

# **The Public Use Microdata Samples of the U.S. Census: Research Applications and Privacy Issues**

**Minnesota Population Center  
and  
Inter-University Consortium for Political and Social Research  
Census 2000 Advisory Committee**

Prepared for  
Census 2000 Users' Conference on PUMS  
Hilton Alexandria Mark Center  
Alexandria, VA

May 22, 2000

This report was prepared by Steven Ruggles, Distinguished McKnight University Professor at the University of Minnesota and chair of the ICPSR Census 2000 Advisory Committee, with the assistance of Catherine Fitch, coordinator of the IPUMS project at the University of Minnesota and Matthew Sobek, managing editor of *Historical Statistics of the United States* at the University of California Riverside. The members of the committee include Ilona Einowski, Assistant Director of the University of California Data Archive; W. Reynolds Farley of the Institute for Social Research, University of Michigan; Halliman H. Winsborough, Interim Director, ICPSR and Bascom Professor of Sociology at the University of Wisconsin; Erik Austin, Director, Archival Development, ICPSR; and Peter Granda, Assistant Archival Director, ICPSR.

## Table of Contents

|                                                  |     |
|--------------------------------------------------|-----|
| <b>Summary</b>                                   | 1   |
| <b>A. Background and Significance</b>            |     |
| A1. What is Census Microdata?                    | 2   |
| A2. Historical Background                        | 3   |
| A3. Strengths of the PUMS                        | 4   |
| <b>B. Survey Results</b>                         |     |
| B1. Design and Execution                         | 5   |
| B2. Statistical Results                          | 6   |
| B3. Qualitative Results                          | 7   |
| B4. The Research Data Center Alternative         | 9   |
| <b>C. Conclusions and Recommendations</b>        | 10  |
| <b>Bibliography</b>                              | 12a |
| <b>Appendixes</b>                                |     |
| A. ICPSR Letter and Survey                       | 13  |
| B. Survey Responses: Quantitative Summary        | 16  |
| C. Survey Responses: Effects on Research         | 26  |
| D. Survey Responses: Thoughts on Confidentiality | 87  |
| E. Survey Responses: Other Comments              | 117 |

## Summary

The Census Bureau is considering significant reductions in the level of subject detail for the Public Use Microdata Samples (PUMS) of the 2000 Census in order to reassure the public about respondent confidentiality. In particular, the Bureau is considering a plan that would:

- group ages into five-year categories for persons aged 65 or older and reduce the topcode for age from 90 to 85;
- reduce the number of ancestry categories from 560 to 105;
- reduce the Hispanic origin categories from 206 to 23;
- reduce the number of identified occupational groups from 505 to 67;
- reduce the number of identified industry groups from 244 to 70;
- reduce the number of language categories from 393 to 83;
- eliminate 298 foreign countries of birth and substitute 14 continents and U.S. possessions.

To provide feedback to the Census Bureau on the potential impact of these changes on the academic research community, the Minnesota Population Center and the ICPSR Census 2000 Advisory Committee carried out a survey of 1,006 users of the Integrated Public Use Microdata Series (IPUMS). This report summarizes the reaction of researchers to the proposed changes and recommends an alternate strategy to ensure respondent confidentiality.

The research community expressed alarm at the proposal to reduce the level of detail in the PUMS. Survey respondents described hundreds of ongoing research projects that will have to be abandoned if the changes go forward in their current form. With remarkable and perhaps unprecedented unanimity, the academic community urges the Census Bureau to maximize historical compatibility of the 2000 census files and to avoid making precipitous decisions that will do permanent damage to social science research in the United States.

The ICPSR Census 2000 Advisory Committee recognizes that some census categories that appeared in the 1990 PUMS are too small, and may pose at least a theoretical, if not a practical, risk of disclosure. These gaps in the confidentiality safety net should be closed. There are some categories in the 1990 PUMS files that represent a national population of fewer than 1,000 persons. But the changes under consideration by the Census Bureau are excessive. Under the proposed system, most detailed categories would represent upwards of a million individuals in the general population. This thousand-fold reduction in the finest available level of detail would severely compromise the scientific utility of the data for academic and policy research.

The goal of ensuring that specific individuals cannot be identified in the census would be equally well served by imposing a minimum population threshold of 10,000 or 25,000 persons for every sensitive census category. Such a 10 to 25-fold reduction in detail for the smallest census categories for 1990 would result in a substantial loss of information, and some researchers will oppose such a drastic reform. Nevertheless, this alternative is far preferable to the radical change under consideration by the Bureau. If a minimum population threshold is adopted and the collapsing of categories is carried out with sensitivity to historical compatibility, the PUMS will remain the single most important source in American social science. It would also provide unprecedented security for respondent confidentiality.

We also recommend that the Bureau close one remaining confidentiality gap. There are a small percentage of cells in the tables of the summary files that contain a single individual when the tables are aggregated over an entire PUMA. We recommend that the Bureau suppress the PUMA codes for such individuals when they appear in the PUMS. This would foreclose the possibility of identifying any particular individual in the PUMS, but would not significantly reduce the precision of the samples because the number of excluded cases would be extremely small.

In the 36-year history of the census microdata samples, no respondent has ever been identified by anyone outside the Census Bureau. We consider such identification to be highly improbable even under the current standards. A minimum threshold for identification of sensitive population categories may be justified, however, on the grounds that it would reassure the public that the risk of disclosure is negligible.

The PUMS files are the crown jewels of American social science. They generate more population research than any other data source, and are the only available data source for a wide range of topics that bear on pressing policy issues. The ICPSR Census 2000 Advisory Committee strongly recommends that the Bureau proceed cautiously, and delay a final decision on the redesign of classifications until the long-form data are available for analysis. The current (as of 22 May 2000) classifications appear to be arbitrary, unsystematic, and hastily prepared. We further recommend that any reduction of detail maximize historical compatibility and be carried out in close consultation with the user community, even if that means a significant delay in the release of the data.

## **A. Background and Significance**

### **A1. What is Census Microdata?**

Most population data are available only in aggregated tabular form. The PUMS are microdata, which means that they provide information about individual persons and households. This makes it possible for researchers to create tabulations tailored to their particular questions. Since the PUMS files include nearly all the detail originally recorded by the census enumerations, users can construct a great variety of tabulations interrelating any desired set of variables. The flexibility offered by microdata is particularly important for historical research because the aggregate tabulations produced

by the Census Bureau are often not comparable across time, and until recently the subject coverage of census publications was limited.

Microdata do pose some limitations, however. Most important, for the period since 1940 census microdata are subject to strict confidentiality measures that limit their usefulness for some applications. The available samples for these years include no names, addresses or other potentially identifying information. To further ensure that no individuals can be identified, the Census Bureau limits the detail on place of residence, place of work, very high incomes, and several other variables. Most important, the microdata records for the period since 1940 identify no geographic areas with fewer than 100,000 inhabitants (250,000 in 1960 and 1970).

## **A2. Historical Background**

The founding fathers of the United States envisaged the census not simply as a tool for apportioning representatives, but rather as a means of gathering information about American society. James Madison, the primary author of the constitution, argued forcefully that the census should “embrace some objects besides the basic enumeration of the population.” Madison pressed for the enumeration of occupations, since it would provide the “kind of information . . . all legislatures had wished for” and “it would give . . . an opportunity of marking the progress of society, and distinguishing the growth of every interest . . . this would furnish ground for many useful calculations” (Magnuson 1995: 12-13). After the first census, Thomas Jefferson and his fellow members of the American Philosophical Society lobbied for the expansion of detail on age, birthplace, and occupation “in order to ascertain more completely the causes which influence life and health” and “the conditions and vocations of our fellow citizens” (Magnuson 1995:15).

From 1790 to 1950, the census expanded dramatically, both in terms of the questions asked and the number of tabulations produced. But the basic mode of access to census data remained unchanged: throughout that period, census results consisted of counts of the number of persons in each geographic area who had a particular characteristic or combination of characteristics. The advent of electronic computers, both within the Census Bureau and on university campuses, allowed a ground-breaking shift in this paradigm for 1960. In an effort to meet the needs of scholars who required specialized tabulations, the Census Bureau created a 1 in 1000 extract of the basic data tapes they had used to create tabulations for the published census volumes (U.S. Bureau of the Census 1964). To preserve confidentiality, the Census Bureau removed names, addresses, and other potentially identifying information.

The 1960 public use sample revolutionized the analysis of the American population and led to an explosion of new census-based research. Not only did it allow researchers to make tabulations tailored to their specific research questions, but it also allowed them to apply new methods to the analysis of census data, especially multivariate techniques. But the sample did have two significant limitations. First, the sample size was relatively small. The 1 in 1000 sample density yielded about 180,000 person records. Given the modest capacity of computers in 1964, this was a lot of cases, but as

researchers began to use the sample for detailed analysis of small population subgroups, its limitations became apparent. Second, the 1960 public use sample provided highly limited geographic information. To ensure confidentiality, the Census Bureau stripped off all information on places below the state level. This meant, for example, that it was impossible to extract a subsample of the New York City population.

Both of these problems were addressed by the 1970 public use samples. The 1 in 1000 density of the 1960 sample was increased dramatically; the Census Bureau provided six independent public use samples for 1970, each of which had a 1 in 100 density. Users who required an exceptionally large number of cases could combine the samples to obtain a six percent density, or about 12 million person records. In addition, the 1970 samples provided a variety of alternate geographic codes, although the Census Bureau still did not identify any places of less than 250,000 population.

In conjunction with the 1970 public use samples, the Census Bureau released a new version of the 1960 public use sample. The Bureau enlarged the sample density from 1 in 1000 to 1 in 100, and at the same time reorganized the coding schemes and record layouts to be compatible with the samples from 1970. This compatibility made it relatively easy for investigators to pool data from 1960 and 1970, and thus incorporate change into their analyses.

By the late 1970s, the public use samples for 1960 and 1970 had become one of the essential tools of American social scientists. It was in this climate that Halliman Winsborough and a group of others at the University of Wisconsin developed the idea of creating historical public use samples for earlier census years. They obtained funding from the National Science Foundation and contracted with the Census Bureau to create 1 in 100 samples for the censuses of 1940 and 1950 (U.S. Bureau of the Census 1984a, 1984b).

In addition to the census microdata samples covering the period 1940 through 1970, the Census Bureau has released samples for the 1980 and 1990 censuses, and plans to create a sample of the 2000 census (U.S. Bureau of the Census, 1982a). The 1980 and 1990 samples included significantly greater geographic and subject content detail than either the 1960 or 1970 public use samples. We now have a continuous series of Census Bureau microdata samples for six census years consisting of anonymized records spanning the period from 1940 through 1990, and the much-anticipated 2000 PUMS will give researchers a seventh sample in the series in 2002 or 2003.

The series of national census microdata files also extends backwards to the more distant past. The individual-level enumerations of the Census are released to the public after an interval of 72 years. Accordingly, historical demographers have created large national samples of these census enumerations for the period from 1850 to 1920.

The Integrated Public Use Microdata Series (IPUMS), created in 1995, harmonizes these early census samples with the seven Census Bureau samples covering the recent period (Ruggles and Sobek 1998). The combination of free and open access, a

user-friendly access system, and integrated comprehensive hypertext documentation has attracted many users to the IPUMS. Since 1995, the IPUMS project has distributed over two terabytes of IPUMS data to users around the world. The project is now distributing about 95 gigabytes of data per month, or an average of 130 megabytes per hour, 24 hours per day. The IPUMS automated data-extraction system has prepared approximately 20,000 custom extracts of IPUMS data since May 1996 for about 3,000 researchers around the world, and is now processing approximately 1,000 data extract requests per month. Even though the database has only been available for only five years, there is already a substantial body of IPUMS-based research. To date, the database has been used in 120 articles, 3 books, and 35 Ph.D. dissertations as well as hundreds of conference papers and research reports. Many of these articles appeared in leading journals such as the *American Economic Review*, the *American Sociological Review*, the *American Historical Review*, *Social Forces* and *Demography*. Most of these studies use the IPUMS to assess recent change in areas of current public policy concern, so compatibility of the 2000 PUMS with the earlier samples is critical.

### **A3. Strengths of the PUMS Files for Social Science Research**

The PUMS have become a mainstay of American social science. Among population scientists, the PUMS files are the single most important source of data. Among articles published in *Demography*—the leading journal in the field—PUMS data was used 36% more often than the next most important source during the period 1994-1998 (Ruggles 1998). The need for PUMS data is not limited to demographers. Economists represent the single largest group of IPUMS users, and according to Joshua Angrist, an economics professor at MIT, “In my view, the PUMS is the most important research dataset in economics.” PUMS data are also a mainstay of academic research in the fields of sociology and history, and they are an indispensable resource for urban planners and policy analysts.

Why are the PUMS files so widely used? The national census files incorporated in the existing IPUMS database have three key strengths: broad chronological scope, large sample populations and fine detail. Social scientists have increasingly recognized that we cannot understand contemporary social behavior without investigating processes of change. It is the relative continuity of Census Bureau classifications that allows for consistent comparisons across many decades. The PUMS is the only source of microdata that allows researchers to assess the effects of policy changes on the American population across significant periods of time or between cohorts, and the IPUMS design makes such investigations comparatively simple.

The second strength of the public use census files is their large size. The number of cases available for each census year ranges from the hundreds of thousands to the tens of millions. This allows the study of small and geographically dispersed population subgroups. Even the largest surveys are too small to allow analysis of small population subgroups such as Native Americans or particular occupational groups. Moreover, there are presently no national surveys large enough to be used for policy research at the municipal level. Finally, the large size of the PUMS files together with their national

coverage permits multi-level analyses of the effects of local conditions on individual and family behavior.

The third strength, fine detail, is an essential complement to both the chronological scope and large scale of the database. Without detailed categories, it would be impossible to make the datasets compatible over the long run. Moreover, it is the detail of the samples that allows us to identify small population subgroups, and to capitalize on the large scale of the PUMS. Therefore, without detailed population categories, the PUMS files would have limited application for either the analysis of change or the study of population subgroups.

## **B. Survey Results**

### **B1. Design and Execution**

To clarify the specific needs of researchers, the Minnesota Population Center conducted a survey of PUMS users during the week of May 9-16, 2000. We emailed a request to fill out the survey to approximately 1,400 persons who recently registered to use the IPUMS data extraction system. The email message and survey form are reproduced in Appendix A. We received 440 responses within 48 hours and 1,006 responses overall. It is difficult to calculate a precise response rate to our request, because we received many responses from persons other than the 1,400 we asked, but we estimate, based on name matching, that approximately 60% of the responses came from registered IPUMS users, yielding a response rate of approximately 43%.

149 survey responses, or 14.8 percent of the total, came from persons who indicated that they had used the PUMS for one or no studies. We eliminated these responses so that we could focus on experienced users of the data. Appendix B Tables 1-3 give descriptive statistics on the remaining respondents. Over two thirds of respondents were academic users, and most of the rest were policy researchers. The largest group of respondents is faculty members, followed by graduate students and nonacademic researchers. The most important fields of research are economics, demography, and sociology. The geographic distribution of respondents was very broad; about two-thirds of respondents came from institutions with five or fewer respondents. Table 4 lists the 28 institutions with six or more respondents. The list includes most of the leading institutions in demography, empirical economics, and public policy.

