Using Paradata to Understand Panel Effects in the Current Population Survey October 2013

John Dixon
Bureau of Labor Statistics

Abstract
This study examines reluctance to respond to the Current Population Survey (CPS). The CPS is administered to a household for four consecutive months, followed by a break of eight months, and then administered for another four consecutive months. These eight interviews (panels) form the basis of this analysis. The respondents concerns and contact issues during the first panel were summarized using a factor analysis (Dixon, 2010). This study examines and compares the factor structure of the later panels to the first. This study also examines paradata, including respondent contact history recorded by interviewers to help better understand the survey experience and willingness to respond. This study found a similar pattern of concerns as previous studies for the first panel; however, the following panels exhibited different patterns than the first.

Key Words: Nonresponse, longitudinal survey, Contact History Instrument

1. Introduction
The Contact History Instrument (CHI) was designed to collect information about each contact attempt made by a field representative (FR); including information about why respondents refuse and what actions the FR took to attempt to obtain the interview (Dyer, 2004).

The present study uses reluctance concerns from the CHI to explore the experience of the respondents with the Current Population Survey (CPS). Since the reasons for nonresponse are obtained from both the nonrespondents and respondents, the relationship between nonresponse and survey estimates can be modeled. The changes in concerns for the different interviews can also give insight into the interview process, help build models for nonresponse studies, and inform potential changes in field procedures. This could lead to less bias or more efficient data collection.

2. Data Sources
Details about the CPS can be found in Technical Paper 66 (Census, 2006). The CPS is the primary source of information on the labor force characteristics of the U.S. population. The CPS uses a multistage probability sample based on the population counts from the decennial census, with coverage in all 50 states and the District of Columbia. The sample is continually updated to account for new residential construction. The proportion of sample households not interviewed in the CPS due to non-contact or refusals typically varies between eight and nine percent. Data may be collected either in person or by telephone, although the first and fifth interviews are supposed to be in person. This study doesn't consider households where data are collected by telephone centers (CATI), but does consider those where the field interviewer chooses to collect data by telephone.

CHI was added to the CPS in 2009 to collect detailed contact history data (Bates, 2004). The interviewer records times and outcomes of attempted contacts, problems or concerns reported by reluctant households, and strategies used to gain contact or overcome

reluctance. This provides a very rich source for studying the interview process. However, this study only used the answers recorded by interviewers in response to a question about reasons for not responding reported by reluctant households. Answers to a question about the strategies employed by an interviewer were not used in this analysis. Data from 2009 through 2011 was matched longitudinally to provide the changes in concerns studied here.

3. Methods

The Contact History Instrument (CHI) collects information about the respondent's reluctance and concerns about participating as well as number of contact attempts and barriers to contact. Understanding concerns that respondents have about participation and reasons behind their reluctance can adjust field methods in an attempt to improve response rates. A factor analysis had been used on the CHI data in previous studies (Maitland, et.al, 2009, Dixon, 2010) to explore the underlying factors in the reported reluctance. This study builds on past work by calculating refusal rates by each of the identified factors, allowing for examination of which factors have the strongest relationship with nonresponse.

Since the CPS is a longitudinal survey, with respondents asked to complete eight interviews, it is important to recognize they may have different concerns at each of the eight waves. Changes in reluctance may be due to the changing circumstances of the respondent and their prior experience with the interview. A factor analysis (Mplus, 2011) was used for each of the eight interview waves to see if the patterns of concerns were similar across the waves. Several models were tried and one model, which was more stable over interviews, was used to characterise the concerns. Changes from interview to interview were examined with cross-tabulation tables.

