Response to Darryl Getter's Comments to the Lindamood, Hanna and Bi Article in the Winter 2004 issue of the *Journal of Consumer Affairs*.

Suzanne Lindamood

The Article

In our recent publication in the *Journal of Consumer Affairs* , "Using the Survey of Consumer Finances: Some Methodological Considerations and Issues" (*The Journal of Consumer Affairs*, Winter 2007, Vol 41, No 2) we identify several methodological issues related to the Survey of Consumer Finances dataset and analyze how several researchers who have published in JCA using that dataset have dealt with those issues. A primary focus of our discussion is the need for authors to specify how they define and measure variables, which we consider standard procedure and necessary for accurate interpretation of results, replication, and comparison to other research. We state:

> As with any research, those who use SCF data need to make appropriate decisions, justify procedures, precisely define variables, and provide sufficient detail so readers can understand the findings and other researchers can replicate the study.

The Commentary

The editor requested three persons prepare commentaries as a response to our article. Two (Jeanne Hogarth and Sandra Huston) commented on the importance the issues addressed for users of the SCF as well as of other datasets. However, in a third commentary, the author (Darryl Getter) saw our article as an attempt to limit researchers in defining and measuring variables and expressed that there is no need for specifying measures because a competent researcher should be trusted. Perhaps the author takes offense at our methodological considerations because he was the author of one of the articles we criticized for lack of methodological specification. In general, his commentary would be of more use if he had offered examples to illustrate his points.

We find most of his comments to be without merit. They actually further illustrate the need for researchers to be familiar with a dataset and to define and explain their measures and methods. Only in this way will the reader know the researcher is creating new variables or perhaps using methods not appropriate for the data.

The following are selections from Getter's commentary, along with our response to it, pointing out why our original points are valid and that the commentator apparently does not understand why they are important.

Race Variables

The SCF public dataset contains four categories of race and ethnicity: White, Black, Hispanic, and Other. One problem that we discuss in our article is the creation of new race variables without specification or explanation. Getter sees this as limiting the freedom of researchers to redefine variables as they see fit. Getter states:

> *"Generally speaking, researchers should have the freedom to define variables the way they see fit in order to answer their particular research*

questions. Readers may feel that more suitable variable definitions could be used to answer a particular research question but that does not necessarily mean that a methodological error was made."

While this may sound like a researcher's right, when a variable in a well known dataset is redefined or reclassified, the researchers should clearly state how they define that variable and justify the reclassification. In our article we criticize researchers who redefine race variables without explanation. Getter, for instance, apparently reclassifies "Hispanic" as "Black" and Lyons apparently reclassifies "Hispanic" as "White" while the SCF considers "Hispanic" as a unique category. Authors who feel they need to redefine a variable "to answer a particular research question" by calling Hispanics "Black" or "White" should so state and discuss why their definitions are important.

In our paper, we refer to federal guidelines (OMB 15) that govern the reporting of race variables in federal documents. While the OMB 15 guidelines are not required of non-federal research, they do offer a reasonable guideline for uniform racial classifications in any research.

Getter excuses researchers who did not justify their race variables by implying that the SCF gathered data contrary to the federal guidelines. He states:

If the SCF did not explicitly break race into ethnicity groups or simply did not collect race information in accordance with the federal guidelines prior to 2004, then researchers should not be criticized . . .

What Getter fails to point out is that **the SCF did collect information in accordance with federal guidelines prior to 2004.** Getter continues:

"With regard to the specific criticism of how previous authors used the race variable, those criticisms seem rather unfair. Researchers are limited to the questions available."

This is true. And, as we point out, if researchers used the SCF variables, they would not violate federal guidelines. We state:

"OMB 15 requires the use of specific categories when reporting race and ethnicity data acquired with a single question format, such as the SCF race/ethnic question for the 1989 through 2001 surveys. The SCF single question format used during those years **has categories identical to those OMB 15 requires, except that the SCF also offered respondents an "Other" category. The SCF race/ethnicity information was collected and made available in a format that enables researchers to comply with the federal directive just by using the SCF categories and names**. "

Getter continues to defend researchers by stating:

> *"The authors disagree with how some of the researchers have defined variables used in their analysis. For example, they recommend that all researchers using the SCF race variables should define them according to guidelines set by the Office of Management and Budget. If a reader believes that a respecification of a variable definition will alter previously published empirical results, then the burden falls upon that reader to perform another analysis and present new findings that incorporate the improved variable definitions.*

Getter does not offer any example of how a researcher's preferred groupings might violate the federal guidelines. In fact, a major reason for specificity is to enable a reader to replicate research. Without specifying how a variable has been "respecified," a reader could not replicate or incorporate improved variables.