### **B2. Statistical Results**

Detailed survey results are presented in Appendix B Tables 5 through 17. The results suggest a remarkable consensus on the importance of the PUMS and the need to preserve as much detail as possible. At the time we prepared the survey, the particulars of the Census Bureau proposal were not yet available, but the survey results are nonetheless clear. Two-thirds of the faculty respondents indicated the PUMS were indispensable for research in their field, and 96 percent of these experienced researchers indicated that historical comparability was very useful or indispensable for work in their field.

The results indicate a clear preference for maximum detail, especially among academic researchers. The respondents were especially concerned with geography, income, and age variables, but approximately 90% of faculty researchers also said that a reduction in the detail available on occupation and race would have “catastrophic” or “very harmful” consequences for research in their field. Moreover, there is no consensus at all on which aspects of occupation should be preserved if the detail of occupation were to be substantially reduced; a plurality think the occupational classification should maximize comparability with earlier classification systems, but many others think the system should focus on socioeconomic status or type of work. It is therefore clear that no single broad classification could meet the needs of most researchers.

At the time the survey was designed, we did not know that the redesign of the PUMS called for eliminating all specific foreign countries of birth. If we had included a query on this proposal, we expect that it might have generated more concern than any other aspect of the proposed plan.

Perhaps the most revealing quantitative indicator was our question that posed a trade-off between promptness of data release and detail of categories. The question was worded as follows:

“The Bureau might make the decisions on the reduction of detail before the data are processed or they might wait until afterwards to allow analysis of the data to determine the need for confidentiality. The latter strategy might allow greater detail, but could result in a significant delay in release of data. I recognize that one would need to know more to make an informed decision, but in general, would you prefer a less detailed dataset released more promptly, or would you be prepared to wait if it might mean a more useful dataset?”

Over 90 percent of all respondents and 94% of faculty respondents indicated that they would rather wait for the release of the PUMS, if this delay offered some prospect of greater detail in the data. The consensus on this point is unequivocal.

### **B3. Qualitative Results**

The qualitative results were even more revealing than the quantitative ones. The first qualitative question asked respondents “Please comment below on ways in which reduction of PUMS detail might affect your research. Be specific as possible. For example, if you specialize in aging research, comment on the sorts of analyses that would be precluded by grouped age data.” We had hundreds of responses to this inquiry, and they are reproduced in full in Appendix C. Many respondents focus their research on particular population subgroups, which they might be unable to identify in a less-detailed sample. Some examples include studies of immigrant groups within metropolitan areas, the changing demographics of hotel and restaurants workers in San Francisco and Los Angeles, the legal profession, and specific Asian-Pacific populations. Policy makers in

particular expressed concern about the loss of the detail needed to study issues such as school voucher programs, welfare reform, residential segregation, urban poverty, and measures of social and economic inequality at the local level.

A few representative examples make the cost of the possible changes palpable. Robert Hauser, a demographer at Wisconsin, wrote that:

As far as I am concerned, elimination of the detail of age, race, ancestry, income, occupation, and geography would essentially eliminate the value of data from the long form. This is a shameful, cowardly, and ludicrous proposal. I hope it will disappear promptly and not be raised again.

The respondents detailed problems that would arise in several research areas. One of the most frequently mentioned topics was aging. Tony Deitz, a social gerontologist from Central Florida, focuses on the damage that would be done by grouping the ages of the aged:

I am a social gerontologist who focuses primarily upon socioeconomic and racial/ethnic minority status in my research. I am interested in and conduct research on a number of areas within gerontology. Recently, for example, I have written manuscripts that deal with mortality among older adults by ethnic group status. It is imperative that I be able to differentiate between people of different ages (i.e. 65 to 66, even). This is particularly important when we begin to look at age effects by ethnic group membership because such things as the mortality crossover effect is believed to occur at different ages for different ethnic groups (even, for example, between different Hispanic populations). So, given that the census represents our nation's most comprehensive survey of the population, it is very important that as much detail as possible be retained so that reliable, valid estimates and reports can be made.

Caroline Hoxby, a Harvard economist, stresses the use of the PUMS for the analysis of education:

For analyzing the rate of return to education, it is absolutely necessary to know a person's exact age, so that we can associate him with the regime (laws, funding) under which he attended school. We also need the exact age for estimating consistent wage equations.

Many respondents expressed concern about occupational categories and income information. Charles Nam, the developer of the Nam-Powers socioeconomic scale of occupations, writes:

The scale depends on a detailed occupation classification (although comparability with the occupational classification in earlier censuses is not important), as well as a fair amount of detail on education and income (determinants of the score for each detailed occupation). Without the detail for those three variables, it would not be possible to create the scale

(nor would it be possible for competing scales, e.g., Duncan SEI, to be constructed).

Donald Bogue, a Chicago demographer, writes about detail across several categories:

The need to introduce age cohorts into demographic analysis requires that age groups be as refined as possible. The occupational categories used in 1990 were about as minimal as acceptable. Income distributions need detail in order to meet the needs of a variety of studies, but also to adjust income distributions between censuses for comparability. Race and ethnicity are powerful variables in differentiating the population (especially for change over the 1990-2000 decade), and the detail could be as great as for 1990, and as comparable as possible.

We also asked respondents for their thoughts about the confidentiality issue. The voluminous results are reproduced in Appendix D. A few common themes run through the responses. Many researchers echo the comment of Halliman Winsborough of the University of Wisconsin: “The first PUMS file came out about 35 years ago. Has anyone been identified? I think not.” Bogue comments

I have never heard of a case of a breach of confidentiality using PUMS data. Is there a documented case on record? If so, what privacy was breached and how serious was it? Abuse of privacy of credit, bank, and other records is rampant, with little effort to control or regulate it. Persons seeking information about particular others would find PUMS about the least productive entry-point for gaining personal information.

Many also pointed out the impossibility of identifying individuals in the PUMS with certainty. Russell Davis, a graduate student at Louisiana State University, writes “I have it open and running right now, and I would challenge anyone to find someone they know.” It was also commonly noted that it does not make sense that anyone would attempt to identify someone in order to learn their income, since financial information on specific individuals is readily available on the Internet at modest cost.

A number of respondents discussed the politics of privacy and the census. For example, Maseo Suzuki, an economist at the Public Policy Institute of California, wrote

I am concerned that the Census Bureau may be under political pressure, since I have read newspaper accounts of Congressmen and/or Senators questioning the need for census data. I think that there is a lot of public concern about privacy, but this needs to be directed at its true source, private-sector internet businesses, and not redirected at academic researchers who would have no interest in trying to identify individuals.

Marcus Stanley, a graduate student in economics at Harvard, had this to add:

It seems to me this is a political stampede without much reason behind it. I know there have been a few people trying to whip up concern about the long form, but I don't know of any case where Census data has been used to identify individuals. . . it is just not a reasonable thing to try to do.

#### **B4. The Research Data Center Alternative**

One possible solution to the problem of confidentiality would be to restrict access to the PUMS to the five Census Bureau research data centers. Although we neglected to ask our respondents if this would be an acceptable solution, several commented on it anyway. For example, an anonymous faculty member wrote:

I am increasingly concerned about efforts to limit availability of high quality data such as PUMS to a broad spectrum of social scientists and other users, and to eliminate important information out of concern about privacy violations. The chances of such violations appear trivial. Insisting on secure-access sites is not a very appealing solution because doing so greatly disadvantages persons not located near such facilities. Restricting access or eliminating important information does not serve the broader purposes collecting such data - better understanding of social conditions and problems, information about the effect of public policies on social conditions, and insights into how to improve matters.

We believe that restricting detailed PUMS to the Research Data Centers would drastically reduce scholarly and policy use of the data. The cost of using a research data center is very high. For example, the two California centers charge \$3,125 per seat per month. Travel, per diem expenses, and housing costs could easily double this figure for anyone not located in a research data center city. Thus, a large research project, involving several person-years of data analysis—as is typical for dissertation projects—could easily cost \$200,000 simply to obtain access to the data. Major projects involving substantial manipulation, such as merging of confidential and public data sources, could cost many times more. We estimate that if we had carried out the IPUMS project at a Research Data Center, it would have added upwards of a million dollars to the cost. In fact, because of the logistics involved, we almost certainly would not have attempted it.

Because of the cost barrier, use of the centers will be restricted to a small number of well-funded researchers, and those researchers will be discouraged from use of the data by the inconvenience of travelling to another city. Even if the cost and travel issues could be overcome, the RDC's would have to be expanded at least 100-fold to accommodate the current number of PUMS users. Accordingly, we do not regard this as a viable strategy.

### **C. Conclusions and Recommendations**

The revision of the PUMS under consideration by the Census Bureau represents the most radical modification of the data series since it was introduced in 1964. It would undermine studies of such key topics as occupational structure and ethnicity, even as they are changing more rapidly than at any time in our history. If the proposed revision were to be carried out, it would represent the most damaging withdrawal of public data in the history of the Census Bureau, and would sharply reverse the 200-year tradition of

continuous improvement in Census Bureau data products which began at the urging of Madison and Jefferson.

We do not believe that there is an adequate scientific basis to justify such a drastic change. There is no evidence that the confidentiality protections of the past 36 years have been in any way inadequate. We know of not a single instance in which privacy has been breached. We find it implausible that anyone would turn to PUMS data files to attempt to uncover sensitive information about particular individuals. Even in the unlikely event that one could locate a unique exact match for a target individual, one could never be certain that the case actually represented that individual. The PUMS files are samples, and there will always be the possibility that another person exists, not included in the sample, who is also an exact match. The only exception to this is the comparatively rare cases in which a unique individual can be identified in a summary file, and that individual also appears in the PUMS. As proposed in the introduction, we suggest that the Bureau suppress the PUMA codes for such individuals, which would close this potential loophole.

Privacy in America is indeed under assault. There are now on the Internet some 500 web sites promising full investigative reports on any individual—including credit ratings, property records, marital status, and other information—for fees ranging from \$35 to \$150. Given this wealth of precise information readily available from a wide variety of private sources, it is simply not plausible that anyone would attempt to crack the PUMS security measures in order to obtain uncertain and outdated data. Thus, we think that fears of census disclosure are greatly exaggerated.

We believe that the risk to privacy posed by the PUMS must be weighed against the social cost of restricting access to information. That cost is very great. The proposed changes would cripple the efforts of social scientists to understand the radical social and economic changes taking place in American society. For example, the study of immigration will be devastated, as not a single foreign country of birth is identified under the proposed design of the 2000 PUMS. Similarly, we will be unable to gauge the transformation taking place in American occupational structure, since if the Census plan is adopted we will no longer be able to identify such key groups as carpenters, plumbers, waiters, bus drivers, licensed practical nurses, family child care providers, electrical engineers, and bartenders. Even lawyers could not be uniquely identified. And without single-year age information on the elderly, efforts to understand the implications of rapid population aging will be seriously handicapped, since there is simply no other data source large enough to examine the issue. The recodes proposed by the Bureau to date appear arbitrary and unsystematic. Shoshone Native Americans and persons of Albanian descent are to be identified, but not persons of Mexican or Chinese birth. There is no explanation of such anomalies, and no description of the criteria the Bureau is employing to define categories.

We recognize that perceptions are important; we have as much interest in ensuring a high census response rate as does the Bureau itself. But we see no historical evidence that confidentiality measures taken in the past have had any impact on census

response rates. When the Census Office first made responses confidential in the 1890 Census, for example, the undercount did not change. And when the threshold for identifying geographic areas in census microdata was lowered from 250,000 to 100,000 in 1980, the change did not even register with the public. It seems implausible that the technical changes proposed by the Bureau would improve response rates in the future. Nor would they satisfy opponents of the Census. We expect that the only real effect of the proposed changes would be to prevent social scientists and policy analysts from doing their research.

Still, we think it is reasonable to collapse small subject categories. We therefore propose that sensitive variable categories that represent small populations—perhaps those that represent fewer than 10,000 or 25,000 persons—be collapsed into larger categories. Such a criterion would significantly reduce the detail of information available in the PUMS. If the 25,000 population criterion were imposed in 1990, for example, it would have eliminated 110 occupational categories, 180 industry categories, 194 language categories, 148 birthplace categories, and 140 ancestry categories. It would nevertheless allow the great majority of ongoing investigations to proceed.

The ICPSR Census 2000 Advisory Committee is also very concerned about historical comparability. One option under consideration by the Bureau is to create a one-percent “national” sample with greater subject detail but sharply reduced geographic detail. We would endorse this proposal if the national sample were to replicate the 1990 subject area classifications precisely and identified ten-PUMA geographic areas incorporating at least one million persons. This file could then serve as a bridge between the PUMS files of the past and the reduced-detail subject classifications in the Census 2000 five-percent sample.

The Bureau has built a laudable reputation over the past 150 years for scientific integrity. It has repeatedly resisted political pressures that would compromise its mission. That reputation is one of the Bureau's greatest assets, and we believe it would be a mistake to sacrifice it merely for the sake of appearances. The Bureau should not allow unreasonable fears to undermine the social science and policy infrastructure of the nation. The risk to confidentiality in the current PUMS samples is minimal; with measured steps to collapse very small categories, the Bureau can eliminate any practical risk of disclosure without foreclosing essential scientific and policy research applications of the data.

## Bibliography

- Magnuson, Diana (1995). "The Making of a Modern Census: The United States Census of Population, 1790-1940." Ph.D. Dissertation, University of Minnesota.
- Ruggles, Steven (1998). "Demographic Data and Data Dissemination in the New Millennium." Paper presented at the annual meeting of the Association of Population Centers, Albany, November 1998.
- Ruggles, Steven, Matthew Sobek, et al. (1998) "IPUMS-98: Integrated Public Use Microdata Series" Minneapolis: Historical Census Project.
- U.S. Bureau of the Census. (1964). Census of population and housing, 1960 public use sample: one-in-one-thousand sample. Washington, D.C.: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1972). *Public Use Microdata Samples of Basic Records from the 1970 Census: Description and Technical Documentation*. Washington, D.C.: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1973). *Technical Documentation for the 1960 Public Use Microdata Sample*. Washington, D.C.: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1982). *Public Use Microdata Samples of Basic Records from the 1980 Census: Description and Technical Documentation*. Washington, D.C.: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1984a). *Census of Population, 1940: Public Use Microdata Sample Technical Documentation*. Washington, D.C.: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1984b). *Census of Population, 1950: Public Use Microdata Sample Technical Documentation*. Washington, D.C.: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1992). *Census of Population and Housing, 1990: Public Use Microdata Sample U.S. Documentation*. Washington, D.C.: U.S. Government Printing Office.

**From:** Steve Ruggles [<mailto:ruggles@hist.umn.edu>]  
**Sent:** Tuesday, May 09, 2000 4:02 PM  
**To:** [ipums-users@ipums.umn.edu](mailto:ipums-users@ipums.umn.edu)  
**Subject:** URGENT! Census 2000 PUMS  
**Importance:** High

Dear IPUMS users:

The Census Bureau is considering significant reductions in the level of subject and geographic detail for the 2000 PUMS files in order to enhance confidentiality. As the Principal Investigator of the IPUMS project and chair of the ICPSR Census 2000 Committee, I have been asked to attend a meeting at the Bureau later this month to report how these changes might affect users. No final decisions have yet been taken, but various measures are under discussion. Although I do not have the specifics of any proposals, I have heard that they include such steps as grouping ages for persons 65 or older, providing only broad occupational groupings, reducing the available detail on race, ancestry, and income, and reducing geographic detail.

