RESEARCH REPORT SERIES
(Survey Methodology #2006-10)

Privacy Concerns and the Census Long Form: Some Evidence from Census 2000

Elizabeth Martin

U.S. Census Bureau
Washington, DC 20233

Report Issued: December 21, 2006

Privacy Concerns and the Census Long Form: Some Evidence from Census 2000

Elizabeth Martin, U. S. Census Bureau, Washington DC 20233[1]

Keywords: confidentiality, decennial census, data uses, respondent reluctance

I. Introduction

Declining public cooperation with the census (Miskura, 1992) and surveys has been attributed in part to increasing public concerns about privacy and confidentiality (Singer, Hippler, and Schwarz, 1992). Such concerns appear to be increasingly important barriers to public cooperation in recent censuses. Concern about privacy and mistrust of confidentiality predicted lower mail response in the 1990 Census, although it had not in 1980, controlling for demographic correlates (Fay, Bates, and Moore, 1991; Kulka et al., 1991; Singer, Mathiowetz, and Couper, 1993). Receipt of a long form increases concern about privacy (Fay, Bates, and Moore, 1991).

This paper examines trends in privacy and confidentiality attitudes during the course of Census 2000, analyzes concerns about long form questions, and explores what information might allay those concerns.

II. Methods

A series of 5 cross-sectional surveys was conducted between March 3 and April 13, 2000, by InterSurvey, Inc. (now Knowledge Networks) under the sponsorship of private foundations (Nie and Junn, 2000). Households were recruited using an RDD sample of household telephone numbers in areas with access to the Web TV network. Those agreeing to participate (57% did so; InterSurvey, 2000) were provided free hardware and Internet access, allowing surveys to be administered using a Web browser and to include multimedia content. Baseline data on non-census topics were collected in late February, then each household was assigned to one of five tracking surveys. A follow-up survey reinterviewed about 300 randomly selected respondents from each of the tracking surveys, as well as about 600 Rs who were only interviewed in the baseline survey. Table 1 summarizes the survey outcomes.

All surveys were self-administered using web TV. The same core instrument was used in each tracking survey. Question wording and order were kept constant so as to produce comparable data over time. (There were slight changes as the census progressed. After census forms were mailed out March 13-16, questions about receipt and handling of the form were added.) Analysis is based on 3 items asked in the tracking surveys:

"Now we would like to get your opinions about the census. Do you happen to agree or disagree with the following statements?

- My answers to the census could be used against me. Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree
- The Census Bureau's promise of confidentiality can be trusted.
- The census is an invasion of privacy."

Results are weighted to reflect sampling probabilities and to adjust for nonresponse. RASCHPLX is used to fit log-linear and Rasch models (Fay and Turner, 1989). Standard errors (where given) are calculated using simple jackknife methods (Fay, 1985, 1998).

Table 1. Survey Outcomes

Survey	Field Date	Completed Interviews	Response Rate
Baseline	2/25-3/8	7334	81%
Survey 1	3/3-9	993	83%
Survey 2	3/10-16	973	82%
Survey 3	3/23-31	719	61%
Survey 4	4/1-7	1004	58%
Survey 5	4/7-13	948	64%
Follow up	5/12 - 6/12	2079	68%

III. Limitations

The Census Bureau participated as a partner in the InterSurvey project in order to gain experience with Web surveys and obtain immediate feedback on whether the Census Bureau's promotion strategy was reaching the intended audiences. These surveys are not part of the formal evaluation of the Census 2000 advertising and promotional campaign, which is based on surveys conducted by NORC (Calder et al., 2001). Preliminary comparisons of InterSurvey and NORC results suggest the results are consistent (see Martin and Rivers, 2001), although more analysis remains to be done.