Distinguishing the Head and Respondent

In the SCF, the respondent is the person who answers the questions. SCF interviewers try to ensure that the respondent is the person identified during the initial contact as the most knowledgeable about the household's finances. The respondent may be either person in couple households. However, SCF always designates the male as household head in opposite sex couple households. As the 2004 SCF Codebook states:

> When the original respondent was someone other than the person determined to be the head in this sense, all data (including response codes) for the two members of the couple were systematically swapped. The variable X8000 indicates which cases have been subjected to such rearrangement.

It is up to the researcher to determine whether the head or the spouse is the respondent. The researcher does this by using variable X8000, the SCF "switch" variable, when appropriate. This extra step must be taken by the researcher, and the researcher should indicate whether demographic characteristics such as age are for the head or the spouse/partner of the head in couple households.

Getter fails to grasp this distinction and further dismisses the idea that there is any occasion to distinguish between the household head and respondent. He states:

> *Lindamood, Hanna, and Bi also criticize authors who do not distinguish between the ''respondent'' and the ''household head.'' Again, there is no reason to suspect that a methodological error has occurred when these terms are used interchangeably in articles. The SCF codes the answers of the household head separately from those of the spouse. Looking at the online SCF codebook, the responses of the household head clearly appear above those of the spouse or second household member. Even if the second respondent answers questions on behalf of the household head, the SCF staff goes through the responses and assigns them to the appropriate individual. Nevertheless, the care with which the SCF staff codes variables*

suggests that it is unnecessary for a researcher to go out of his or her way to talk about this distinction. All the researcher must do is to pay close attention and choose the correct variable code from the codebook.

Getter's answer indicates a lack of familiarity with the structure of the dataset. If the wife is the respondent, the wife is not **answering questions on behalf of the husband**, as Getter proposes. Analyzing data of the respondent is not a matter of choosing the correct variable code – a researcher must do more as stated above. To use head and respondent interchangeably is simply imprecise and wrong.

As Hogarth stated in her commentary:
"Depending on the research question under study, researchers need to know when to maintain distinctions of respondent, head, and other family members and when aggregation is both acceptable and preferable. . . .our unit of analysis may be the household (e.g. a household's net worth or level of indebtedness), but many of the control measures we want to use (age, race, and education) are specific to an individual within the household The solution to this is, as pointed out by Lindamood, Hanna, and Bi, is to clearly define our measures and use the appropriate terms in our writing."

Which Implicates the Researcher Uses

The SCF deals with missing data by creating a range of possible responses for a missing variable, This method results in five complete but distinct datasets for each respondent, and each of these datasets is known as an implicate. It is possible to use just one implicate in an analysis, but it is possible that the results will be biased, especially with respect to variables that had a greater rate of nonresponse. The SCF recommends using all five implicates, which are combined by a method known as "RII." We also recommend that researchers report whether they used all five implicates and whether they used RII to combine the implicates.

Getter does not agree as to the importance of this disclosure, and essentially says the reader should just assume that the correct methods have been used. Getter states:
"The suggestion that SCF researchers should always discuss whether or not all five implicates are being used is simply unnecessary. Most researchers would presume all five implicates are being used, and the only reason to bring up this issue would be if a researcher had a good reason for not using all five implicates. Generally, the SCF staff recommends using all implicates at all times. Of course, to avoid incorrectly reporting a sample size inflated by a factor of five, authors may prefer to divide the total number of observations by five and report a sample size closer to the truth.

It would take a sentence or two to state that the RII technique recommended in the codebook had been used, etc. We give examples of articles that did not provide even a sentence specifying the methods for use of implicates and weighting. We criticized Getter's article for its lack of methodological specificity, not for any particular method used or not used. Unfortunately, there was not enough information in his article for us to tell which methods were used.

Describing Methods Versus Contacting the Authors

Getter suggests that it is unreasonable to criticize authors for failing to describe important details of their methods, because readers can always contact the authors about missing details. We have found that this method does not always work, as even authors who are still alive may have discarded or misplaced important research records. We agree that it is impractical to list every possible detail of research methods, but our suggestions would be unlikely to add more than 1% to the length of an article. To add a sentence stating, e.g., the multivariate analyses were unweighted and used only Implicate 1, is not burdensome.