I would like to get your feedback on these issues before I meet with the Bureau so I can fairly represent the concerns of IPUMS users. I would therefore appreciate it if you could complete the brief survey at <http://www.ipums.umn.edu/~ipums/survey.html>. I need the results as soon as possible. The survey will remain open until May 16th, but it will help greatly if you could complete the survey today or tomorrow. A high response rate will strengthen the credibility of the results. I appreciate your help on this, especially given the very short notice.

If you know of people who have used the IPUMS but who are not registered to use the data extraction system, I would appreciate it if you would forward this message to them.

If you do not want to receive email from IPUMS and would like us to remove you from our list of registered users, please email [ipums@hist.umn.edu](mailto:ipums@hist.umn.edu) and ask to be removed, and accept my apologies for clogging your inbox.

Yours,

Steven Ruggles

~~~~~  
Steven Ruggles  
Minnesota Population Center  
537 Heller, University of Minnesota  
271 19th Avenue South  
Minneapolis, MN 55455  
FAX: (612) 624-7096  
PHONE: (612) 624-5818  
~~~~~

(ip'-ums)

## Survey of IPUMS Users on Census 2000

### Personal Information

☐ Check here if you want your responses to be kept confidential

First Name

Last Name

Institution

E-mail address (required to eliminate duplicate responses)

Field

- ☐ Demography
- ☐ Economics
- ☐ History
- ☐ Sociology
- ☐ Geography
- ☐ Other academic
- ☐ Government
- ☐ News Media
- ☐ Private Industry
- ☐ Other nonacademic

Educational Status

- ☐ Faculty
- ☐ Academic researcher
- ☐ Support staff
- ☐ Postdoc
- ☐ Graduate student
- ☐ Undergraduate student
- ☐ High school student
- ☐ Nonacademic

What sort of research do you primarily use PUMS data for?

- ☐ Academic research
- ☐ Policy research
- ☐ Market research
- ☐ Journalistic research
- ☐ Other research

1. How important is the PUMS for your research in your field?

- ☐ indispensable   ☐ very useful   ☐ useful   ☐ somewhat useful   ☐ not useful

2. How many times have you used the PUMS data in your research?

- ☐ never   ☐ once   ☐ 2-5 times   ☐ 6-10 times   ☐ over 10 times

3. How important is it to maximize data compatibility across census years?

- ☐ indispensable   ☐ very useful   ☐ useful   ☐ somewhat useful   ☐ not useful

4. If the 2000 PUMS were to reduce detail in each of the following subject areas, how harmful would it be for YOUR research in your field?

age

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

income

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

race

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

ancestry

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

occupation

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

geography

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

5. If the 2000 PUMS were to reduce detail in each of the following subject areas, how harmful would it be for OTHER research in your field?

**age**

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

**income**

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

**race**

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

**ancestry**

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

**occupation**

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

**geography**

- ☐ catastrophic
- ☐ very harmful
- ☐ harmful
- ☐ somewhat harmful
- ☐ not harmful

6. What is the minimum number of occupational categories necessary for research in your field?

- ☐ 505 titles (same as 1980 and 1990)
- ☐ 160 titles (same as 1950)
- ☐ 25 broad categories
- ☐ Occupations are not needed

7. If only a broad occupational classification can be provided, which aspect of occupation should be stressed?

- ☐ Socioeconomic status (e.g. simplified Duncan SEI score)
- ☐ Type of work (e.g. clerical sales, manufacturing, etc.)
- ☐ Simplified industrial classification
- ☐ Maximize historical comparability
- ☐ Maximize international comparability
- ☐ Other (specify in comment section below)
- ☐ No opinion

8. What level of geographic detail is required for research in your field?

- ☐ PUMAs of 100,000 population (same as 1980 and 1990)
- ☐ PUMAs of 250,000 population (same as 1970)
- ☐ Metropolitan areas
- ☐ States
- ☐ Regions
- ☐ Geographic codes are not needed

9. The Bureau might make the decisions on the reduction of detail before the data are processed or they might wait until afterwards to allow analysis of the data to determine the need for confidentiality. The later strategy might allow greater detail, but could result in a significant delay in release of data. I recognize that one would need to know more to make an informed decision, but in general, would you prefer a less detailed dataset released more promptly, or would you be prepared to wait if it might mean a more useful dataset?

- ☐ I would be prepared to wait
- ☐ I would prefer less detail sooner
- ☐ No opinion

10. Please comment below on ways in which reduction of PUMS detail might affect your research.

Be specific as possible. For example, if you specialize in aging research, comment on the sorts of analyses that would be precluded by grouped age data.

A large rectangular text input area with a light gray border and a vertical scrollbar on the right side.

11. Do you think it is plausible that PUMS data would be used to identify the characteristics of specific individuals?

☐ Yes ☐ No

12. If you have any thoughts or suggestions on issues of privacy or confidentiality in the PUMS, please elaborate:

A large rectangular text input area with a light gray border and a vertical scrollbar on the right side.

13. Other comments:

A large rectangular text input area with a light gray border and a vertical scrollbar on the right side.

14. Would you like us to keep you informed about developments on the 2000 PUMS files as they unfold?

☐ Yes ☐ No

Submit Completed Survey

[| Home](#) | [What is the IPUMS?](#) | [What's new in the IPUMS?](#) |  
[| Get IPUMS data](#) | [Get IPUMS documentation](#) | [Citation and use](#) | [Revision history](#) |  
[| IPUMS research](#) | [Staff](#) | [Links](#) | [A note to genealogists](#) | [Site map](#) |

## Survey Results: Quantitative Summary

**Table 1**  
**Distribution of Respondents by Field**

| <b>Field</b>      | <b>%</b> |
|-------------------|----------|
| Demography        | 12.0     |
| Economics         | 30.1     |
| History           | 5.5      |
| Sociology         | 20.2     |
| Geography         | 2.3      |
| Other academic    | 11.9     |
| Government        | 9.8      |
| News media        | 1.4      |
| Private industry  | 3.4      |
| Other nonacademic | 3.3      |

**Table 2**  
**Distribution of Respondents by Position**

| <b>Position</b>     | <b>%</b> |
|---------------------|----------|
| Faculty             | 36.7     |
| Academic researcher | 10.7     |
| Support staff       | 4.1      |
| Postdoc             | 1.7      |
| Grad student        | 26.3     |
| Undergrad           | 2.5      |
| Nonacademic         | 18.0     |

**Table 3**  
**Distribution of Respondents by Use of Data**

| <b>Use of Data</b>    | <b>%</b> |
|-----------------------|----------|
| Academic research     | 64.3     |
| Policy research       | 21.9     |
| Market research       | 4.2      |
| Journalistic research | 1.9      |
| Other                 | 7.7      |

**Table 4**  
**Institutional Affiliation of Survey Respondents**

| <b>Institution</b>                     | <b>Frequency</b> |
|----------------------------------------|------------------|
| Harvard University                     | 34               |
| The Urban Institute                    | 24               |
| University of Wisconsin - Madison      | 24               |
| University of Chicago                  | 22               |
| University of Minnesota                | 20               |
| University of California - Berkeley    | 20               |
| Princeton University                   | 18               |
| University of Washington               | 16               |
| University of California - Los Angeles | 14               |
| Rutgers University                     | 13               |
| Penn State University                  | 11               |
| Columbia University                    | 10               |
| University of Pennsylvania             | 10               |
| University of Michigan                 | 10               |
| Massachusetts Institute of Technology  | 9                |
| Ohio State University                  | 8                |
| Stanford University                    | 8                |
| University of Southern California      | 8                |
| University of Texas - Austin           | 8                |
| Vanderbilt University                  | 8                |
| University of Maryland                 | 8                |
| Florida State University               | 7                |
| Indiana University                     | 7                |
| University of Illinois at Chicago      | 7                |
| Arizona State University               | 6                |
| Louisiana State University             | 6                |
| Cornell University                     | 6                |
| <b>Subtotal of top respondents</b>     | <b>342</b>       |
| Institutions with 5 or fewer responses | 664              |
| <b>Total responses</b>                 | <b>1006</b>      |

**Table 5**  
**How Important Is The PUMS For Your Research In Your Field?**

|                  | Indispensible | Very Useful | Useful | Somewhat Useful | Not Useful | <i>N</i> |
|------------------|---------------|-------------|--------|-----------------|------------|----------|
| Faculty          | 66.9          | 29.4        | 2.7    | 1.0             | 0.0        | 296      |
| Acad. Researcher | 55.2          | 36.8        | 6.9    | 1.1             | 0.0        | 87       |
| Graduate Student | 50.9          | 42.0        | 6.2    | 0.9             | 0.0        | 226      |
| Support Staff    | 45.5          | 39.4        | 12.1   | 3.0             | 0.0        | 33       |
| Undergraduate    | 45.0          | 30.0        | 20.0   | 5.0             | 0.0        | 20       |
| Nonacademic      | 36.1          | 46.5        | 13.9   | 3.5             | 0.0        | 144      |

**Table 6**  
**How Important Is It To Maximize Data Compatibility Across Census Years?**

|                  | Indispensible | Very Useful | Useful | Somewhat Useful | Not Useful | <i>N</i> |
|------------------|---------------|-------------|--------|-----------------|------------|----------|
| Faculty          | 64.5          | 31.8        | 3.0    | 0.7             | 0.0        | 296      |
| Acad. Researcher | 59.8          | 32.2        | 5.7    | 2.3             | 0.0        | 87       |
| Graduate Student | 60.2          | 35.0        | 3.1    | 1.3             | 0.4        | 226      |
| Support Staff    | 45.5          | 42.4        | 6.1    | 6.1             | 0.0        | 33       |
| Undergraduate    | 40.0          | 45.0        | 10.0   | 5.0             | 0.0        | 20       |
| Nonacademic      | 33.8          | 44.1        | 14.5   | 6.9             | 0.7        | 145      |

**Table 7**  
**What Is The Minimum Number of Occupational Categories**  
**Necessary for Research In Your Field?**

|                  | 505<br>Titles | 160<br>Titles | 25<br>Categories | Not<br>Needed | <i>N</i> |
|------------------|---------------|---------------|------------------|---------------|----------|
| Faculty          | 58.5          | 29.6          | 10.5             | 1.4           | 294      |
| Acad. Researcher | 54.4          | 16.5          | 24.1             | 5.1           | 85       |
| Graduate Student | 48.8          | 35.4          | 13.4             | 2.4           | 225      |
| Support Staff    | 63.3          | 20.0          | 13.3             | 3.3           | 32       |
| Undergraduate    | 35.3          | 41.2          | 17.6             | 5.9           | 19       |
| Nonacademic      | 44.9          | 23.9          | 23.2             | 8.0           | 142      |

**Table 8**  
**If Only A Broad Classification Can Be Provided,**  
**Which Aspect of Occupation Should Be Stressed?**

|                  | Socioecon.<br>Status | Type of<br>Work | Indust.<br>Classif. | Max.<br>Hist.<br>Compar. | Max.<br>Internat'l<br>Compar. | Other | No<br>Opinion |
|------------------|----------------------|-----------------|---------------------|--------------------------|-------------------------------|-------|---------------|
| Faculty          | 15.9                 | 31.2            | 4.0                 | 39.5                     | 2.5                           | 2.9   | 4.0           |
| Acad. Researcher | 20.0                 | 28.8            | 11.3                | 23.8                     | 5.0                           | 2.5   | 8.8           |
| Graduate Student | 23.4                 | 26.3            | 6.8                 | 30.2                     | 2.4                           | 2.9   | 7.8           |
| Support Staff    | 10.7                 | 42.9            | 10.7                | 21.4                     | 0.0                           | 3.6   | 10.7          |
| Undergraduate    | 27.8                 | 44.4            | 0.0                 | 5.6                      | 11.1                          | 0.0   | 11.1          |
| Nonacademic      | 8.5                  | 45.1            | 9.2                 | 24.6                     | 0.7                           | 0.7   | 11.3          |

**Table 9**  
**What Level Of Geographic Detail Is Required For Research In Your Field?**

|                  | PUMA<br>100,000 | PUMA<br>250,000 | Metro<br>Areas | States | Regions | Not<br>Necess. | <i>N</i> |
|------------------|-----------------|-----------------|----------------|--------|---------|----------------|----------|
| Faculty          | 80.4            | 6.2             | 7.2            | 4.5    | 1.0     | 0.7            | 291      |
| Acad. Researcher | 80.5            | 1.1             | 11.5           | 4.6    | 1.1     | 1.1            | 87       |
| Graduate Student | 62.4            | 6.6             | 19.0           | 7.1    | 4.0     | 0.9            | 226      |
| Support Staff    | 90.3            | 0.0             | 6.5            | 3.2    | 0.0     | 0.0            | 31       |
| Undergraduate    | 31.6            | 10.5            | 36.8           | 10.5   | 10.5    | 0.0            | 19       |
| Nonacademic      | 79.5            | 2.7             | 12.3           | 5.5    | 0.0     | 0.0            | 146      |

**Table 10**  
**Would You Be Willing To Wait To Get A More Detailed Dataset?**

|                  | Rather<br>Wait | Less<br>Detail | <i>N</i> |
|------------------|----------------|----------------|----------|
| Faculty          | 94.2           | 5.8            | 278      |
| Acad. Researcher | 88.0           | 12.0           | 83       |
| Graduate Student | 94.0           | 6.0            | 218      |
| Support Staff    | 86.7           | 13.3           | 30       |
| Undergraduate    | 81.3           | 18.8           | 16       |
| Nonacademic      | 84.6           | 15.4           | 136      |

**Table 11**  
**How Harmful Would A Reduction Of Detail Be For Research In Your Field?**  
**By Variable**  
**[Academic Faculty]**

| Variable   | Catastrophic | Very Harmful | Harmful | Somewhat Harmful | Not Harmful |
|------------|--------------|--------------|---------|------------------|-------------|
| Age        | 51.3         | 34.3         | 11.9    | 2.2              | 0.3         |
| Income     | 58.1         | 31.4         | 8.5     | 1.6              | 0.5         |
| Race       | 54.4         | 31.9         | 8.8     | 3.3              | 1.6         |
| Ancestry   | 34.8         | 26.7         | 18.8    | 13.0             | 6.8         |
| Occupation | 49.5         | 34.5         | 9.8     | 4.7              | 1.4         |
| Geography  | 57.7         | 31.7         | 7.7     | 1.9              | 1.1         |

**Table 12**  
**How Harmful Would A Reduction Of Detail Be For Research In Your Field?**  
**Age Variable**

| Variable        | Catastrophic | Very Harmful | Harmful | Somewhat Harmful | Not Harmful | N   |
|-----------------|--------------|--------------|---------|------------------|-------------|-----|
| Faculty         | 60.1         | 30.9         | 6.9     | 2.1              | 0.0         | 288 |
| Acad. Reseacher | 37.8         | 40.2         | 19.5    | 1.2              | 1.2         | 82  |
| Grad. Student   | 49.5         | 34.5         | 14.5    | 1.4              | 0.0         | 220 |
| Support Staff   | 27.6         | 44.8         | 24.1    | 3.4              | 0.0         | 29  |
| Undergraduate   | 31.6         | 42.1         | 5.3     | 15.8             | 5.3         | 19  |
| Non-academic    | 31.6         | 45.9         | 15.0    | 6.0              | 1.5         | 133 |