4. Results

Before looking at the pattern of concerns across waves, it is important to look at the overall pattern of reporting to understand the factor analysis results since lower frequency concerns are more difficult to estimate for the factor structure. In the CHI, the most common concern expressed by respondents during the first interview was "busy" (Figure 1), followed by "schedule difficulties", and "not interested". Other notable concerns were "time the interview takes" and "privacy concerns". Since many of the categories are low frequency, this is a challenge for analysis.
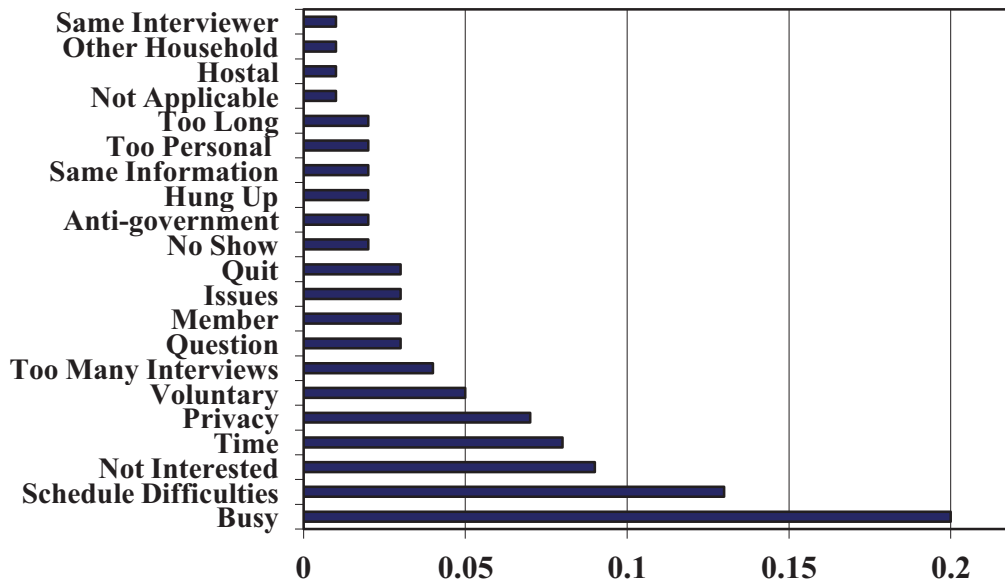

Figure 1; Relative frequency of concerns in Wave 1.

	Hostile	Time	Busy	Privacy
Hostile	0.719	0.057	-0.041	0.186
NotInt	0.55	0.229	0.139	0.223
Hangup	0.708	0	0.083	0.05
Samefr	-0.515	0.165	0.335	0.293
Time	0.176	0.506	0.388	0.01
SameInf	-0.069	0.832	-0.074	0.124
TooPers	-0.031	0.696	-0.107	0.34
TooMany	-0.01	0.87	0.011	-0.007
TooLong	0.03	0.839	0.161	-0.156
Quit	0.298	0.558	0.094	0.134
Busy	0.169	0.169	0.692	-0.056
NoShow	0.128	-0.019	0.727	-0.005
Schedule	-0.163	0.028	0.73	0.04
Privacy	0.051	0.153	0.025	0.73
AntiGov	0.21	0.054	-0.076	0.707
Question	0.04	0.096	0.089	0.527
NotApp	0.16	0.199	-0.112	0.539
OthHH	0.016	-0.059	0.194	0.676
	Hostile	Time	Busy	Privacy

Table 1; Pattern of concerns by factor for the first time in sample.