An important limitation is the low cumulative response rate of about 30%, taking into account recruitment success and response rates for baseline and tracking surveys. This is much lower than the Census Bureau would accept in its own surveys. The characteristics of respondents correspond fairly closely to population data from the Census Bureau's Current Population Survey, except that individuals with less than a high school education are under represented and voters are over represented: 69% of InterSurvey respondents of voting age in 1996 said they

[1]This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

voted in the 1996 presidential election, compared to 54.2% of voting age persons in the Current Population Survey voting supplement. Since voting is highly correlated with census participation, InterSurvey respondents over represent people likely to attend to and participate in the census.

InterSurvey excluded nontelephone households, and recruitment was less successful in households with unlisted telephones. The likely effect is a slight understatement of privacy concerns (see Martin, 2000).

On the other hand, because the surveys were not sponsored by the Census Bureau or any government agency, the data are probably less biased by the survey auspices than other surveys. Census Bureau surveys elicit more favorable opinions about the census and data confidentiality than surveys by outside organizations (Martin, 2000). To the extent that government sponsorship biases cooperation or responses, these surveys are less affected by this bias than other surveys.

The likely effects of nonresponse and sample biases are to produce estimates of participation that are too high and estimates of the extent of privacy concern that are too low. Although estimates of *levels* are biased, estimates of *trends* should not be, because the same core instrument was used in all 5 surveys, and nonresponse and sample biases affect them all to roughly the same degree.

IV. Results

Section A reports trends in the level of concern about privacy and confidentiality during Census 2000. In section B, a privacy and confidentiality scale is created based on three items. Section C uses the follow-up survey results to explore concerns about long form questions.

A. Trends during Census 2000

Table 2 presents results from the five tracking surveys. Consistent with past research, trust in the confidentiality of census data was high, and most people did not perceive the census as an invasion of privacy. However, the 10% who agreed or strongly agreed that their census answers could be used against them through mid-March increased significantly to 15-17% thereafter. Over 50% agreed or strongly agreed that "the Census Bureau promise of confidentiality can be trusted," with about a third unsure, throughout the census period. Finally, 10% agreed or strongly agreed in early March that "The census is an invasion of privacy," and this increased significantly to 20% in mid-April. Similar changes occurred in response to these items in surveys before and after the 1990 census (Fay, Bates, and Moore, 1991; Martin, 2000).

Table 2. Trends in privacy attitudes during Census 2000

	Mar 3-9	Mar 10-16	Mar 23-31	Apr 1-7	Apr 7-13
My answers to the census could be used against me					
Strongly agree	2%	3%	5%	4%	4%
Agree	9	7	12	10	11
Neither ...	31	30	29	28	31
Disagree	42	42	37	43	38
Strongly disagree	16	18	17	15	15
The Census Bureau promise of confidentiality can be trusted					
S. agree	14%	13%	15%	15%	15%
Agree	37	42	36	37	35
Neither	36	33	32	34	33
Disagree	12	8	14	12	13
S. disagree	3	4	4	3	5
The census is an invasion of privacy					
S. agree	3%	2%	6%	6%	6%
Agree	7	8	12	10	14
Neither	27	26	27	28	31
Disagree	47	46	36	43	35
S. disagree	16	18	19	13	14
N	993	973	719	1004	948

B. Creation of Privacy and Confidentiality Scale (PCS)

In order to determine whether the items form a scale, the Rasch measurement model (1960/1980) was fitted to the cross-classification of the three items. Each item was dichotomized by combining "agree" with "strongly agree," "disagree" with "strongly disagree," and dropping the middle category ("neither agree nor disagree"). Three log linear models were fitted to the data:

The model of independence posits that responses to each item are independent of responses to the other two.

The two-factor model posits that responses to each pair of items are associated, with no higher order interactions.

The Rasch model posits that a single underlying dimension accounts for associations among items. Conditional on the unobserved value of the underlying dimension, responses are independent.