**Table 13**  
**How Harmful Would A Reduction Of Detail Be For Research In Your Field?**  
**Income Variable**

| Variable        | Catastrophic | Very Harmful | Harmful | Somewhat Harmful | Not Harmful | N   |
|-----------------|--------------|--------------|---------|------------------|-------------|-----|
| Faculty         | 64.0         | 29.4         | 6.3     | 0.3              | 0.0         | 286 |
| Acad. Reseacher | 46.9         | 35.8         | 13.6    | 3.7              | 0.0         | 81  |
| Grad. Student   | 59.5         | 28.8         | 9.0     | 1.8              | 0.9         | 222 |
| Support Staff   | 34.5         | 48.3         | 13.8    | 3.4              | 0.0         | 29  |
| Undergraduate   | 36.8         | 47.4         | 5.3     | 5.3              | 5.3         | 19  |
| Non-academic    | 35.8         | 50.7         | 10.4    | 3.0              | 0.0         | 134 |

**Table 14**  
**How Harmful Would A Reduction Of Detail Be For Research In Your Field?**  
**Race Variable**

| Variable        | Catastrophic | Very Harmful | Harmful | Somewhat Harmful | Not Harmful | N   |
|-----------------|--------------|--------------|---------|------------------|-------------|-----|
| Faculty         | 60.1         | 29.7         | 8.1     | 1.8              | 0.4         | 283 |
| Acad. Reseacher | 46.9         | 35.8         | 11.1    | 4.9              | 1.2         | 81  |
| Grad. Student   | 55.2         | 29.9         | 8.1     | 4.5              | 2.3         | 221 |
| Support Staff   | 26.7         | 43.3         | 20.0    | 6.7              | 3.3         | 30  |
| Undergraduate   | 36.8         | 52.6         | 0.0     | 0.0              | 10.5        | 19  |
| Non-academic    | 29.3         | 41.4         | 15.0    | 6.0              | 8.3         | 133 |

**Table 15**  
**How Harmful Would A Reduction Of Detail Be For Research In Your Field?**  
**Ancestry Variable**

| Variable        | Catastrophic | Very Harmful | Harmful | Somewhat Harmful | Not Harmful | N   |
|-----------------|--------------|--------------|---------|------------------|-------------|-----|
| Faculty         | 41.3         | 26.9         | 18.2    | 10.1             | 3.5         | 286 |
| Acad. Reseacher | 21.5         | 26.6         | 22.8    | 16.5             | 12.7        | 79  |
| Grad. Student   | 35.2         | 26.0         | 19.2    | 12.3             | 7.3         | 219 |
| Support Staff   | 13.3         | 30.0         | 20.0    | 26.7             | 10.0        | 30  |
| Undergraduate   | 21.1         | 26.3         | 5.3     | 26.3             | 21.1        | 19  |
| Non-academic    | 7.6          | 22.7         | 23.5    | 22.7             | 23.5        | 132 |

**Table 16**  
**How Harmful Would A Reduction Of Detail Be For Research In Your Field?**  
**Occupation Variable**

| Variable        | Catastrophic | Very Harmful | Harmful | Somewhat Harmful | Not Harmful | N   |
|-----------------|--------------|--------------|---------|------------------|-------------|-----|
| Faculty         | 54.2         | 34.9         | 8.1     | 2.8              | 0.0         | 284 |
| Acad. Reseacher | 43.0         | 31.6         | 13.9    | 10.1             | 1.3         | 79  |
| Grad. Student   | 51.8         | 32.4         | 10.8    | 2.7              | 2.3         | 222 |
| Support Staff   | 20.0         | 46.7         | 6.7     | 23.3             | 3.3         | 30  |
| Undergraduate   | 26.3         | 47.4         | 10.5    | 5.3              | 10.5        | 19  |
| Non-academic    | 20.8         | 45.4         | 17.7    | 10.8             | 5.4         | 130 |

**Table 17**  
**How Harmful Would A Reduction Of Detail Be For Research In Your Field?**  
**Geography Variable**

| Variable        | Catastrophic | Very<br>Harmful | Harmful | Somewhat<br>Harmful | Not<br>Harmful | <i>N</i> |
|-----------------|--------------|-----------------|---------|---------------------|----------------|----------|
| Faculty         | 64.9         | 28.1            | 5.9     | 0.7                 | 0.3            | 288      |
| Acad. Reseacher | 55.0         | 27.5            | 12.5    | 3.8                 | 1.3            | 80       |
| Grad. Student   | 52.5         | 36.7            | 7.2     | 2.3                 | 1.4            | 221      |
| Support Staff   | 46.7         | 40.0            | 10.0    | 3.3                 | 0.0            | 30       |
| Undergraduate   | 36.8         | 31.6            | 15.8    | 5.3                 | 10.5           | 19       |
| Non-academic    | 50.0         | 37.1            | 11.4    | 1.5                 | 0.0            | 132      |

**Responses to Question 10: Please comment on ways in which reduction of PUMS detail might affect your research. Be specific as possible. For example, if you specialize in aging research, comment on the sorts of analyses that would be precluded by grouped age data.**

**Confidential Response**

I am particularly interested in data available by American Indian reservation areas and trust lands, as Census is really the only place where data are available at that level of analysis.

**Alair MacLean, University of Wisconsin - Madison, Sociology (Graduate Student)**

I do research on occupational change over time. Providing only broad occupational groupings would make it impossible for me to update the research I've already done, which depends on comparability based on the IPUMS recode to the three digit occupational categories of 1950.

I also do research on education and immigration, with specific reference to metropolitan areas. At larger geographic levels, differences are often smoothed over or not apparent. If this level of geographic detail were no longer available, I would not be able to do historical comparisons.

**J. Wesley Null, The University of Texas at Austin, Other Academic (Graduate Student)**

I conduct research on the ages and levels of educational attainment for teachers. Reductions in the level of specificity for each of these categories would severely hamper the levels of analysis that I could complete.

**Confidential Response**

The PUMS is the best source of reliable income data by age, race, and geography since it is the most complete with the largest sample size. To lose this would cut the most meaningful variable of interest in the field of economics.

**Dowell Myers, University of Southern California, Other Academic (Faculty)**

Comparability across time is critical for analysis of trends in meaningful ways. Any added limitations for 2000 data jeopardize all our past understandings.

Our knowledge of occupational mobility could disappear if we cannot identify specific occupational niches. Without exact ages and races we cannot define groups in ways that are consistent with prior studies.

**Confidential Response**

My area of research is in urban planning, hence it is important for me to access data at a high level of geographic detail. Given that neighborhood oriented programs are now becoming popular, understanding the structure of neighborhoods and localities is essential to our field. Also, the detailed structure of neighborhood populations and comparability across space and time is critically important.

**Christina Paxson, Princeton University, Economics (Faculty)**

I am currently doing research on socioeconomic status, health, and mortality. The PUMS provides extremely useful information on socioeconomic status that can be matched (by birth cohort, race, and geographical location) to health and mortality information. Removing or reducing geographical identifiers renders the data nearly useless for this purpose. It is also crucial to know the exact year of birth (i.e. age) of each respondent.

Reducing the amount of detail in income would also make it more difficult to construct accurate measures of the distribution of incomes within groups of individuals, to assess the relationships between income distribution and health.

**Jeff Hayes, University of Wisconsin, Sociology (Graduate Student)**

Aging researchers need to be able to distinguish different age groups over 65.

Modeling incomes would be much more difficult -- both to estimate and to explain findings to various audiences -- if incomes were grouped beyond the use of topcodes.

Detailed occupations are very important to much of the work in my field. I suppose we would all live if the 1950 classification scheme (or a similar one) were used, but I like the IPUMS strategy of having it all! That is, I like having both the 1950 based for time series, but I also think that sometimes one needs an occupational structure that reflects the period in which the data are collected.

**Jesse Rothstein, University of California, Berkeley, Economics (Graduate Student)**

I have recently been working on a project using 1960-1990 PUMS data to estimate differences in wages across industries. The central challenge in the empirical implementation of this research was the construction of comparable industry and occupation measures from the four census data years. In the end, the only thing that allowed the research to go forward was the fact that the 1990 PUMS had the most detailed industry and occupation codes of the four data sets; if it had been less detailed, I wouldn't have been able to do the research. As it was, the limited detail in earlier years severely limited the usefulness of estimates for those years.

**Steve Doig, Arizona State University, Media (Faculty)**

Journalists plan to use PUMS to do stories on the widest possible array of topics. Any limiting of detail will invariably degrade the accuracy and quality of the data and the resulting stories.

**Charles Haynes, The University of Texas at Austin, Demography (Graduate Student)**

My research delves into the various factors which influence migration decisions and the probability of becoming an interstate migrant. It is therefore essential to my work that I have detailed data on race/ethnicity and ancestry, state, age, and income. Reducing data availability in those areas would be highly detrimental to my work. Occupation and industry categories, for me, are much less important. I typically only use major categories.

**Eric Monkkonen, UCLA, History (Faculty)**

I worked on age standardized homicide and use the IPUMS censuses to establish age distributions for relatively small geographic units. I regroup the age data into different categories, depending on the categories of the various comparison populations. Although age categories are relatively stable recently, it is clear that they will change as different standards come into use. As a minimum, age needs to be recorded by year.

**Rachel Deyette, Harvard University, Other Academic (Graduate Student)**

I have used the PUMS to examine participation in voucher programs. These programs generally have quite specific criteria -- age, income, geography, public school enrollment, etc. For research purposes, it is CRITICAL that the appropriate characteristics be available to IDENTIFY the eligible groups, and it is extremely important that as many detailed characteristics about these groups -- occupation, working status, citizenship/residency status, age, receipt of public assistance/social benefits, race, etc. -- be available so that we can see who is taking advantage of these programs. Especially with a political hot potato like vouchers, we really need to know who is benefiting and who is being left behind in education.

**Kristen Peterson, Brown University, Sociology (Graduate Student)**

My current research involves Japanese American occupational concentration. If the race or ancestry categories were limited, I would not be able to accurately identify my study population. Having an aggregate category, such as "Asians", would be unacceptable due to the diversity among Asian American groups. Also, Japanese Americans have very high rates of intermarriage, and having both the race and ancestry variables have been essential to capturing this population.

Because I'm researching occupational networks, it is important to have the most specific occupational and industrial classifications. For example, knowing that Japanese Americans are concentrated in a particular industry without knowing what type of occupation (or vice-versa), I would be unable to identify anything meaningful about occupational networks. The specificity of BOTH of these variables is necessary for my research. In addition, providing only broad geographical categories would also inhibit the study of networks or communities.

**Karen Smith, The Urban Institute, Other Nonacademic**

I do retirement research. Loss of age above 65 would render the PUMs useless for retirement analysis.

**Dan Devroye, Harvard University, Economics (Graduate Student)**

With regard to question 7: There is simply no way to choose among these different aspects of occupational status. Each of the categories -- SES, industry, and type of work -- is vital for research in labor economics and related fields. My own research relies on data that are both historically comparable and comparable across countries, and I shudder to think that either time series or cross section research should be compromised. What we need is more detail, not less!

**Confidential Response**

My research deals with educational markets and their variation across metropolitan areas. I sometimes need detail at the PUMA level, and I would be greatly affected if I couldn't get that level of detail. At least, please keep the metro area identifiers.

Elimination of race and income would be a great source of trouble, too. Educational quality differs greatly across income levels and ethnic groups. If these variables were eliminated, I would be unable to capture this very important source of variation.

**Matthew Sobek, University of California, Riverside, History (Researcher)**

Occupations are the single most important social identifiers in the census, speaking to a whole range of behaviors and life chances. Moreover, occupations are available in detailed form in every PUMS, allowing long-term measures of economic and social status before such questions as income are available. There is no substitute for historical analyses of social structure, economic attainment, etc. Even in the modern period, occupation has much information not available from other questions such as income. Detailed occupations allow the application of socioeconomic scores, which have proved to be powerful analytical tools in sociology, history, and economics.

Grouping age data would cripple cohort analyses tracking particular groups as they age across the censuses. In defining generations, one cannot be constrained by pre-determined age groupings. Studies of retirement and aging more generally would be compromised by less detail -- and this when an aging population poses one of the greatest public policy concerns of the new century.

**Confidential Response**

I study the segregation of men and women across occupations and industries. Without information on detailed occupational categories, it would literally be impossible to study this important topic. In fact, it is well known in my field that the "detailed" occupational categories are too gross to obtain more than a guestimate of the actual degree of occupational or industrial segregation.

Note, too, that for much of the research in this area, one analyses will only use three, perhaps four variables out of the following list: detailed occupation, sex, race, education, maybe age and industry. Assuming that one only has national-level or even state-level data, it is virtually impossible to identify a particular person based on these variables alone. Yet files like the PUMS are indispensable, because cross-tabulations of these variables are not available in published tables.

**Richard Campbell, University of Illinois at Chicago, Sociology (Faculty)**

I do specialize in aging and the loss of age detail would be extremely unfortunate at a time when the number of the "old old" is expanding dramatically. Regarding the occupation issue, again if you want to study occupational data on the elderly, you need detail. It is hard for me to answer question 7 because it depends in part on what is provided for income. Assuming decent income data were available, I would want the occ. data to maximize historical comparability.

**April Eaton, University of Washington, Sociology (Graduate Student)**

If the data on occupations were to become less detailed, I would not be able to continue a project on the labor market impact of immigration. Further work depends on the ability to compare the labor market structure of MSAs in 2000 with that on 1980 and 1990.

**Peter Meyer, Northwestern University, Economics (Graduate Student)**

I have been just today downloading data on particular occupations in the 1880 census to see if technological change at different time periods has some similar effects on income inequality. I will be downloading similar data for 1970-1990 soon. I need to be able to distinguish occupational categories in the recent time period to do this -- the difference between electrical engineers and computer scientists or mechanical engineers is probably minor for most research but it may really matter to studies of technological change because we \*know\* that electrical engineers are really ahead of the others in receiving effects from the invention of the microprocessor.

I study technological change partly because I think it is so important, and will continue to have major effects on development worldwide indefinitely. So I hope there will not be a reduction in occupational precision; we may need that sort of evidence to understand the spread of technologies.

**Kim Rueben, Public Policy Institute of California, Economics (Nonacademic)**

I do work on the effects of policy on state and local government. I am working on a project examining teacher shortages across different states and districts. I was looking forward to using the pums 2000 data to examine changes in the number and qualifications of public school teachers as compared to other people employed in similar areas. If information on occupation was reduced it would be impossible to examine these questions.

I am also beginning work on a project using 1990 census information to investigate how linking public school attendance to housing decisions affects peoples decisions. Studying similar issues across states using the 2000 census will be more informative given in the last decade more states and districts have adopted both open enrollment systems and licensed charter

schools. Investigating these questions and how different ethnic communities are served by having more school choice will be incredibly important and impossible to study if information on individuals' ethnic and racial status is not released.

