staff (in this article referred to as "missing values") or actually given by the respondent. Generally, the more cases in which the respondent fails to provide an answer, the less reliable the data. Therefore, researchers should consider whether to remove cases with missing values if they find a number of such cases for key variables. Few researchers who use the SCF report on whether they examined their variables for missing values, however, the SCF provides information that a researcher can consult to determine if the variable was given by the respondent or SCF calculated. This information is contained in variables referred to as "shadow variables," and are also referred to as "J" variables. In the SCF dataset, variables containing data for analysis are indicated by the prefix "X." Each X variable also has a "shadow variable" indicated by the prefix "J." For example, "X6809" is the variable for racial/ethnic identity while "J6809" is its shadow variable.

The coding of the shadow variable indicates whether the data in the "X variable" was provided by the respondent (coded "0") or was provided by the SCF pursuant to their methods for dealing with missing and inconsistent responses. The SCF codebook contains a detailed list of codes and explanations for values other than "0" in the "J" variable. The 2016 codebook lists numerous possible reasons why an X variable might not represent a respondent's actual information, along with codes for why a J variable code is other than "0." From these explanations a researcher can judge the relative reliability of the value and, depending on the use of the variable in the particular analysis, determine whether to eliminate cases with missing values (J values other than 0). We recommend that for variables of primary interest, researchers

analyze the J variables to determine the frequency of missing values and the reasons for J values other than "0." Shin and Hanna (2015), for instance, after performing this type of analysis of the J variable, excluded all households with imputed values for the racial/ethnic variable, since that was the main focus of their analyses.

The Public Dataset Versus the Full Dataset

The Federal Reserve Board releases a public version of each SCF dataset, which is the version that all researchers must use. However, the tables released by the Federal Reserve Board are based on an internal (full) version of each dataset. "Variables not included in the public data set are available only to members of the Microeconomic Surveys Statistical Unit at the Federal Reserve Board" (Bricker et al., 2017, p. 37). A small number of households in each internal dataset are excluded from the public dataset. For instance, in the 2016 SCF, 6 households in the internal dataset were excluded from the public dataset because the households had a net worth that would qualify them for the *Forbes* magazine list of the 400 wealthiest families in the U.S. (Dolan & Kroll, 2016). Therefore, the public dataset for 2016 has 6,248 households compared to the 6,254 households in the full dataset.

The SCF codebook listing for each variable indicates whether the information in the public dataset is different from the information in the full dataset. It is important to understand the restrictions for some variables in the public dataset. For instance, for the variable X6809, the racial/ethnic identity of the respondent, four categories that are included in the full dataset are combined into an "Other" category in the public dataset (Lindamood et al., 2007), resulting in four categories: White, Black, Hispanic, and Other. Hanna and Lindamood (2008) concluded that most of the respondents listed in the Other racial/ethnic category in the SCF public dataset were Asian or Pacific Islander. Location variables provide another example of the difference between the public dataset and the full dataset. For instance, a variable indicating the census region of residence is not included in the public datasets after 1998.

A third example denoting the difference between the public dataset and the full dataset is related to the occupation and job status variables. The full dataset version of the variable X7401 has answers corresponding to the Census 4-digit occupation codes, enabling one to identify very detailed occupational levels, e.g., "Nuclear Technicians." However, the public dataset only has six codes for occupation, one of which is an exceptionally broad category including all teachers, managers, and health practitioners.

Weighting of Analyses

The SCF provides a weight variable, X42001,which adjusts the weight for each household so that sample statistics reflect the population of households in the U.S. Weighting of analyses is more important in the SCF than in most other survey datasets because the sampling for the survey includes both an area probability sample of the general population and a list sample provided by the Internal Revenue Service. This dual sampling method results in a combined sample that includes a higher proportion of high wealth households (SCF Staff, 2017a) than possible with only a probability sample. The dual sampling method enables the SCF to obtain more accurate estimates of household wealth than are possible using only a probability

sample, but also requires that researchers use weighting in analysis to avoid producing biased results.