Table 3. Goodness of fit for three log linear models

Model	Goodness of fit (Jackknifed χ^2)
Independence	χ^2=24.5, df=4, p<.0001
Two-factor	χ^2= .33, df=1, p=.22
Rasch	χ^2= -.75, df=2, p>.5

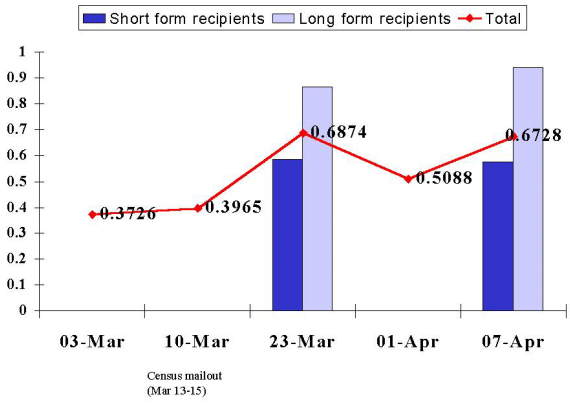
The model of independence is rejected, indicating the items are significantly correlated. The two factor model fits acceptably well. The Rasch model provides an excellent fit to the data (p>.5). Thus, the data are consistent with the hypothesis that a latent dimension accounts for responses to all three items. This implies they can be scaled. The scale value for each case equals the number of concerned responses, so the scale takes on values from 0 (no concern) to 3 (all three responses indicate concern). (In a Rasch scale, all items are given equal weight; Duncan, 1984.) Neutral responses are excluded from the scale, and cases with missing or neutral values for any item are dropped from the analysis.

For all respondents, the distribution of the PCS is:

73.5%	0 (no concern)
10.9%	1
6.4%	2
9.2%	3 (high concern)

Fig. 1 graphs mean PCS over the 5 surveys; the bar charts show results separately for long and short form recipients in the two surveys conducted after the mailout for which this information is available.

Fig. 1. Trend in mean level of PCS during Census 2000



The level of PCS rose significantly after the March 13-16 census mailout, especially among long form recipients.

C. Implications of Concerns about Long Form Questions

As was true in 1990, evidence suggests that privacy concern and receipt of a long form both reduced cooperation in Census 2000. Respondents in the follow-up interview were asked what they had done with their census forms. Table 4 presents their responses, controlling for type of form (SF=short form, LF=long form) they received and their prior level of privacy concern. (Due to the small sample sizes, values of the PCS are collapsed, with 0 = no privacy concern, and values of 1, 2, or 3 = some privacy concern.)

Table 4. Mail response behavior, by level of privacy concern and form type

"When you...filled out your census form, did you...	No concern		Some	
	SF	LF	SF	LF
Answer every Q relevant to my household	98%	82%	89%	59%
Left some blank that I didn't want to answer	1	10	8	26
Left some blank I didn't know how to answer	0	7	0	0
Did not return the census questionnaire	<1	2	2	14
Total	100%	100%	100%	100%
N	394	60	101	25

Note: Rs who did not receive a form, received both long and short forms, or did not work on the form are dropped, as are Rs who were not interviewed in a tracking survey

Table 4 shows that people who received a long form or were concerned about privacy were more likely to return an incomplete census form, or fail to return it at all. Results of fitting log-linear models to the four-way cross-classification of PCS (2 levels), form type, education, and mail response behavior show that the effects of privacy concern and long form receipt are highly significant (p<.002) and additive, controlling for the effects of education (not shown). (Education is negatively correlated with privacy concern and has a positive effect on mail response rates.) The results are consistent with research showing that privacy concerns reduced mail response in 1990 (Singer, Mathiowetz, and Couper, 1993; Fay, Bates and Moore, 1991).[2]

What concerned people about the long form? In the follow-up interview, long form recipients were asked, "Which ... questions, if any, did you feel reluctant or find difficult to answer?" Not surprisingly, income led the list by a wide margin, checked by 55% of long form recipients. Most questions were checked by fewer

[2]In the first two surveys, privacy concern was measured before the mailout, and in the last three, it was measured afterward. Privacy concern significantly predicts self-reported response behavior in both cases.

respondents: physical/mental disability (checked by 19%), employment (17%), race (16%), housing (13%), names of people who live here (10%), and education (10%); 38% checked "none of these."