#### **Confidential Response**

Our research on aging requires full age identification. The differences between younger seniors (65 to 70) and older seniors (85 or older) are profound.

#### **Confidential Response**

I used the PUMS and PUMA data to look travel patterns and disability for elderly. A category of over 65 is not detailed enough since the disability rates increase as one ages over 65. As the elderly pop. increases, this data will be more important.

#### **Katherine Condon, Florida International University, Other Academic (Faculty)**

Issues of aging need more detail than just a "65+" or "60+" age category. I have given presentation where I have been told "65 isn't old" Yet, then one turns around and senior citizen discounts start at age 60. It is increasingly clear to me that being able to have 85+ age category, and eventually will need 100+ age category will be absolutely imperative given the increasing longevity of people. In Miami-Dade County the percentage of 85+ elders of the 60+ population has increased from 4.6% in 1970 to 9.0% in 1990. It is the 85+ that have the highest probabilities of frailty. Being able to plan for services in the community for such "needy" populations is important.

While the majority of elders in the past have been white and non-Hispanic, Miami-Dade County was already showing its unique demographic flavor in 1990 with 40% of elders being Cuban, and another 8% being non-Cuban Hispanic. Approximately 9% of elders in Dade County were black non-Hispanic. Unlike it's neighboring county of Broward in which over 90% of the 60+ population was white non-Hispanic. Not being able to get this kind of racial/ethnic break-out would be absolutely devastating to doing policy research for the county. Further, it has been projected nationally that the number of minority elders will increase astronomically in the next 40 years. While many urban areas may not have large heterogeneous elder population already (unlike Miami) being able to have the racial/ethnic along with age breakouts as detailed as possible has been absolutely important in doing policy research and examining need in the Dade County.

With respect to income, I realize that this is a very touchy issue for many individuals to answer, particularly among elders. This is also key to being able to help make decisions on policy issues as well as understand health care service needs. The difference in percentage of elders who have physical limitations for those who live above poverty and those who live below poverty was astounding even for me as a health researcher.

Historical consistency is very important - I just finished a paper using IPUMS data from 1970, 1980, and 1990 to examine historical demographic trends of elders in Miami-Dade County. I was not able to use 1960 data because there was no county identifier. It would have been very useful to have 1960 data because this census occurred before the Cuban influx into Miami-Dade County. Being able to have at the very least a county identifier is important for policy planning since in Florida if any planning is done it is done at the county level ... not state-wide.

#### **Lisel Blash, Public Research Institute/San Francisco State University, Other Academic (Researcher)**

PUMS is absolutely vital to my work. It is so difficult to find such detailed information at this geographic level (PUMA's). I have used these data to look at the economic impact of a local living wage ordinance; a historical survey of the changing race, ethnicity and pay scale of hotel and restaurant workers in San Francisco and Los Angeles; at the changing demographics in the SF Bay Area; at the need for special services for people with limited English speaking skills (by county), at the need to establish adult care facilities with specific cultural and language needs in one PUMA in San Francisco, etc. I have looked at the demographics of people who commute into San Francisco compared to the demographics of those who both live and work here; I have been able to ask questions like "Which racial and ethnic groups in the city are most likely to live in overcrowded households?" I have been able to look at the specific demographics of people who work in below-"living wage" jobs by industry and by occupation. If the geographic detail were eliminated, I would not be able to do any of this work. If the occupational information were removed or made more broad, the value of these data for policy analysis on jobs and wages would be greatly diminished. Less detailed information on race and ethnicity would certainly impact my ability to do relevant research on the San Francisco Bay Area, since we have such a diverse population.

#### **Frank Braconi, Citizens Housing & Planning Council & New York University, Economics (Faculty)**

I am primarily involved in urban labor research and community development policy. It is critical for me to be able to relate changes in incomes, work status, and occupations to specific sub-metropolitan geographic areas. If the detail on those issues is diminished, it would impair our ability to track changes in urban economies and their effects on specific communities, and also to evaluate government policies aimed at promoting community development.

**Ileana Ciobanu, Harvard University, Economics (Undergraduate)**

In general, more information is better than less information in economic research. Economists need to control for as many factors as possible in order to better argue that the independent variable of interest causes, and is not merely correlated with, the dependent variable of interest. Economists can better control for as many factors as possible by having more detailed rather than less detailed data. Thus, in my opinion, reducing the detail of available data would greatly undermine economic research.

**Celina Schocken, UC Berkeley- Graduate School of Public Policy, Other Academic (Graduate Student)**

As a graduate student in public policy, I research in several areas, so I cannot say exactly how it would affect my research. However, having less detail means being less able to measure some of the changes in public policy.

**Mark Smith, The MEDSTAT Group, Economics (Nonacademic)**

My research using PUMS concerns the effects of welfare programs on household composition. Age and race/ethnicity distinctions are key elements because the effect of welfare differs substantially among women of different ages and racial/ethnic categories. At a time when the United States is aging and becoming more diverse ethnically, reducing the level of detail in any of these areas would be quite harmful to future social-economic research.

**Darius Lakdawalla, RAND, Economics (Researcher)**

My research involves relative wage comparisons between different occupations. Eliminating the detailed occupation coding makes it impossible to extend my work to apply to 2000 Census data.

**Joseph Doherty, UCLA School of Law, Other Academic (Researcher)**

We recently examined the cost of living for garment workers in Central Los Angeles. It would have been impossible without PUMS, which allowed us to estimate housing and utility costs and household composition among garment workers. We also used CPS to fill in the gaps, but its geographic limitations make it only marginally useful. No other data source is comparable to PUMS.

**John Johnson, University of Illinois, Economics (Faculty)**

The use of detailed data is really one of the strengths of the PUMS sample. The number of important and vital questions that academic researchers can answer would be greatly reduced by cutting back on regional or localized data. Any number of evaluation projects- including issues on the impact of localized labor markets, welfare reform, minimum wage- would be severely limited by cutting back on the detailed data available. I am strongly against it.

**FlorenceMae Waldron, University of Minnesota, History (Graduate Student)**

The staff of the Minnesota Population Center must already expend a great deal of time and energy to make historical data compatible across years. The proposed reductions in detail would, I believe, affect compatibility to earlier years to such a great extent in my work that the 2000 data would become totally useless. In order to study a population as precisely focused as persons of French Canadian ancestry currently residing in New England, many of whom are quite elderly, it is essential that I and other researchers have access to highly detailed data on age, geographical region, and racial and ethnic background. Without this level of detail, the 2000 PUMS is unusable for our research; the same would be true of researchers examining contemporary ethnicity, race, and immigration issues within numerous populations throughout the United States.

**Confidential Response**

I conduct research on the geriatric population in Manhattan for the purpose of tracking trends, identifying needs, and developing programs. As most people know, people over 64 are not one group. Not only are they very different based on their living circumstances, past occupational history and income, but they are also very different within different subset age groups. Combining age groups or eliminating/streamlining some of the other variables that have been mentioned would make my type of research impossible.

**Dan Black, Syracuse University, Economics (Faculty)**

I try to ascertain the earnings of various racial and ethnic groups. Restrictions on income questions might render the PUMS useless

**Confidential Response**

In education and immigration research, research is increasingly observing trends at the level of PUMAS as a way of controlling for unobserved factors that might affect outcomes such as earnings, education, mobility etc. Aggregating beyond the level of the PUMA would set back social science research, and policy analysis, by decades. Further, aggregating age groups and racial/ethnic groups will make it extremely difficult to examine social and economic trends. What are they thinking?!!

**Scott Drewianka, University of Chicago, Economics (Post-Doc)**

Reduction in detail could render the PUMS essentially useless for me. Detailed information on age, education, income, labor supply, and occupation is essential for virtually any good empirical labor economics study. Much of my own work also involves factors that vary across locations, so it is extremely important to be able to identify the geographic location of observations as finely as possible.

**Rebecca Sandefur, University of Chicago, Sociology (Graduate Student)**

Most of the research in stratification, inequality, and work and occupations is moving toward recognizing the need for GREATER detail about occupations, rather than less detail. At the very least, a less fine-grained coding scheme would need to allow the researcher to identify not only type of work, but also detailed industry.

As the population of this country ages, more and more people end up in the 65+ category, and these people become less "selected" on other traits. Research -- and policy!!! -- cannot afford to lose the information we would lose by treating this group as homogenous with respect to age.

**Confidential Response**

The reduction of detail would make it far more difficult to look at how important factors such as income vary across ages and occupations as well as geographical areas. Census data is a valuable indicator in educational research because it allows broad comparisons across a population with highly specified characteristics. Reducing the level of detail is always harmful in this sense. For example, reducing age data may make it impossible to split groups correctly as to whether they were of school age when certain legislative acts were passed.

**Confidential Response**

The PUMS is the ONLY source with enough cases to analyze some of the fastest growing immigrant groups in the United States! If we are to be able to study the long-term impact of immigration in the United States and the adaptation of immigrants from various groups to the United States, we must have detailed information on ancestry, race and country of origin! Some point to the Current Population Survey as an alternate and cheaper source for such detail but while the CPS samples may be sufficient for some larger groups or panethnic groups (Asians as a whole, for example) they are NOT large enough for analyzing individual immigrant groups.

Research on immigration and race/ethnicity that comes from PUMS data has applications for academic research but also important policy implications as well! (Age is important too since the impact of migration varies across the life course but I will let those in aging research speak to that specifically).

**David Clark, Marquette University, Economics (Researcher)**

I use the PUMS to model individual migration propensities, and well as factors which determine wage levels. Occupation and locational attributes are definitely needed for that research. Reducing the geographic resolution would severely diminish the usefulness of this data set. There are alternatives, but none have the scope of geographic coverage combined with the detail of occupational and industry status that the PUMS provides. PUMS is seen as a very rich data set among urban and regional economists, and I strongly urge you to maintain consistency between the Census years.

**Patricia Kelly Hall, University of Minnesota, History (Graduate Student)**

I analyze long-term trends in immigration and internal migration. The variables you describe as under threat of detail loss are some of the most important personal characteristics associated with migration differentials. The loss of the mother's/father's birthplace question was already a severe blow to generational analysis of migration trends. Further loss of personal characteristics would be -- as the survey so aptly states -- "catastrophic." I understood the purpose of the PUMS was to provide social scientists with access to individual-level data. It sounds like the entire concept is at risk. At least for my work, that may well be the case.

**Carole Shammas, University of Southern California, History (Faculty)**

Income and wealth are so poorly sampled already by govt agencies that further grouping would make them all but useless. As the growth of the old, old has been the big news over the past 25 years, it makes little sense from a policy standpoint to start obscuring that change by grouping older ages.

**Joy Williams, Environmental Health Coalition, Other Nonacademic (Nonacademic)**

We are an environmental justice organization. We are interested in the toxic chemical exposures of people both at work and in their communities. We can make broad generalizations about chemical exposure from occupation, but, the more process-specific this information is, the more accurately we can gauge what the chemical exposure is. We are also interested in

tracking the racial and economic status of the neighborhoods with high levels of pollution, and in identification of neighborhoods with high concentrations of children under 6, the group at highest risk for lead poisoning.

#### **Confidential Response**

Now I am studying the effects of immigration on the U.S. economy. To do this, I need ancestry, birth place, age, year of immigration. Moreover, to investigate the economic performance of immigrants in the U.S., I need the income level and the type of occupation of them. It is obvious that the detailed data help my research.

#### **David Burress, University of Kansas, Economics (Researcher)**

We build impact models for regional effects of development within Kansas. We need the PUMAs to be smaller, not larger, so we can localize effects better. We also look at effects of development on human capital by occupation and industry within Kansas. If the occupation categories are too coarse, then we can't distinguish between types of industry, so we lose the ability to recommend industrial targeting.

#### **Susan Carter, University of California, Riverside, Economics (Faculty)**

Some of my research is focused on gender, race, and ethnicity differences in occupational attainment. All the standard measures of difference are quite sensitive to the level of detail by which occupations are reported. A reduction of PUMS occupational detail would severely restrict scholarly work in this important area.

#### **Kate Antonovics, University of Wisconsin, Economics (Graduate Student)**

My research focuses on racial discrimination. Detailed information on race, education, age and income are all absolutely essential for the work that I do. Broad groupings make rigorous empirical work extremely difficult. For example, it is essential for my work to have data on hourly wages. Because the "number of hours usually worked per week last year" variable is only available in later Census years, I am forced to use weekly wages. Given the tremendous variation in the number of hours per week that people tend to work, this measure of wages is extremely noisy.

#### **Lawrence Nitz, University of Hawaii at Manoa, Political Science, Other Academic (Faculty)**

Hawaii has only a handful of PUMAs. As it is, the 100,000 rule combines three islands which are nearly two hundred miles apart (Maui, Lanai, Molokai v. Kauai). We have a largely Asian-Pacific population in which ancestry is as important as race—since there are different Chinese cultures, varying sources of immigration from Asia, and radically different characteristics among folks of the same ethnicity and/or heritage as a function of their respective waves of immigration to Hawaii.

One of the most pressing problems in a small state with a tradition of friendliness is demonstrating systemic failure. When our friends in the business world complain about unqualified job applicants streaming from the high schools, we have no external way of validating either the skill level of the job or that of the candidates. Occupational data should be reflected in a way that maps to some small extent onto the average skill and disposition categories developed in the new Dictionary of Occupational Titles.

Until we can document the relationship between the general skill levels associated with jobs and the wages and benefits paid, we cannot respond adequately to questions of equitable distribution of educational facilities or expectations.

#### **Joelle Saad-Lessler, Columbia University, Economics (Graduate Student)**

The more you reduce PUMS detail, the fewer observations we have to work with, and that makes all the results of the research less robust. When you group age data, we are precluded from being able to do significant amounts of research by age category. This will become more and more significant of a restriction in today's labor market, where people are not retiring early, and are working longer and living longer as well. For example, if the age category 65 and older replaces more age detail, we would be precluded from conducting any research about labor participation trends among those over 65, by their age. Any such research will have to content itself with grouping a 65 year old with an 80 year old, thereby reducing the usefulness of any findings. This will affect policy on the reform of the social security system, for example. It will also affect research on the connection between working and longevity, or working and health outcomes. Finally, with the changing skill requirements of the labor market, we would be precluded from learning more detail about the rate of skill depreciation, and the value of experience.

#### **Leora Friedberg, University of California, San Diego, Economics (Faculty)**

I used the IPUMS to analyze the impact of Old Age Assistance on retirement. This was a very informative study because OAA benefits varied across states. Social Security benefits do not vary across states, so it is difficult to understand how Social Security benefits, which depend on income and other personal characteristics, affect retirement ages. The study of OAA showed definitively that benefits affect when people retire. This helps us understand the extent to which Social Security has lowered retirement ages, which has in turn increased the drain on the Social Security system.

**Patricia Oslund, University of Kansas, Economics (Researcher)**

1. Research on migration. We have on-going models that include parameters for sensitivity of migration to overall household income and to wage income. Reduction in geographic scope would make this analysis less meaningful, and reduction in income detail might make it impossible.