Although weighting the sample produces estimates that reflect all households in the U.S., one should be careful when using weighting, as applying the weighting factor X42001 to the sample results in an apparent sample five times as large as the number of households in the U.S. Table 3 shows how the application of various weighting variables results in different apparent sample sizes. If researchers create a new weight variable, WGT, by dividing X42001 by 5, the application of that weight in the 2016 SCF dataset results in an estimated sample size of 125,982,000, which is approximately the same as the July 1, 2016 Census estimate of 125,819,000 households in the United States (U.S. Census, 2018). Application of this researcher-created WGT weight variable is inappropriate for statistical tests of significance, as standard statistical software such as SAS will assume the sample size is very large. The NWGT variable (which we created and is shown in Table 3) should be used for simple statistical tests when not using an RII technique (see discussion below). For tests of statistical significance when using an RII technique (e.g., Shin & Hanna, 2017; and Table 4 in Hong & Hanna, 2014), a researcher-created weight variable based on the population weight should be used (AWGT in Table 3).

Table 3.
The Apparent number of U.S. Households, as Calculated by Alternate Weighting Approaches, 2016 SCF Dataset (Census estimate: 125,819,000).

Number of Households Calculated Using:		
X42001	WGT=X42001/5	NWGT*
629,910,000	125,982,000	6,248

Note. *NWGT formula
x42001/(5*20163.523)
For bivariate statistical tests using RII procedures, the variable AWGT = NWGT * 5 should be used. Census estimates of the number of households: U.S. Census (2018)

Researchers should consider the use of variables in their analysis and then use appropriate weights, which differ depending on the purpose of the analysis. Table 4 provides simple illustrations based on an indicator we created, whether net worth was greater than one billion dollars. Using a weight we created, WT5 (set equal to 0.2), we found that there are eight billionaires in the 2016 sample. The proportion of billionaires in the U.S. is estimated at 0.0004% of the population, which implies that with random sampling, we would expect no billionaire households to be included in the SCF database (weighted N=0.025). Because a disproportionally high percentage of the sample is wealthy, a weight reflecting the population weight (NWGT) should be used to obtain distributions reflecting all U.S. households. By using the weight WGT, one can obtain estimates of the numbers of particular types of households in the U.S. For instance, we estimated that in 2016 there were 505 households with a net worth over one billion dollars in the U.S.

Weighting of Multivariate Analyses

The issue of whether to weight multivariate analyses when using SCF datasets has been controversial (Kennickell, 2003). In a discussion of the controversies, Lindamood et al. (2007) compared unweighted logistic regression analyses to the same analyses weighted by scaled population weights. They found that effects in the weighted RII analyses were more likely to be conventionally significant (at 5% level) than in the unweighted RII analyses. They also suggested that it would be best to report exact significant levels for effects rather than the simple indicators that are traditionally used.

Table 4.
The Number and Proportion of Households with Net Worth Over $1,000,000,000. As Calculated by Application of Alternate Weights, 2016 SCF Dataset

Weighting Method:	Net Worth > $1,000,000,000
Averaged across implicates (WT5)	
N (actual number in sample)	8
Proportion	0.1280%
Population weight/5, scaled to make total N=sample size (NWGT)	
N (estimated number in sample)	0.0250
Proportion	0.0004%
Population weight/5 (WGT)	
N (estimated number in U.S.)	505
Proportion	0.0004%

Note. Calculated by authors based on analyses of the 2016 SCF.

Shin and Hanna (2017) provided a detailed discussion of the weighting issue when using an imputed dataset along with the issue with the complex survey design of the SCF. They presented comparisons of unweighted RII regressions to RII regressions with population weights applied and also to RII regressions with both population weights and replicate weights applied. They concluded that unweighted RII regressions provide the most conservative estimates of effects, but ideally, unweighted analyses and analyses with both population and replicate weights should be conducted, and if the results differ substantially, researchers should reconsider their models.

Conclusions

Researchers using any dataset should describe their methods and assumptions in enough detail to allow for replication and also carefully discuss issues in analyses related to the complexity of the dataset. The SCF datasets are complex, thus any publication using such a dataset that does not provide a detailed discussion of the issues mentioned in this article should be considered deficient, as interpretation of results may be flawed. The salient issues addressed in this article include the primary economic unit versus the household, identifying the respondent

versus the head, limitations of variables in the survey, imputation and implicates, shadow variables, the public dataset versus the full dataset, weighting of analyses, and the use of replicate weights. Researchers using SCF datasets can use the list of salient issues mentioned above as a checklist to ensure that their methods properly consider the complexities of the SCF. Readers and reviewers of manuscripts using the SCF should keep these salient issues in mind when evaluating methods described in the manuscripts.