Reluctant respondents were asked to select one or more reasons for their reluctance from the list shown in table 5 (columns sum to more than 100%).

Table 5. Reasons for reluctance to answer questions

	"Why were you reluctant or unable to answer the questions on ...?			
	Income	Disa-bility	Race	Hous-ing
1. Don't believe my answers will be kept confidential	46%[a]	42%[a]	20%[b]	24%[b]
2. None of the government's business	53%[bc]	60%[b]	43%[c]	79%[a]
3. There is no purpose for Q	51%[a]	44%[a]	58%[a]	56%[a]
4. Q is insulting	12%[b]	23%[a]	25%[a]	29%[a]
5. Difficult to provide info.	12%[a]	7%[a]	15%[a]	13%[a]
6. Don't like giving info. about others	22%[ab]	33%[a]	15%[b]	30%[a]
N	153	52	51	35

Note: Across each *row*, percentages sharing a letter in their superscripts are not significantly different (α=.10).

The most common reasons for reluctance were "it's none of the government's business" and "there is no purpose for the question." Different questions aroused reluctance for different reasons, however. Thus, in row 1, concern about confidentiality was a reason for 42-46% of respondents who were reluctant to provide income and disability information, while fewer (20-24%) gave this as a reason for reluctance to answer race and housing questions. In row 2, housing questions were perceived as "none of the government's business" by a much larger fraction of reluctant respondents than disability or income, probably due to the perceived inappropriateness of the government asking about toilets and plumbing. All four items were perceived as having no purpose by about half of reluctant respondents (row 3). Housing (as well as disability and race) were insulting to about a quarter of those who were reluctant to answer them, significantly more than the 12% reluctant to answer income for this reason (row 4). Item difficulty was not given as the main source of reluctance for any of the questions (row 5). Finally, 22-33% didn't like giving disability, housing, or income information about other people; this was less common as a reason for reluctance about race (row 6).

Thus, the profile of reasons for sensitivity varies among questions, although lack of purpose and "it's none of the government's business" are common to all of them.

If respondents understood the purpose for asking census questions, would they feel less reluctant to answer them? Some hints at an answer are provided by responses to the question, "What information would have helped you understand the census better?" Rs checked one or more items from the list shown in table 6 or typed a response (many volunteered, "didn't need any more information"). (This question was asked early in the survey, so Rs were not sensitized to concerns about particular questions.)

Table 6. Percentage who want different types of information, by level of privacy concern and form type

"What info. would have helped you understand the census better?"	% of all Rs ✓ing each	No concern		Some	
		SF	LF	SF	LF
Uses of the data	48% (1.12)	53%	37%	58%	30%
Reasons for particular Qs being asked	57 (1.19)	46	72	67	86
Procedures for protecting data	38 (1.57)	29	28	51	40
Didn't need more info. (vol.)	7 (.56)	13	6	5	1

Overall, the most commonly requested information (by 57%) is "reasons for particular questions being asked," followed by uses of the data (48%). This result mirrors the finding (from table 5) that the most common complaint about particular questions was their lack of purpose. Table 6 shows striking differences in information needs according to Rs' prior level of privacy concern and the type of form they received. Log linear analyses (not shown) of cross-classifications of responses by education, privacy concern, and form type confirm the following patterns observed in table 6:

Form type strongly influenced information needs. Long form recipients were *less* likely than short form recipients to want information about the uses of the data (40 versus 51%), but far *more* likely to want to know reasons for particular questions being asked (75 versus 54%). There was *no difference* in desire for information about procedures for protecting data.