2. Social accounting model. We have an on-going social accounting matrix model that uses PUMS data (we aggregate to state level) to find: average full time wages by industry; breakdown of income by household income category by industry - for each industry, what share of labor income goes to households with total income < 40th percentile of total income, etc. We need income and industry detail for this. Geographic detail is less important here.

**Reeve Vanneman, University of Maryland, Sociology (Faculty)**

For the last several years I and my colleagues have used the PUMS, the detailed occupational (EEO) tables, and the STF3c, building up from county records in the STF3C and EEO and from PUMAs in the PUMS to analyze how individual and metropolitan area characteristics JOINTLY affect earnings, labor force participation, and occupation. We have completed a set of papers analyzing the impact of local labor market occupational segregation and the percent of female occupations in the local labor market. We have published this in the top sociology journals and continue to generate new research. All of this will become impossible with the kinds of restrictions you suggest. Just as we get the statistical ability to do multilevel analyses, the Census is thinking about not including sufficient geographic detail to make this possible. What a shame.

**Stephanie Aaronson, Columbia University, Economics (Graduate Student)**

There are several issues. First of all, I work primarily on young people. Since the living and working circumstances of young people change very quickly, it is important to be able to identify them by specific age. Even just grouping 16 to 18 year olds together would obscure a great deal of variation.

With respect to occupational data it is very difficult to detect changes in the distribution of occupations if the occupations are highly aggregated. A similar issue arises with geographic data. Furthermore, much analysis relies on cross-geographic variation in the data for identification. The fewer geographic areas that can be identified the worse the statistical analysis.

**Confidential Response**

I am a labour economist with a particular interest in geographic labour markets. Comparability of geographic areas is vital, as is the ability to control for differences in workforce and employment composition.

**Confidential Response**

If the data surrounding age, especially for children and race/ethnicity is reduced, our agency would have a difficult process in providing critical demographic information for policy makers and legislative bodies. The loss of the data would greatly impact the ability to design and implement programs focused on overrepresentation or underserved population especially for children's issues.

**Gabriela Galescu, Cornell University, Sociology (Graduate Student)**

I am working on a structural model of occupational segregation by race and city. David Grusky, my adviser, did a similar research on cross-national occupational segregation by sex and the results showed that there is no such thing as a general pattern of segregation, that the underlying structure of that phenomenon is more complex than previously thought.

I have started on working on a similar research on occupational segregation by cross-cities. Now I'm only working on 13 big cities, but my research will be really significant only when covering most of the cities of the U.S. Also I need detailed race and occupation data in order to give meaningful results. I am claiming that segregation is more subtle than "black people don't get as good jobs as the white ones", but how more subtle and interesting the pattern is depends on the data I can use.

Besides, I grouped the 505 occupations in 42 categories to make the analysis easier for me for a final paper due one of these days and I was just amazed how much of the actual difference I lost even at the low-level of cross-tabulation. Really, for what I am doing, detailed data are absolutely fundamental.

**Huey-Chi Chang, University of Wisconsin-Madison, Sociology (Graduate Student)**

Age: In the study of the living arrangement of elderly immigrants, I won't be able to identify precisely at what age these immigrants came to US even with the information on the year of entry. Without detailed age information on the elderly, I won't be able to use age as an indicator to explore how the aging process or changes in the health condition might influence living arrangement. Age is an important variable especially if we would like to disentangle the cohort, period and age effects. Unless we can develop a grouping scheme that is consistently used throughout all such time variables (or across all censuses for synthesis cohorts), any arbitrary grouping may render the information useless.

Occupation: 505 titles are preferred. Detailed classification is more likely to insure the comparability with older censuses. It's easier to collapse categories and to reorganize the data with detailed classification.

#### **Confidential Response**

One project I am currently working on involves examining the differences in income and education of individuals residing in large, mid-sized, and small cities. This analysis requires a significant level of detail of the data for geographic variables as well as income, race, and age variables.

#### **Charles Wetherell, University of California, Riverside, History (Faculty)**

Single year of age is becoming increasingly important in the demography of the aging population as well as in social and economic analysis of the elderly behavior. Fine grained geographic analyses are also becoming more important as sociologists and social psychologists try to assess environmental and neighborhood effects. As a Co-PI of a five-year, NIH funded study of functioning in Mexican-American and Euro-American families in Southern California, my work would be seriously impaired by any reduction in information in the 2000 PUMS

#### **Gunnar Almgren, University of Washington, Demography (Faculty)**

In fact, reductions in the number age categories would greatly reduce the validity of mortality and morbidity rate estimates by age. This is a particular concern for researchers attempting to link social structure with health outcomes.

#### **Robert Adelman, University at Albany-SUNY, Sociology (Graduate Student)**

PUMS data are, without question, an integral part of my research. As a sociologist/demographer interested in racial and ethnic inequality in the United States, any reduction in PUMS detail will hurt my (and many others') ability to perform adequate analyses of important, policy-related issues.

#### **Lindsay Rossano, The Joseph L. Mailman School of Public Health at Columbia University, Sociology (Researcher)**

For public health research (specifically social epidemiology) such as I do, detailed information with regard to SES is vital. This, of course, includes information with regard to race, education, occupation, and where such people live. As it stands we know that census data likely underestimates low income and minority populations. Efforts to study these populations will only be further hindered by generalizing the data available.

#### **Confidential Response**

I need to be able to match PUMS data to specific local public policies in order to assess the policy impacts. More detail about geographic/political jurisdictions greatly eases this task.

#### **Confidential Response**

Loss of detail is loss of information. Loss of information means statistics are less likely to capture micro behavior.

#### **Carolyn Liebler, University of Wisconsin at Madison (sociology & demography), Sociology (Graduate Student)**

I do research on American Indians, including people who say their ancestry is partly American Indian. I am interested in the meaning behind reporting a tribe. Also, I use tribal affiliation to add context to my work.

Also, I do research on the meaning and measurement of race. If you were to combine categories in the race question, I would not be able to understand the differences, for example, between people who are American Indian/White/Black and those who are American Indian/White. The 2000 Census race question is very exciting to people who study the meaning of race because it gives us significantly more information than we have ever had about people's racial identification.

Whether or not you combine categories in the race question, the ancestry question is also key to understanding the complex racial nature of our country. In my research, I have found that people who say they are American Indian by both race and ancestry are different in interesting ways from those who say they are American Indian by race only. I hope to use the 2000 PUMS data to see if people who are American Indian/White by race and only American Indian by ancestry are different from those who are American Indian/White and only White by ancestry.

#### **Mark Ellis, University of Washington, Geography (Faculty)**

My research involves making comparisons of the social and economic situation of immigrants in major cities of immigrants over time. I spent considerable time (with NSF money) maximizing the compatibility of the 1980 and 1990 microdata geographies to build metropolitan areas that were similar in both years (see Ellis, Mark, Michael Reibel and Richard Wright (1999) "Comparative Metropolitan Area Analysis: Matching the 1980 and 1990 Census Public Use Microdata Samples for Metropolitan Areas". *Urban Geography* 20: 75-92). This is essential to remove any effect of geographic boundary changes on measurement of migration, employment and so on. I have used these synthesized metro areas (and others have too) to publish a number of papers on immigration/migration linkages, immigrant wages, and immigrant employment. I will need to

build synthetic metro areas in the 2000 PUMS that match those I built for 1980 and 1990. This will make results from the 2000 PUMS compatible with my earlier work and allow me to infer trends about immigrants in specific metro areas with greater confidence. Reduced information on race and ancestry and income may also have devastating effect on immigrant analysis if one cannot identify the specific economic circumstances of groups or if these data are not compatible with earlier years.

#### **Confidential Response**

My thesis research focuses on the effect of family migration on wage growth, and specifically how this effects married women differently from other workers. The proposed changes would harm my ability to use the census in the following ways: 1. Geographic areas differ in terms of number of jobs available in various fields. Not having labor market information (PUMAs) would limit the ability to make use of this information. 2. Labor markets are thinner in certain occupations. Having specific occupational information helps identify fields most helped/hurt by family migration.

#### **John Sandberg, The University of Michigan, Demography (Graduate Student)**

I work primarily with age, race and education data to make population comparisons across time, and geographic region. Aggregated age-especially at older ages would be disastrous for analyses of the older and oldest-old (one of the fastest growing segments of the population). The dynamic relation of income and wealth accumulation among these elderly as they pass through these years is of critical importance, and would be hamstrung by grouping the data as proposed.

#### **Michael Katz, University of Pennsylvania, History (Faculty)**

My colleague Mark Stern and I have a contract from the Russell Sage Foundation to write a book putting the 2000 census in the context of social and economic trends in the 20th century. The PUMS is the basic source. Detail is crucial both to perform the level of analysis we need on the 2000 census and to compare trends across time. A reduction in the level of detail would cripple our ability to write meaningfully about major changes across time.

#### **Confidential Response**

Reduction in the detail of IPUMs data would be very detrimental to my research. My studies involve the effects of inequality on socioeconomic indicators by race and ethnic backgrounds. Since there is an increasing amount of interest in how different groups are affected by inequality, you can see that detail by race/ethnic background/ancestry are important to my work.

Furthermore, I use income to calculate measures of inequality. Yet income is not the only measure of inequality. Detail on education and occupation are also important to provide robust measures of inequality. In addition, age detail is indispensable because in comparing different groups based on race, ethnicity, etc., one must age standardize. Clumping 65+ excludes analysis of the elderly population. During a time in which there will be more 65+ people, limiting analysis to the younger age cohorts will disallow the many studies that could provide useful policy for Social Security. Also, geographic detail is important because this allows me to combine IPUMS data with other data sets at the aggregate level when I need to utilize indicators such as the infant mortality, death rates, etc. Aggregate measures at geographically differing levels are essential to providing sound testing and robust results.

Concerning the timing release of the data, if less detailed data is available earlier, followed by more detailed data later, that would be very good. This allows for preliminary analysis followed by more substantial investigation.

Finally, I am sure legitimate researchers will not divulge results such as to reveal the personal information of individuals. If the fear is that researchers do this unknowingly, perhaps the Department of Human Subjects at each institution should expend more efforts in disseminating rules concerning privacy.

#### **David Belkin, City of New York Independent Budget Office, Government (Nonacademic)**

Given the paucity of alternative sources of earnings and income data at the city level, the construction of such data out of PUMS detail is absolutely vital for urban economic research. Not only should there be NO reduction in the level of geographic detail, there would on the contrary be great value in having GREATER geographic detail.

#### **William Donlin, University Of Minnesota, History (Undergraduate)**

Grouping ages for the elderly would definitely work against my research. I am currently working on a project that includes World War II Veterans in the second half of the 20th century. This group would be adversely affected with these new modifications and my project would loose continuity for the latest census year.

#### **Confidential Response**

PUMS data with detailed geographic codes are essential to policy researchers who are trying to understand the baseline characteristics of the populations subject to new policies. Limiting the availability of these data would make it necessary to

collect more data for individual policy evaluations-- an expensive and inefficient approach given that much of the necessary is potentially available from PUMS.

**Steve Holden, Johns Hopkins University, Economics (Graduate Student)**

Most important is the regional variables, race, age and gender.

**Confidential Response**

Much of our current research requires identification of lawyers or judges. Our research also requires clear identification by race and income

**Ernest Schreiber, Lancaster New Era, Media (Nonacademic)**

Any decrease in detail makes the data less useful. When the bureau goes to the expense and effort to compile good data, any program that degrades its usefulness makes utterly no sense.

**Robert Hauser, University of Wisconsin-Madison, Demography (Faculty)**

As far as I am concerned, elimination of the detail of age, race, ancestry, income, occupation, and geography would essentially eliminate the value of data from the long form. This is a shameful, cowardly, and ludicrous proposal. I hope it will disappear promptly and not be raised again.

**Cristian deRitis, Johns Hopkins University, Economics (Graduate Student)**

I consider historical evolutions of wage patterns in my research. Research in this area is significantly compromised by vague definitions of skill, education and occupation with the data currently available. Results are challenged regularly on the basis of historical consistency, relevance and significance. To broaden categories further would only serve to decrease the power labor market analyses may have.

**Seth Sanders, University of Maryland, Department of Economics, Economics (Faculty)**

It is imperative that all of the data be entered even if restrictions are necessary. First, there are many questions of data quality that can only be answered by cross referencing data items and checks for consistency and error are greatly diminished when the data is preclassified. The exact coding of age seems quite important as so many studies use the age structure of families in creative ways. For example, without exact ages identifying twins would be impossible, identifying when kids meet compulsory school age is impossible and tracing benefits on social programs that are age dependent is also impossible. The geographic detail would be a total disaster if it is obscured. It would just about eliminate the field of rural sociology and greatly limit policy studies in which geography has a role.

**Eugene Hammel, University of California Berkeley, Demography (Faculty)**

Reduction in specificity at higher ages will make it very difficult to track ongoing changes in survivorship at these ages, just as that population is increasing and just as the results of analysis become more important.

**Confidential Response**

As a geographer it is particularly spatial distribution on the neighborhood level I am interested in. PUMS detail is thus EXTREMELY important

**Confidential Response**

For my own research, the 2000 census is not all that important as I mainly look at the 1940 census, but I could see that any reduction would harm the general trend of scholarly research.

**Amanda Brown, UC Berkeley, Other Academic (Graduate Student)**

I am studying occupational effects of ethnicity and immigrant status, so it is very important for me to be able to see the ethnic break down by as many occupational categories as possible.

**Confidential Response**

Ability to track racial trends in lieu of new multi-race question. Ability to track social/demographic/ local research for business, industry and government.

**Anupam Tyagi, University of Utah, Economics (Graduate Student)**

An area of my research interest is social and economic inequalities. This requires detailed data on all aspects listed above that characterize a social and economics group. Loss of details means loss of criteria on which such groups can be differentiated to begin such research. Loss of geographic details will be a big loss since my future research will heavily depend on geographic variations in the sample. It has been delightful to use the well organized samples by IPUMS. With the

long form in 2000 Census I was hoping to have more detailed information in the samples available for research, NOT LESS!!!

#### **Confidential Response**

Since I do geographic research, an increase in PUMA size would have a great impact on geographic detail in less densely settled areas. The effect would also be a headache in regional studies.

#### **Confidential Response**

Socioeconomic studies of behavior in group contexts requires information about relatively small differences between individuals: just as "non-white" doesn't capture much information about culture, "over \$50,000" doesn't capture much about standard of living--and neither helps one understand the socio-economic environment faced by the individual or group in American society. Ask your Lexus-driving friends what it's like to own a new Ferrari and you'll see what I mean...

#### **Gunnar Thorvaldsen, University of Tromsø, History (Faculty)**

Since the most reliable detail on ancestry comes from old people, grouping them agewise reduces the value of the data set for immigration and ethnicity research substantially.

#### **Confidential Response**

I would only use PUMS data in comparative research with Britain, therefore the more detail available, the easier comparisons would be made.

#### **Barbara Reskin, Harvard, sociology, Sociology (Faculty)**

Among my research areas is the study of trends in occupational sex and race/ethnic segregation, using 3-digit census codes. There would be no way to monitor these trends or explain cross-sectional variation in trends with more aggregated data. It is well known that occupational segregation (like God) is in the details, and when one sacrifices detailed data, one obtains the illusion of considerably less segregation. It would be impossible for me and others who pursue related research to continue to monitor segregation trends. This is critical because segregation is both the cause of other forms of gender/race/ethnic inequality and it is an important indicator of the extent of inequality.