Four factors were identified based on concerns expressed for the first interview in the Contact History Instrument (Table 1). The "Hostile" factor included hostile behavior, hangups, "not interested", and in a negative relationship, wanting the same FR from the previous interview. The "Privacy" factor included concerns about privacy, expression of anti-government sentiment, not understanding the survey, not thinking the survey was applicable to them, and other household members advising the respondent not to participate. Although the "Time" and "Busy" factors may seem similar, looking at the concerns in each factor, "Time" seems more related to being overwhelmed or over burdened while the "Busy" seems to relate to the respondent not being able to find the time. In the subsequent interviews the concerns moved between "Time" and "Busy", with the factors becoming more correlated. It made more sense to combine them into a factor "Burden", which was comprised of "too busy", "takes too much time", "breaks appointment", and "scheduling difficulties". The fourth factor; "Other", contained the concerns that didn't fit in the other factors; "other", and "talk only to a specific household member".
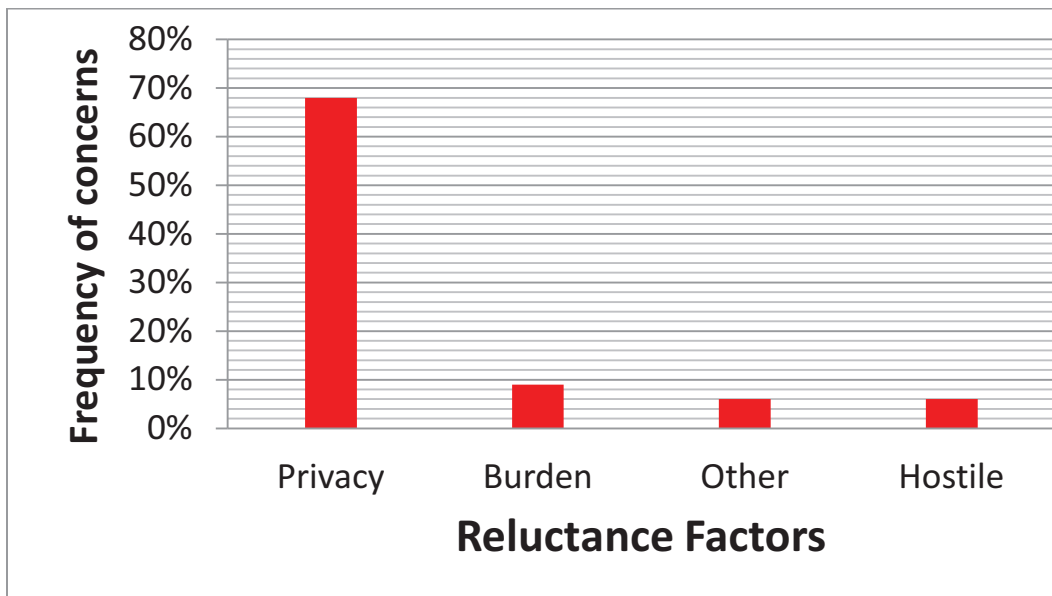

Figure 2; Frequency of reluctance factors.

Figure 2 shows the relative frequency of the 4 factors. Privacy was by far the most frequent type of concern factor expressed, with privacy comprising 68 percent of concerns, burden frequency was 10 percent, other and hostile 6 percent each. The concerns included in the privacy factor were the most common, but that factor had the lowest refusal rate. The Hostile factor on the other hand, made up of relatively rare concerns, had the highest refusal rate.
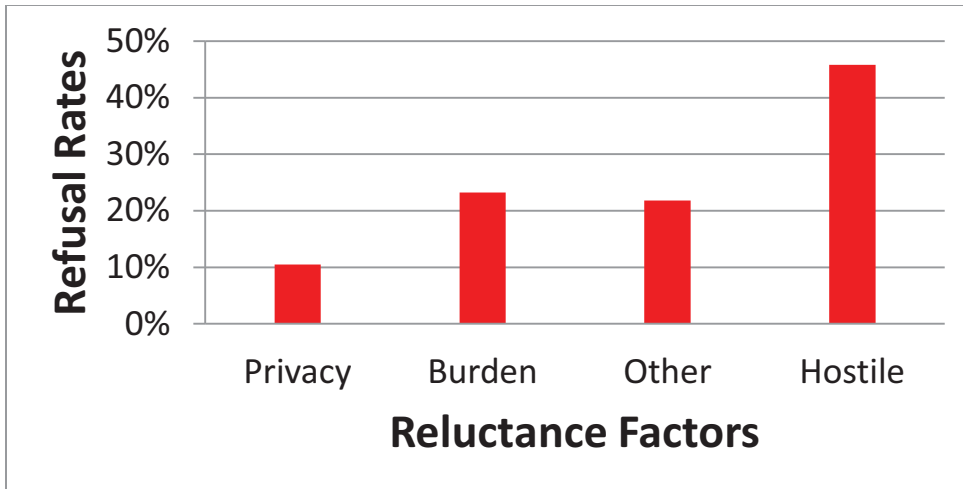
Figure 3; Concerns and refusal.