Level of privacy concern influenced information needs. People who were concerned about privacy were *more* likely to want information about reasons for particular questions (67 versus 51%) and procedures for protecting

individual data (46 versus 28%) than those who were not concerned. There was *no difference* in desire for information about data uses.

Effects of form type and privacy concern were additive.

Education was significantly related to need for information. College graduates were *less* likely, and those with some college, *more* likely to want to know the reasons for questions, while those with a high school education or less were intermediate.

Would reluctance be reduced by providing the information respondents seem to be asking for? Reluctant respondents were given a brief explanation of the uses of the data for the question they were reluctant to answer, then asked how they felt about answering it (see Table 7).

Table 7. Willingness to answer questions after explanation

"If you knew that census information on income is used to allocate funds to school districts and help compute the Consumer Price Index, how would you feel about answering the question?"		
Total	100%	S. E.
Feel more like answering	31	5.00
No different, DK	63	4.77
Feel less like answering	6	2.34
"If you knew that info. on physical and mental disabilities is used to make sure that communities can provide adequate services for people with disabilities such as the elderly, how would you feel about answering the question?"		
More like answering	41	7.86
No different, DK	56	8.09
Less like answering	2	2.46
"If you knew that the Voting Rights Act requires that census info. on race be used to make sure all people have equal representation in Congress, State legislatures, and local governments, how would you feel...?"		
More like answering	20	6.28
No different, DK	68	8.57
Less like answering	12	5.75

For all three questions, most (56-68%) respondents were unmoved by an explanation of the purpose. For income and disability, a significantly larger fraction felt *more* like answering than felt *less* like answering, suggesting that providing explanations might yield a net improvement. For race, the 20% who felt *more* is not statistically different from the 12% who felt *less* like answering, so there is no net improvement. The questions vary somewhat in the beneficial effect of an explanation: the 41% who felt more like answering disability questions after they were explained is significantly larger than the 20% who felt more like answering the race question. This difference may occur because the explanation given for disability was more persuasive than the one given for race, or it may reflect differences in the extent and reasons for sensitivity of the two questions.

VI. Conclusions

As previous research has found, long form receipt and level of privacy concern both appear to have influenced cooperation with the census. People who received the long form or had concerns about privacy were less likely to fill it out completely and mail it back. This result is consistent with research conducted on the 1990 census.

Respondents expressed the greatest reluctance to answer long form questions (particularly income) but also some short form items, such as race and names. The perceived lack of purpose and "it's none of the government's business" were the most commonly cited reasons for reluctance, with some questions (income, disability) also arousing considerable concern about the confidentiality of the answers. These results are consistent with, and help explain, the heightened privacy and confidentiality concerns among long form recipients.

The results suggest that reluctance to answer particular questions is related to uncertainty or skepticism about why the questions are asked and how the data are used. For some questions, reluctance may be reduced somewhat by explaining the questions's purpose and use. This inference, however, is based on hypothetical questions, and may not accurately predict how people would really respond. Indeed, results from an experiment (Junn, 2000) caution against drawing conclusions about the likely effect of better explaining the uses of census data. She manipulated the introduction prior to administering a simulated census long form. The experimental treatments were (1) positive information about uses of census data (positive treatment), (2) a series of leading questions suggesting privacy concerns (negative treatment), or (3) no information (control). The effect of the positive treatment was mixed. Attitudes about the census were more positive than in the negative or control treatments, but respondents were actually more likely to skip some questions (annual cost of electricity, real estate taxes), and less likely to skip others (income) than the control group. Thus, the effect of explaining the reasons for census or survey questions on census cooperation remains unclear.

It is important to caution the reader that the results reported here may not be generalizable, based as they are on surveys with sample limitations and high rates of nonresponse. The results should be taken as suggestive, not definitive; it would be useful to replicate and expand the research using more general samples.