#### **Robert Kaufman, Ohio State University, Sociology (Faculty)**

My own current and long-time research using PUMS data analyzes employment segregation and earnings gaps among race-sex groups. In my current research and analyses (funded by a multi-year NSF grant under the Human Capital Initiative and now by a grant from the W.E. Upjohn Institute for Employment Research), I constructed 1961 labor market positions (comparably defined in 1990 and 1980) starting from 6 digit occupation-industry combinations and aggregating combinations only as needed to have sufficient cases in a position to analyze positions as units. Reduction in the occupation or industry detail would be disastrous for this and other similar research in the field because employment segregation is very sensitive to the level of data available. More broadly the extensive and growing research literature on sex segregation has relied heavily on measurement of sex composition at the detailed occupation level. Reducing occupational detail would seriously hinder such research and analyses of trends would be rendered impossible. Similarly, segregation researchers have just recently begun to move beyond black-white-hispanic comparisons to study segregation using more detailed definitions of race and ethnicity. Again such research would be rendered impossible if race-ethnic detail were eliminated.

#### **Confidential Response**

I have analyzed occupational group trends over time, and the reduction of data on occupations would be difficult to continue doing this historical analysis.

I also plan to look at racial and ethnic differences in occupational mobility over time, for my next paper and possibly in future research. By not maintaining historical consistency in occupation, race and ethnicity, or income, my research would be jeopardized.

Finally, changes in age groups, particularly for the elderly 65+, would inhibit me from doing future research with the IPUMS on changing work trends among the elderly.

#### **John Kenny, University of Chicago, Other Academic (Graduate Student)**

My research looks at immigrant associations. Without accurate information about ancestry and occupation there is no way I can accurately estimate the size of an immigrant community or the level of occupational segregation among immigrants from particular countries. Detailed occupational classifications are especially important. Segregation can't be found in broad occupational categories.

**Louis Chauvel, Institut d'Etudes Politiques de Paris, Sociology (Faculty)**

For age period cohort analysis of standard international classification of occupation, I need a complete Occupation code. I need also exact age for the calculation of exact cohorts. Discrepancies in these information will disturb the measure and break long term series.

**Carl Schmertmann, Florida State University, Demography (Faculty)**

(1) Aggregating the elderly into larger age groups would completely eliminate our ability to study many important policy questions.

Some of my recent research with PUMS data concerns living arrangements of men and women 60+ years old, by single year of age. There are important changes in living arrangements at ages 75+ that would be undetectable in age-aggregated data. For an example of analysis that would be compromised, see C.P. Schmertmann, M. Boyd, W. Serow, and D. White. "Elder-Child Coresidence in the United States: Evidence from the 1990 Census", Research on Aging, Jan 2000.

As the US population ages in the early 21st century, it would be extremely irresponsible of the Census Bureau to eliminate age detail from one of the primary sources of policy information on elders.

(2) Migration research demands geographical detail and large samples. There is no source other than PUMS data for many migration-related issues. Aggregation to higher geographical levels would destroy our ability to study many important policy questions, particularly those related to intra-state population distribution.

**Gordon Dahl, University of Rochester, Economics (Faculty)**

Income measures are vital for the research economists carry out. Reducing detail on the key income variables would greatly reduce the value of this dataset for economists.

**Robert McCaa, U. of Minnesota, History (Faculty)**

The proposed changes seem ill-considered. The brouhaha this past Spring was not about microdata, it was about questions on the long form.

Before the census bureau embarks on a radical new policy that could seriously harm one of the largest, most useful, and most used social science databases in the world, there should be a study to assess confidentiality risks in the microdata.

The confidentiality threshold must be extremely high (say the odds of linking microdata to another database at 1 in a million), but there should be some study of the risks before hand.

**William Collins, Vanderbilt University, Economics (Faculty)**

Much of my research concerns changes over the course of the 20th century in racial income, education, and occupational gaps. This kind of research is crucial for understanding how policy has (and has not) affected differences in well-being across groups. The IPUMS data has been indispensable to this research because the empirical analysis demands detailed information on age, race (and ethnicity), nativity, education, occupation, industry, migration history, and location. Moreover, the analysis requires a strong degree of comparability over time. More generally, compromising the richness of the IPUMS data would substantially hamper economists efforts to do the research that guides (or should guide) policy evaluation and design. Having worked with the data extensively, I'm fairly certain that any gains associated with more stringent confidentiality restrictions will be far outweighed by losses in research quality and quantity.

**Steve Coble, Vermont Agency of Commerce and Community Development, Government (Nonacademic)**

In Vermont the population is very small, 600,000 for the state. We currently have only 5 pums areas for the entire state. Fewer geographical areas would reduce our ability to identify areas of need for the programs that the Agency of Commerce offers.

**Monte Aaker, Minnesota Housing Finance Agency, Government**

The reduction in detail would cripple our efforts to target housing resources. We are trying to provide housing resources to households traditionally underserved by the private market. We need this detailed data at the smallest geographic area possible. Any reduction would harm our efforts to better target our resources. We need the age detail to provide better location decisions for our housing investments. Housing lasts a long time and having detailed data is needed for long term planning. This is especially important with the aging of the baby boom.

**Confidential Response**

I have used the PUMS data for two purposes within the last month. First, to detail the number of school-aged children by race in the Philadelphia Metropolitan Area (city and suburbs) for the last 100 years. Losing detailed age or geographic information would have made this analysis impossible. The second way in which the data were used was to predict private

school attendance across PUMA areas of different race and socioeconomic compositions for people of different races, occupations and poverty/income levels. Losing any of this information--but particularly increasing the size of the PUMA areas--would have made the analyses far less useful. As it is, it may be that PUMA areas are too large in size (and don't match with other geography in useful ways) for this particular analysis to be useful enough to be considered for publication or policy-related issues.

**Susan De Vos, University of Wisconsin-Madison, Sociology (Researcher)**

A major activity at present is to examine the living arrangements of elderly Hispanics. Age is very important as the "old old" (80+) is often of special concern. Much hinges on being able to examine people by detailed age, even among this oldest group. Also, ancestry is very important as there are marked differences among Hispanic groups. Finally, source of income is very important as SSI and Social Security are differentially received. Whether income comes from investments or wages also is crucial, but if this detail were kept, categorizing income in increments of \$5,000 (current \$) would probably be okay. On the other hand, occupation is dealt with simply and my research probably would not suffer from a much less detailed coding. The same is true for geographic location.

**Confidential Response**

USDOT sponsors long-term research on travel forecasting models that rely heavily on Census data, including PUMS. Commercialization of these research products, under TRANSIMS, is commencing this year and data availability will strongly influence the success of this effort.

These models are used by States and local agencies to identify transportation needs and develop transport programs. In addition, new requirements on social equity analyses will have to be supported by demographic data that can be compared in a time series using a consistent geography.

**Irene Bloemraad, Harvard University, Sociology (Graduate Student)**

My own research is on immigration. I need to have the most detailed place of birth and ancestry codes as is possible to have. If census data are only released with the major immigrant/ancestry countries (as is done in Canada), it is almost impossible to track changes in immigrants' demographic and socio-economic characteristics over time since the key immigrant groups change. For example, Portuguese immigrants were numerous in New England in the 1960s and 1970s. If, however, no identifying information is provided on them in the 2000 census, it is impossible to see how they are doing, nor how their descendants are doing. Since the Portuguese on average have low levels of education, it is important from a public policy and academic standpoint to know whether they were nonetheless able to achieve economic success (what the data imply) or whether they have had problems. This would give us an indication of what might happen for other immigrant groups with low levels of schooling. Given these concerns, it is also very important to have the most detailed information possible on income, quite detailed information on occupations, and extremely detailed information on year of migration to the US, citizenship status and age. I am currently comparing immigrant populations in the US to those in Canada, an important exercise for understanding the institutional dynamics of immigrant integration. Canadian data are often aggregated in large groups (e.g. just 28 place of birth/ancestry codes are supplied) which makes my research very, very difficult. It has gotten to the point that I have formally requested access to the confidential Canadian statistics (those NOT released as PUMS) because otherwise my research is impossible. If the US Census Bureau were to reduce the detail of the PUMS data, many academic researchers would desire assurances that they could have access to the detailed data through other channels, such as going to the Census Bureau in the summer to run statistical models. Does the Bureau (and the federal government) have the money to set up a mini-statistical center in order that academics and other interested officials can use the more detailed data? If so, how fair is it to restrict access to such data to a few privileged individuals? The beauty of IPUMS is that many people, in the US or elsewhere, professors or graduate students, can use the data. Greater use increases the chances for significant academic findings that could be important for public policy.

**Caroline Hoxby, Harvard University, Economics (Faculty)**

For analyzing the rate of return to education, it is absolutely necessary to know a person's exact age, so that we can associate him with the regime (laws, funding) under which he attended school. We also need the exact age for estimating consistent wage equations. The geographic detail that was provided in 1980 and 1990 was too aggregated for many of the projects related to urban poverty, crime, and education. If it becomes more aggregated, it will be a disaster.

**Allen Glicksman, Polisher Research Institute of the Philadelphia Geriatric Center, Sociology (Researcher)**

I try to understand the way immigrants assimilate into the American mainstream, especially in terms of employment. This has important policy as well as academic implications. Without ancestry, we cannot see how members of different ethnic origins are doing. Without specific employment data comparable with previous census data, we cannot look at assimilation over time.

**Myron Gutmann, University of Texas at Austin, History (Faculty)**

I work on relationships between population and environmental conditions, which requires as much geographic detail as possible (because areas are always so different). This is especially so in rural areas, where the population is sparse and PUMAs are spatially large. If PUMAs were much larger than they were in 1990, it will be impossible to do research on this important area of the relationship between agricultural land use and population, and therefore it will be impossible to make a policy contribution in this area that is vital to Americans who live in cities and on farms.

**Confidential Response**

(The first sentence was removed for the purpose of maintaining the author's confidentiality.)

Most of our patrons ask for detailed age groupings of the variables of interest. The variables of income and occupation are used most by HR students and income and rent and other expense variables by students of Marketing or finance. Sociology researchers seem to use a wide cross section of variables depending on their area however age and income seem to appear frequently there as well. Education levels are also used as grouping variables in many analyses.

**Confidential Response**

I use the income variables to make sure that other survey data are reporting reasonable amounts. Thus a reduction in precision for the PUMS data with respect to income could greatly reduce the ability to determine precision in other data sets. Additionally, I use geographic data in order to analyze specific proposals affecting only certain areas of the country. Again, a reduction in geographic precision would reduce the ability to effectively analyze policies.

**Confidential Response**

I have been studying differences in the gender wage gap across metropolitan areas over time. Clearly, grouped income information would completely preclude this kind of analysis, and grouped geographic information would reduce its usefulness.

**Ruth Turley, Harvard University, Sociology (Graduate Student)**

I'm specifically interested in city-level, race-specific variables (i.e., income by race, education by race, etc.). I cannot proceed without race-specific data.

**David Cutler, Harvard, Economics (Faculty)**

Perhaps there could be a process where researchers could apply for more detailed data and use it with special permission, if they have a secure computer. I know that HRS/AHEAD use this method. The Census Bureau also has Census locations around the country that people can use. That might be another solution.

**Confidential Response**

My research includes earnings differences by sex and race. I also examine the effect of working conditions and job injury risk on earnings. I could not estimate reliable earnings equations without adequate data on income, occupation, race, and age. Some specific examples: I used PUMS detailed occupational and industry data to calculate gender-specific injury incidence rates by 3-digit occ and ind. The BLS injury data is provided at the 3-digit level and I needed the corresponding number of workers in the 3-digit category. I anticipate doing research on discrimination in hiring and the PUMS racial and occupational data provided information on the availability of potential workers. I have a Ph.D. student who is looking at compensating differentials for commuting and is using the detailed geographical information.

**Confidential Response**

We are interested in behavior of pre-retirement and post-retirement individuals that directly affect their economic outcome. Grouping data for age or income will destroy what microdata are good for in this instance. Race and ancestry information are also crucial in determining how policy can affect different groups, and it is also crucial for us to retain this level of detail

**Confidential Response**

I frequently need to know information on target audiences based on age, income, occupation, and occasionally race. If any of these statistics are affected, it could have impact on my decisions.

**Mark Hayward, Population Research Institute, Penn State University, Demography (Faculty)**

Social scientists are making great strides in incorporating geo-referenced concepts into their research. Spatial-temporal perspectives, though relatively new, hold considerable promise for new insights into endemic social problems. Increasingly sophisticated applications of GIS technology and spatial statistics are the new analytic tools of social scientists working in this area. Without high quality, geographically-refined data, the application of these tools will be seriously hampered. Potential scientific advances will be eliminated.

**Confidential Response**

I specialize in analysis of racial and ethnic fragmentation. A reduction of the available categories would greatly harm the meaning and informativeness of the heterogeneity indexes I construct.

**Olivia Mitchell, University of Pennsylvania, Economics (Faculty)**

To do any sort of retirement research or analysis on the wellbeing of the older population it is indispensable to have detail on age. To do policy research on the most important of all national programs -- Social Security, Medicare and Medicaid, long-term care, tax policy -- one must have access to geographic detail, income and wealth, race, and age. I cannot emphasize strongly enough how critical it is to NOT group age data.

**Confidential Response**

My opinion is that confidentiality can be preserved by large groups at the local level and it is useless to reduce detail in other variables like age, occupation ... which will affect the precision or make impossible thousands of studies. If small geographic units are of interest to many, it is necessary to use large samples to have significant results, so that PUMS are not relevant. Trying to accommodate everyone might destroy everything.

**Confidential Response**

I do not actually do research, I am a statistical consultant and help the faculty and students do their research. My answers here are based on the research I have helped my clients with. We have many faculty and students (graduate and undergraduate) who use the PUMS data for many different purposes. What would be catastrophic to one may be of no consequence to another. Without a doubt, though, income and geographic location would seriously affect everyone.

**Stewart Tolnay, University at Albany-SUNY, Sociology (Faculty)**

Reduction in race or socioeconomic detail would seriously affect my research, especially if comparability with previous PUMS files was compromised. Also, I often merge individual- and household-level data with information for geographic areas (for example MSAs) to do multilevel analyses. Therefore, changes in the level of geographic detail could also interfere with my research objectives.

**Richard Wright, Dartmouth College, Geography (Faculty)**

I do research occupational segregation by race/ethnicity/nativity. I also use the income data to study immigrant impacts on the native born. I use PUMAs to build local labor markets. If these data were compromised from the versions we have for 1990, that would compromise my ability (and that of related research) to investigate employment, income, and migration effects of the foreign born. These are crucial policy issues facing the USA

**Richard Kleban, Ohio State University, Sociology (Graduate Student)**

I focus on understanding the relationship between social disadvantage, race, and community crime. Without detailed race, occupation, income, and geography information I am unable to measure the key concepts that are thought to link social disadvantage, race, and crime. To lose this detail would seriously handicap our efforts to understand how the social disadvantage of minority communities leads to high crime.