References

Bieker, R. F., & Yuh, Y. (2015). Homeownership and financial strain following the collapse of the housing market: A comparative study on loan delinquencies between Black and White households. *Journal of Financial Counseling and Planning*, *26*(1), 79-93.

Bricker, J., Bucks, B., Kennickell, A., Mach, T., & Moore, K. (2011). Surveying the aftermath of the storm: Changes in family finances from 2007 to 2009. (FEDS Working Paper No. 2011-17). Washington, D.C.: Financial and Economic Discussion Series.

Bricker, J., Dettling, L. J., Henriques, A., Hsu, J. W., Jacobs, L., Moore, K. B., ... Windle, R.A. (2017). Changes in U.S. family finances from 2013 to 2016: Evidence from the Survey of Consumer Finances. *Federal Reserve Bulletin*, *103*(3), 1-41.

Bricker, J., Henriques, A., Krimmel, J., & Sabelhaus, J. (2016). Measuring top wealth shares estimating top income and wealth shares: Sensitivity to data and methods. *American Economic Review*, *106*(5), 641–645.

Chalise, L., & Anong, S. (2017). Spending behavior change and financial distress during the Great Recession. *Journal of Financial Counseling and Planning*, *28*(1), 49-61.

Dolan, K. A., & Kroll, L. (2016, October 4). Inside the 2016 Forbes 400: Facts and figures about America's richest people. *Forbes*, 1. Retrieved from: https://www.forbes.com/sites/kerryadolan/2016/10/04/inside-the-2016-forbes-400-facts-and-figures-about-americas-richest-people/#2217e6cd7517

Fisher, P. J. (2016). Differences in credit card use between White and Hispanic households. *Journal of Financial Counseling and Planning*, *27*(1), 199-211.

Fisher, P. J., & Anong, S. T. (2012). Relationship of saving motives to saving habits. *Journal of Financial Counseling and Planning*, *23*(1), 63-79.

Hanna, S. D., Ji, H., Lee, J., Son, J., Letkiewicz, J., Lim, H., & Zhang, L. (2011). Content analysis of *Financial Services Review*. *Financial Services Review*, *20*(3), 237-251.

Hanna, S. D., & Lindamood, S. (2008). The decrease in stock ownership by minority households. *Journal of Financial Counseling and Planning*, *19*(2), 46-58.

Hanna, S. D., & Lindamood, S. (2010). Can one spouse accurately represent couple risk tolerance? *Consumer Interests Annual*, 56.

Hanna, S. D., Yuh, Y., & Chatterjee, S. (2012). The increasing financial obligations burden of U.S. households: Who is affected? *International Journal of Consumer Studies*, *36*(5), 588–594.

Hanna, S. D., Zhang, L., & Kim, K. T. (2017). Do worker expectations of never retiring indicate a preference or an inability to plan? *Journal of Financial Counseling and Planning*, *28*(2), 268-284.

Heckman, S. J., & Hanna, S. D. (2015). Individual and institutional factors related to low-income household saving behavior. *Journal of Financial Counseling and Planning*. *26*(2), 187-199.

Hong, E.O., & Hanna, S.D. (2014). Financial planning horizon: A measure of time preference or a situational factor? *Journal of Financial Counseling and Planning*, *25*(2), 184-196.

Hong, E. O., Hanna, S. D., & Kim, K. T. (2013). Which household have adequate emergency funds after the Great Recession? Estimates based on an income measure and a spending measure. *Proceedings of the Academy of Financial Services*.

Huang Jr, T., Xu, X., & Chiang, T. F. (2016). Household expectations for future economy and risk-taking attitudes. *Journal of Financial Counseling and Planning*, *27*(1), 109-121.

Ji, H., & Hanna, S.D. (2012). Factors related to the financial vulnerability of small business owner-manager households. *Journal of Personal Finance*, *11*(1), 49-77.

Ji, H., Hanna, S. D., Lawrence, F.C., & Miller, R. K. (2010). Two decades of the *Journal of Financial Counseling and Planning*. *Journal of Financial Counseling and Planning*, *21* (1), 4-14.

Kennickell, A. B. (1996). Codebook for 1983-1989 Panel SCF. Washington, DC: Board of Governors of the Federal Reserve System.

Kennickell, A. B. (2003). Codebook for 2001 Survey of Consumer Finances. Washington, DC: Board of Governors of the Federal Reserve System.