Figure 3 shows the refusal rate for the different groups of concerns. It is interesting to note that the frequency of the concerns mentioned is opposite the refusal rate.
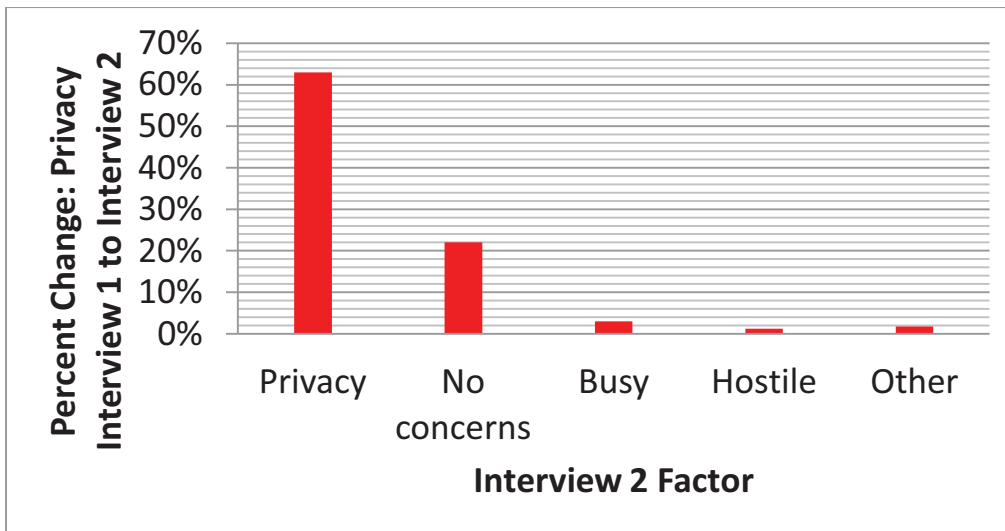

Figure 4; Percent change in Privacy concerns from first interview to second.

Figure 4 shows the change for those who expressed privacy concerns on the first interview. For example, 60 percent of the concerns included in the Privacy factor expressed in the first interview were also expressed in the second interview. Any of the concerns that were in the "privacy" factor would be counted as privacy concerns, so a household could change what they said, but still end up in the same factor. The next largest change is respondents who reported privacy concerns in the first interview and no concerns in the second. Less than 10 percent of respondents reported privacy concerns in the first interview and busy, hostile, or other concerns in the second.
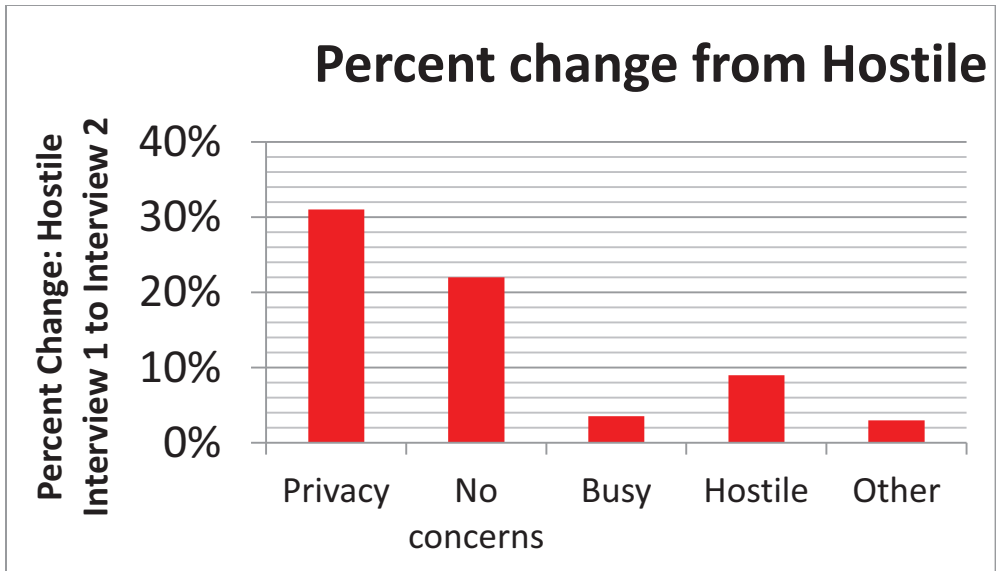
Figure 5; Hostile concerns change from first interview to second.