Clearly, there is much to learn about the public's need for information about the census, and the effects of providing it. On a smaller scale, respondents may have

unmet needs for information about the purposes and uses of data in ongoing demographic surveys, which usually ask far more extensive questions even than the census long form. The results suggest that further empirical investigations of the nature and reasons for respondents' reluctance and concerns about privacy, as well as research investigating possible strategies for addressing their concerns, would be worthwhile. This should include experimental research to develop alternative methods and messages that provide respondents the information they seem to want, to evaluate how effectively the information addresses respondents' concerns. It is important to note that research to understand and address privacy concerns should not be solely focused on the topics of privacy and confidentiality. The results reported here suggest that concerns about privacy related in the public's mind to questions and uncertainties about the purpose for asking census questions; therefore, research on these topics needs to address both questions of purpose and concerns about privacy.

ACKNOWLEDGMENT

Thanks to Yuki Ellis and Tom Mayer for helpful comments.

REFERENCES

Calder, B., Malthouse, E., Porras, J., and Wolter, K. (2001) Evaluation of the Census 2000 Partnership and Marketing Program. Report prepared under contract for the Census Bureau by NORC.

Duncan, O. D. (1984) "Rasch measurement: Further examples and discussion." Ch. 12 in C. F. Turner and E. Martin (eds.) Surveying Subjective Phenomena, Vol. II. New York: Russell Sage.

Fay, R. E. (1985) A Jackknifed Chi-Squared Test for Complex Samples, Journal of the American Statistical Association 80 (389):148-157.

Fay, R. E. (1998) VPLX Program Documentation. Washington DC: Bureau of the Census.

Fay, R. E., Bates, N., and Moore, J. (1991) "Lower mail response in the 1990 Census: A preliminary interpretation." Pp. 3-32 in Proceedings of the Annual Research Conference. Washington DC: Census Bureau.

Fay, R. E., Carter, W., Dowd, K. (1991) "Multiple causes of nonresponse: Analysis of the Survey of Census Participation." In Proceedings of the Social Statistics Section, American Statistical Association.

Fay, R. E., and Turner, C. F. (1989) RASCHPLX: Rasch Analysis for Contingency Tables with Complex Sample Designs, Program Documentation. Census Bureau.

InterSurvey. (2000) Census 2000 Tracking Study. Palo Alto CA: InterSurvey. June 29, 2000.

Junn, J. (2000) "The influence of negative political rhetoric: An experimental manipulation of Census 2000 participation." Paper presented at the annual meeting of the Midwest Political Science Association, 21 April 2000.

Kulka, R. A., Holt, N. A, Carter, W., and Dowd, K. (1991) "Self-reports of time pressures, concerns for privacy, and participation in the 1990 mail census." Pp. 33-54 Proceedings of the Annual Research Conference. Washington DC: Census Bureau.

Martin, E. (2000) "Changes in public opinion during the census." Paper presented at the Census Advisory Committee of Professional Associations, Virginia, Oct. 19, 2000.

Martin, E. and Rivers, E. (2001) "A look at some preliminary results of evaluating the Census 2000 Integrated Marketing Strategy." Paper presented at the annual meeting of the American Association for Public Opinion Research, Montreal, Quebec, May 18, 2001.

Miskura, S. M. (1992) "Mail response/return rates by type of form - 1970, 1980, and 1990." May 29, 1992, 2KS Memorandum Series. Census Bureau.

Nie, N. and Junn, J. (2000) Census Monitoring Study: Summary Findings. Palo Alto, CA: InterSurvey.

Rasch, G., (1960/1980), Probabilistic Models for Some Intelligence and Attainment Tests. Chicago IL: University of Chicago Press.

Singer, E., Hippler, H., and Schwarz, N. (1992) "Confidentiality assurances in surveys: Reassurance or threat?" Public Opinion Research 4:256-268.

Singer, E., Mathiowetz, N. A., and Couper, M. P. (1993) "The impact of privacy and confidentiality concerns on survey participation." Public Opinion Quarterly 57:465-482.