Kennickell, A. B. (2008). The bitter end? The close of the 2007 SCF Field Period. (SCF Working Paper). Washington, DC: Federal Reserve Board. Retrieved from https://www.federalreserve.gov/econresdata/scf/files/asa2008complete1.pdf

Kim, K. T., & Hanna, S. D. (2015). Do U.S. households perceive their retirement preparedness realistically? *Financial Services Review*, *24*(2), 139-155.

Kim, K. T., & Hanna, S. D. (2016). The severity of the impact of the stock market crash of 2008-2009 on the wealth of U.S. workers. *Journal of Financial Planning*, *29*(2), 52-58.

Kim, K. T., Wilmarth, M. J., & Henager, R. (2017). Poverty levels and debt indicators among low-income households before and after the Great Recession. *Journal of Financial Counseling and Planning*, *28*(2), 196-212.

Lee, J. M., & Hanna, S. D. (2015). Savings goals and saving behavior in the context of Maslow's hierarchy of needs. *Journal of Financial Counseling and Planning*, *26*(2), 129-147.

Lee, J., & Kim, K. T. (2018). The rise of payday loans and impaired credit after the Great Recession, *Journal of Family and Economic Issue*, *39*(2), 360-369.

Lei, S., & Yao, R. (2016). Use of financial planners and portfolio performance. *Journal of Financial Counseling and Planning*, *27*(1), 92-108.

Letkiewicz, J. C., & Hanna, S. D. (2013). The impact of stock and real estate returns on the propensity of households to meet the Capital Accumulation Ratio Guideline. *Journal of Financial Counseling and Planning*, *24*(1), 34-44.

Lindamood, S., Hanna, S. D., & Bi, L. (2007). Using the Survey of Consumer Finances: Methodological considerations and issues. *Journal of Consumer Affairs*, *41*(2), 195-214.

Lofquist, D., Lugaila, T., O'Connell, M., & Feliz, S. (2012). Households and families: 2010. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Montalto, C.P., & Sung, J. (1996). Multiple imputation in the 1992 Survey of Consumer Finances. *Financial Counseling and Planning*, *7*, 133-146.

Montalto, C. P., & Yuh, Y. (1998). Estimating nonlinear models with multiply imputed data. *Financial Counseling and Planning*, *9*(1), 97-101.

Rha, J., Montalto, C. P., & Hanna S. D. (2006). The effect of self-control mechanisms on household saving behavior. *Financial Counseling and Planning*, *17*(2), 3–16.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(434), 473-489.

SCF Staff (2017a). Codebook for the 2016 Survey of Consumer Finances. Retrieved September 27, 2017 from https://www.federalreserve.gov/econres/files/codebk2016.txt

SCF Staff (2017b). SCF changes. Retrieved October 1, 2017 from https://www.federalreserve.gov/econres/files/2016_scf_changes.txt.

Schooley, D. K., & Worden, D. D. (2016). Perceived and realized risk tolerance: Changes during the 2008 financial crisis. *Journal of Financial Counseling and Planning*, *27*(2), 265-276.

Shin, S. H., & Hanna, S.D. (2015). Decomposition analyses of racial/ethnic differences in high return investment ownership after the Great Recession. *Journal of Financial Counseling and Planning*, *26*(1), 43-62.

Shin, S. H., & Hanna, S.D. (2017). Accounting for complex sample designs in analyses of the Survey of Consumer Finances. *Journal of Consumer Affairs*, *51*(2), 255-283.

U.S. Census Bureau (2018, May). America's families and living arrangements: 2016. (Table H1. Households by type and tenure of householder for selected characteristics). Washington, DC: Current Population Survey, 2016 Annual Social and Economic Supplement. Retrieved from https://www2.census.gov/programs-surveys/demo/tables/families/2016/cps-2016/tabc1-all.xls.

Wang, W. (2012). The rise of intermarriage: Rates, characteristics vary by race and gender. *Social and Demographic Trends*. Washington, DC: Pew Research Center.

Yao, R., Hanna, S. D., & Lindamood, S. (2004). Changes in financial risk tolerance, 1983-2001. *Financial Services Review*, *13*(4), 249-266.

Yao, R., Gutter, M.S., & Hanna, S.D. (2005). The financial risk tolerance of Blacks, Hispanics and whites. *Financial Counseling and Planning*, *16*(1), 51-62.

Yuh, Y., & Hanna, S. D. (2010). Which households think they save? *Journal of Consumer Affairs*, *44*(1), 70-97.