In contrast, Hostility concerns were not persistent from the first to second interview (Figure 5). Only 10 percent of respondents who reported Hostility concerns in the first interview reported them again in the second Most interview one Hostility concerns became Privacy concerns in the second interview. The next largest change is to "no concerns".
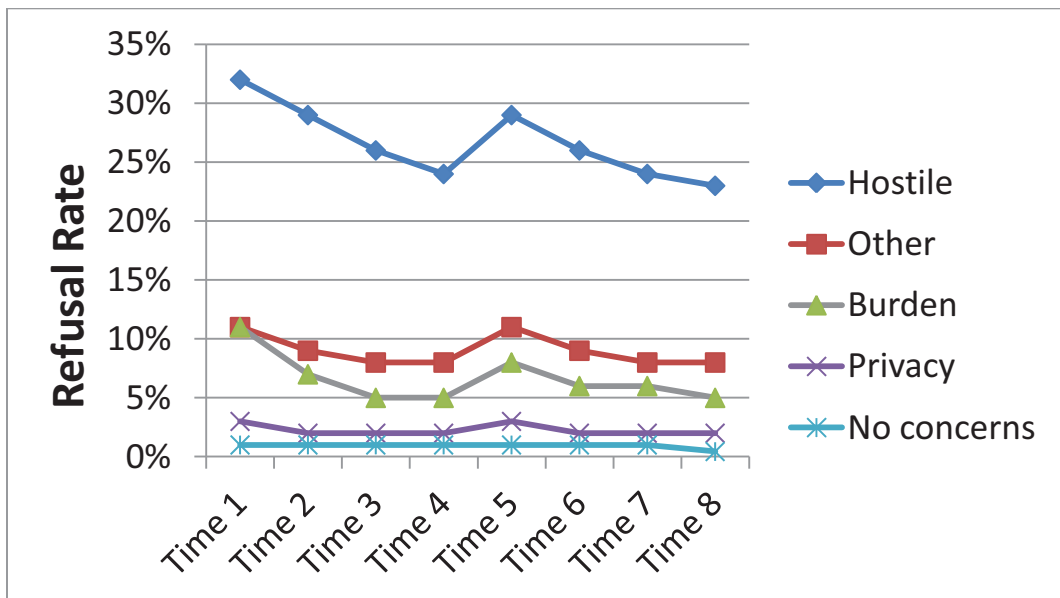

Figure 6; Refusal rates for different concerns over time.

Looking at the relationship between concerns and refusal over time (Figure 6) shows that Hostile concerns had the highest refusal rate across all interviews. However, because it was so rarely reported, it had little overall effect on the overall refusal rate. On the other hand, privacy was the most frequent concern, but had a very low refusal rate.

In the CPS, there is an 8 month time gap between interview 4 and 5. As would be expected, refusal rates increase in interview 5 as a result. Looking at the concerns, the increase in refusal rates between time 4 and 5, was similar for all concerns (even no concerns).

The increase in refusals was slightly higher for those who expressed hostility and burden concerns in the 4th interview; they were the more likely to refuse in the 5th, While those who expressed burden concerns and refused in the 4th were the most likely to cooperate in the 5th.
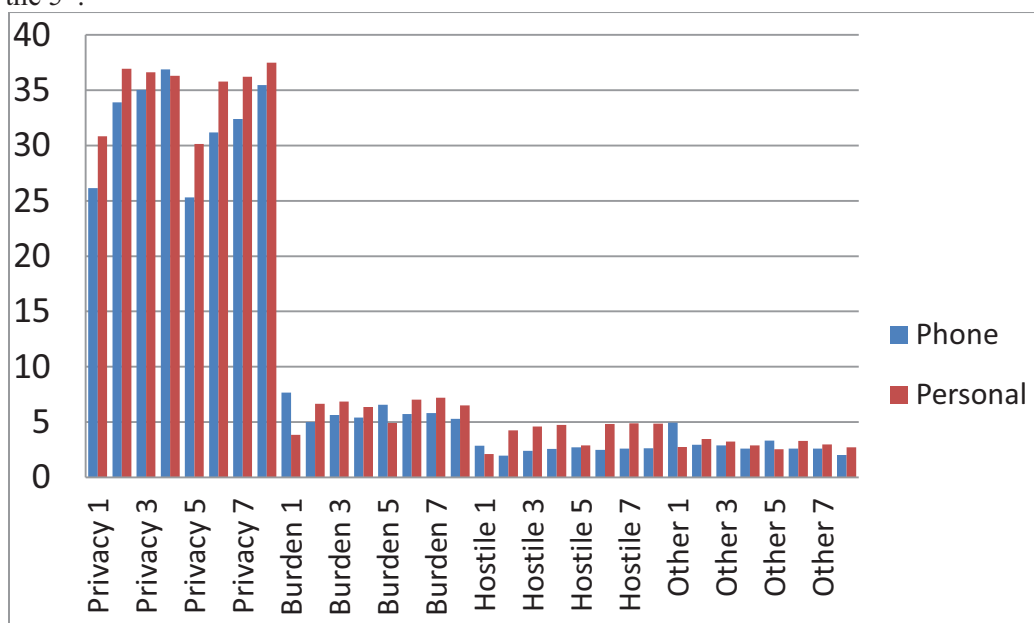

Figure 7; Mode differences in concerns.

Figure 7 shows the different concerns expressed when the interviewer collected data over the phone or in person. The first time in sample has about 20% of the interviews collecting data by phone. The 2nd through 4th are closer to 70%. The fifth time, after the 8 month break, is supposed to be in person, but 40% of the cases collect data by phone. It returns to 70% after that. I expected more concerns to be expressed by phone (since it's less engaging than a personal visit), but that wasn't the case. The first bars show that for successful interviews, in the first interview about 25% of the phone contacts had privacy concerns, while more than 30% had privacy concerns for personal visits. The concerns expressed stay in the same pattern except for the first and fifth times. The personal visits usually had proportionately more complaints. The fourth interview had a slightly higher rate of privacy concerns by phone than by personal visit, as did the burden concerns and other concerns for the 1st and 5th interview. Hostile concerns were slightly higher by phone for the first visit, then higher by personal visit for the rest.

5. Discussion

The CHI data showed factor patterns, which could describe broad areas of concern. Those patterns predicted refusal about as well as the individual concerns. Hostility and burden concerns were most related to refusal, but privacy concerns had little relationship. Since privacy was by far the most frequent concern, it had the potential to create high refusal rates. The interviewers seem adept at addressing those concerns enough to get the interviews.

The changes in concerns showed which concerns remained constant, and which were more situational. Privacy was most common, but didn't relate to nonresponse very well. Hostility appeared to be transitory, but with a high likelihood of refusal. Burden concerns interacted with the length of time between interviews. Those who expressed burden concerns and refused were more likely to participate after the long gap, but those who cooperated were more likely to refuse after the gap. The "rest" seemed to help change their minds.

There were a number of surprises in the data. Personal interaction lead to less hostility than phone interactions. The mode of data collection made little difference in the types of concerns expressed.

The CHI data is limited in that it only reflects the concerns expressed by respondents. Some of the most common concerns may mask the real reasons, for example, "busy" may hide concerns about privacy, which weren't expressed to the interviewer.
Replicating the models with another survey may help make the model more general, and give different insights which would help with the CPS.
Relating the changes in concerns to estimates from the survey may help explain phenomena like "rotation group bias" (Bailar, 1975).

References

BLS Handbook of Methods, http://www.bls.gov/opub/hom/
Bates, N. (2004). Contact Histories: A Tool for Understanding Attrition in Panel Surveys., A paper presented at the 59th Annual AAPOR Conference, May 11-13
Bates, N., Dahlhamer, J., and Singer, E. (2008). "Privacy Concerns, Too Busy, or Just Not Interested? Exploring the Relationship Between Doorstep Concerns and Nonresponse Bias", *Journal of Official Statistics*, Vol. 24.4, December.
Dahlhamer, J., Simile, C., and Taylor, B. (2006). "Exploring the Impact of Participant Reluctance on Data Quality in the National Health Interview Survey (NHIS)", paper presented at the Statistics Canada Symposium; Methodological Issues in Measuring Health.
Dixon, J. (2010). "Assessing Nonresponse Bias and Measurement Error Using Statistical Matching", presented at the Joint Statistical Meetings.
Dyer, W. (2004). Contact History Instrument (CHI).
http://www.blaiseusers.org/2004/papers/03.pdf

Henley, M., and Bates, N. (2006). "Using Call Records to Understand Response in Longitudinal Surveys", paper presented at the American Association for Public Opinion Research.

Maitland, A., Casas-Cordero, C. and Kreuter, F. (2009). An evaluation of nonresponse bias using paradata from a health survey. Pp. 2250-2255 in Proceedings of the Section on Survey Research Methods. Alexandria, VA: American Statistical Association.

MPLUS (Version 6.11). [Computer Software]. Los Angeles, CA: Muthén & Muthén.

U.S. Census Bureau. 2006. *Current Population Survey: Design and Methodology*. Technical Paper 66. Washington, DC: Government Printing Office. http://www.census.gov/prod/2006pubs/tp-66.pdf