**Tomas Jimenez, Harvard University, Sociology (Graduate Student)**

Questions about race and ethnicity are vital to my research. The demographics of this country are changing quite rapidly and complete data on race and ethnicity are crucial for understanding the impact that these changes are having and will have on the United States.

**Ellen Meara, Harvard Medical School, Economics (Faculty)**

A significant portion of my research deals with income inequality. The PUMS is really the best source of information for deriving reliable, comparable, measures of inequality over time. By reducing the level of detail on income, geographic detail, age, or race, this would limit one's ability to assess at what level inequality might matter (states, MSAs, PUMAs, within race, within age groups, etc.).

**Susan Brudvig Other Nonacademic**

We use the data for school districts' applications. Geographic detail is INDISPENSIBLE for local areas and policy decisions. Also, we have found socio-economic characteristics to be CRITICAL in accurately projecting school-age populations for small geographies. We need this data for local governments!

**Giovanni Peri, Bocconi University, Milano, Economics (Faculty)**

I study local human capital externalities in US cities, therefore reducing information on income and schooling would be very harmful. I inquire the effect of the neighbors on the productivity of a worker, so that the very detailed geographical level is very useful.

**Confidential Response**

Much of my research concerns issues of racial segregation and urban poverty. Without detailed information on race, income, and geographic identification it becomes substantially more difficult to test hypotheses. For example, I am interested in determining the impact of central city gentrification on the poor. To adequately conduct this important research, it is essential to know (1) what city, if any, a respondent lives in, (2) what part of that city, if it is large, (3) detailed information on that person's occupation and income. Item (3) is what allows me to differentiate persons of low "current" income from persons of low "permanent" income. My work on segregation and its impacts on individual outcomes requires detailed geographic identification and accurate measures of income.

**Confidential Response**

We use PUMS for a variety of purposes. The whole point is to get the kind of detail and cross tabulations that are not obtainable in standard tables. My past research on the elderly in Minnesota consistently shows powerful differences among the elderly population when broken down by age, i.e. people 90+ versus those 70-74. These in income, living arrangements, etc., show service providers that just talking about "the elderly" is not very useful because of these great differences within the elderly group.

**Confidential Response**

Most of my research involves analysis of income and occupation so any reduction of that information would make it necessary to use more detailed data that are not as extensive as the PUMS. Also, as the population ages and becomes more diverse it will be necessary to look at these analyses in terms of age and race which would again necessitate detailed PUMS data.

**Darcy Hango, Ohio State University, Sociology (Graduate Student)**

I do population/demography research in sociology and find it very distressing that ages might be grouped into categories. The researcher needs the flexibility to be able to choose their own cutpoints for ages. For example, grouping together ages 15 to 19 is not very useful if you are trying to study high school completion.

**Roland Daeumer, The Pennsylvania State University; Population Research Institute, Demography (Post-doc)**

The Social Security administration expects us to evaluate pension policies. This implies the necessity to analyze retirement decisions at every age of individuals 65 years of age and older. Grouped age data would severely reduce the predictive quality and accuracy of our forecasts regarding retirement decisions of the elderly population and their outcomes. This is especially the case if we are to evaluate retirement policies intended to alter pension and social security eligibility. Thus, the successful implementation of policies aimed at securing the financial integrity of government subsidized health care and retirement programs will be at stake. will be

**Confidential Response**

I perform analysis on economic opportunity which requires detailed information on the income and occupational categories, and then I look into certain racial/ethnic/age categories by MSA and analyze changes over time. So, all of the variables that are in question would affect my research, and compatibility of the data over time is extremely important to me.

**Confidential Response**

This would drastically reduce the ability of researchers attempting to conduct cross-level research in many areas of sociology.

**Confidential Response**

I have used census data in many areas of my research. Calculation of single year of age specific labor force and disability rates has been important in my work on the effect of health on early labor force exit, the geographic detail has been essential for my work on local labor markets, the industry detail has been important for understanding changes in local labor markets, the occupational detail has been essential for my own work on racial earnings differentials, and has been essential for some of my students work on gender differentials. The geographic and age detail has also been central to my work on health differentials across rural and urban populations.

The main advantage of the census over the SIPP or CPS is its large sample size. This allows researchers to use the data to study phenomena that one can not study with the CPS. The geographic, occupation, age etc detail available on the Census is crucial for all of this.

**Seth Ovadia, University of Maryland College Park, Sociology (Graduate Student)**

My current research is on occupational and residential segregation across metropolitan areas. My plan for the next stages of this research are highly dependent on the availability of 2000 Census data which are comparable with the 1980 and 1990 data, as change over time is an essential part of both my and other segregation researchers' plans for doing policy-relevant analysis. Any changes in the occupational and geographic data formats would make such work simply impossible and cut off a vibrant research tradition before it could take root.

**Confidential Response**

As an economist who does research on the changing economic position of women, it is absolutely necessary to know not only the age of each individual, but particularly their specific occupation (to understand whether the differences in economic outcomes are due to occupational segregation versus lower wages, etc. within occupation). In addition, it is important to have specific income information; grouped income data would make it very difficult to back out women's relative positions.

**Suzanne Duryea, Inter-American Development Bank, Economics (Researcher)**

My research utilizes the PUMS to establish a baseline of outcomes for comparison with household survey data and census data for Latin America. We are currently examining issues of social exclusion in Latin America and it is critical that reliable information on race and ethnicity be available in the PUMS. Perhaps more importantly for my research is the potential indirect effect of the elimination or reduction of data coverage in the PUMS. In high-level dialogues with governments and statistical offices in LAC, the PUMS is used as a model survey for detailed socio-economic questions. Other governments will be much less willing to include critical questions in their surveys if the US government restricts either the collection or accessibility of certain socio-economic and demographic variables.

**Jamie Strickland, Western Kentucky University, Geography (Faculty)**

I am a geographer who is primarily involved with aging issues. Therefore, I am extremely concerned over a possible decision to reduce BOTH the geographic detail as well as the age detail in the 2000 PUMS. PUMS represents one of the very few datasets which allows me to examine older populations at the state and substate level. I would like to stress that I consider anyone over the age of 40 as being a part of the "older" population. Only considering people over the age of 65 as being part of the "older" or elderly population is incredibly short-sighted. Specifically, my research considers the changing nature of retirement and its potential impacts on the spatial mobility of older persons. It is becoming increasingly common for retirees to be under the age of 65. In addition, not all individuals over the age of 65 have left the paid labor force. These processes do not occur evenly over space nor do they have the same economic and social impacts in all locations. As such, there are important local, state and national policy implications for this research that absolutely require a detailed age and geographic coding. I certainly respect the Census Bureau's desire to maintain a high level of confidentiality for respondents, but I do not feel that presenting a "watered down" version of PUMS for 2000 will accomplish this goal.

**Eric Millis, Federal Reserve Bank of Dallas, Economics (Support)**

There is a particular interest now in the effects of policies encouraging the elderly population to enter the labor force. Clumping all 65+ categories together, which I have heard being mentioned as a possibility, would be a significant setback to such studies.

**Gregory Reeves, The Kansas City Star, Media (Nonacademic)**

We use PUMS for detailed information on race and income. Those variables are crucial.

**Scott Adams, Michigan State University, Economics (Graduate Student)**

The research that I do tends to focus on the comparisons of outcomes at the city level. For example, one project that I am working on concerns access to health care for the elderly. Because certain characteristics of cities (like public transportation) influence this and vary by city, I exploit this cross-city variation. Census data is quite helpful in terms of providing the basic characteristics of city residents, such as race, age, and income. I was hoping the 2000 census would allow me to do this. I could use the Current Population Survey (and have done so), but some of my estimation techniques require very large samples.

**Joshua Rosenbloom, University of Kansas, Economics (Faculty)**

I have used PUMS data for two purposes. One is labor economics research on the impact of childbearing on female labor force participation and work effort. This research used the occurrence of twins in the first birth as a "natural experiment" but we had to stop our analysis after 1980 because the lack of data on quarter of birth for children made it impossible to identify mothers of twins in census data after that date. The other use is for economic history research on a variety of aspects of the geography of American labor markets. I am interested both in measuring economic integration--for which geographic,

occupational, ethnic and income detail is needed--and in studying long-run trends in migration for which all of the same is needed along with detailed information about age and family structure.

**Confidential Response**

Since I am interested in neighborhoods and neighborhood effects, linking IPUMS' individual records to neighborhood level data is very important to me. The Census is the only source of reliable neighborhood characteristics.

**Liana Sayer, University of Maryland, College Park, Demography (Graduate Student)**

I have used PUMS data to examine differences in rural and urban poverty between the North and South regions of the U.S. I have also used the PUMS-L, which allows examination of rural non-metro labor market areas. If the PUMS contained broader geographic areas, I would be unable to examine how the link between labor market areas and poverty rates for rural areas has changed since 1990.

**Thomas Wilson, Florida Atlantic University, Sociology (Faculty)**

I am interested in the interplay of race, class, and gender, with particular emphasis on class. The usefulness of IPUMS data for me resides largely in its detailed class indicators, including occupation, income, SEI, etc. A reduction in this detail would be catastrophic. Further, it is hard to imagine how confidentiality could be compromised by maintaining the current level of detail in these data.

**David Kantor, Institute for Policy Studies, Johns Hopkins University, Other Academic (Support)**

1: I'm not prepared to answer question 5. 2: The geographic identification is poor enough already. We don't need to make it any less detailed. I would prefer in IMPROVEMENT in the geographic identification.

**Confidential Response**

We use The Census data for Program administration, Community level Industrial Recruitment and other Economic development activities. We are the source link with the local & regional Chambers of Commerce, Industrial & Workforce boards. Prior to the 1990 census we have request special reports for the local areas and wait for them to be compiled. Since the development of the IPUMS Web server, we have been able to access the information in a more direct timely manor. And I am represented by the Oklahoma Second District Congressman who started this mess. But he has never listened to anyone anyway.

**Joshua Angrist, MIT, Economics (Faculty)**

A key feature of my work is the ability to identify cohort (exact age), states of residence, and states of birth. Also important is the comparability of data on income, employment status, weeks worked, and schooling.

**Confidential Response**

I do research on the structure of wages, and changes in wage inequality: a reduction in ANY of these characteristics essentially makes the Census 2000 data useless for research in this area. Perhaps this is exactly what proponents of reducing these details would want, because I find it absolutely crazy to think that characteristics like age, ancestry, race, and occupation could ever be used to identify specific individuals. These variables are vital to expanding our knowledge of how the labor market works. The confidentiality issue is a red herring.

**Confidential Response**

I specialize in labor market research -- anything related to earnings or occupation must be as finely detailed as possible. Race, age, and gender are also important as control variables.

Geographic information and ancestry are less important for my research.

**John Hipp, University of North Carolina, Chapel Hill, Sociology (Graduate Student)**

My research focuses on internal migration and its effect on labor markets. Thus, geographic detail is extremely important for me in looking at the location of movers. In addition, information on both occupation and income are necessary for this. As well, much of this research looks at migration by racial/ethnic groups, so losing this level of detail (i.e. particular Latino groups) would dramatically reduce the usefulness of the data.

**David Cotter, Union College, Other Nonacademic (Faculty)**

I am particularly concerned about the reduction in detail in the geographic codes. These have been absolutely critical to both of my major research projects which examine variations in gender inequality and poverty across labor markets. Without this information there would be no way to construct reliable estimates of several key indicators, nor would it be possible to extend

this research into the future without the PUMS. Much of the work we have done already with the 1980 and 1990 PUMS needs to be extended with the 2000 data, and it plays a central part in several grants we have proposed to NICHD and NIH.

**Confidential Response**

I am a university statistical consultant but have helped people work through compatibility issues between 1980 and 1990 data for studying ethnicity and education, and changes in employment demographics between these years. Addition of the 2000 data will require as much consistency as possible; we already had several problems with compatibility between 1980 and 1990!

As the US population ages it seems nonsensical to group ages over 65. How else are we going to understand changes in aging structure from 2000 forward?

**Confidential Response**

I use single year of age data and data on race and ethnic group to construct life tables in specific geographic areas. Grouped age data would make this difficult.

**Confidential Response**

I currently work with a combination of ancestry, race, and age variables. Grouping age would preclude me from determining whether the age structure of various elderly ethnic populations differ. For example, are elderly Japanese Americans much older, on average, than elderly Chinese Americans, due to dramatic differences in foreign-birth, duration in the U.S., and life expectancy? Are Japanese Americans concentrated in the "oldest old" group (those 85+), compared to elderly Chinese Americans who may be more likely to be clustered in the 65-74 group? What of elderly Italian Americans, compared to Mexican Americans?

Reducing geographic detail would also complicate research on, for example, Marriage Markets, which generally attempts to look at the most detailed information to ascertain area data most closely linked to risk populations. For example, having an area of 100,000 is a much more reasonable marriage market to contemplate than an SMSA.

**Mark Stern, University of Pennsylvania, History (Faculty)**

In our study of historical change in work, family, and community during the 20th century, Michael Katz and I would find geography "most sacrificeable" because we will be using the pums to examine metro regions as our basic geographical unit. However, losing detail on individual characteristics that make it difficult to compare 2000 with the earlier files would be "catastrophic" (to use the conventional term from the survey) to the project as we conceive it.

**Susan Losh, Florida State University, Sociology (Faculty)**

I have already run into excised variables on NSF datasets (race, region). It has been \*catastrophic\*. I am now trying to get the complete data because my results will not be academically acceptable with these variables omitted.

Please keep the data intact. Don't group if at all possible because different researchers will group data different ways depending on their research agenda.

I think the identifiability issue is somewhat overemphasized. Just be sure the geographic units are large enough, perhaps truncate a few outliers (the local millionaire) and keep these research resources generally intact!

Thank you

**Mary Washington, Lehigh University, Sociology (Faculty)**

I specialize in the historical changes in racial classification, mixed-race/ethnic identity and the formation of pan-racial and ethnic groups as they relate to social status. Reduction in the detail of racial categories and occupational classifications reported, as well as single year age would greatly harm my ability to make historical projections or.

**James Cassell, University of North Carolina at Chapel Hill, Sociology (Graduate Student)**

I am particularly interested in changes in the race/ethnic distributions of income, occupations and industries over time for low-levels of geography. The PUMS is the only source of such data that I'm aware of.

**Jeff Elsworth, Purdue University, Economics (Graduate Student)**

I do wage and occupation research in the hospitality industry. Loss of this detailed occupation data would severely limit my investigation of what drives wages and other occupation specific factors in the food, lodging and travel industry.

**Confidential Response**

I do a lot of small area economic and health related research. Therefore the any reduction in degree of detail with regards to geographies a/or detail on individuals would be very harmful.

**Kristina de los Santos, Ohio State University, Sociology (Graduate Student)**

My work focuses on Asian Americans and examines specific ethnic groups under this rubric. PUMS is one of the only sources with large #'s of Asian Americans, and therefore is indispensable to conducting quantitative analyses.

**Confidential Response**

I use geographic data to overlay macroeconomic variables on my microdata. For instance, unemployment rate in the area that a person lives in often proves useful in explaining labor market behavior.

**Confidential Response**

My research focuses on Asian Pacific Americans and specifically, on APA neighborhood formations and characteristics. Reduction in PUMS will seriously affect my ability to look at various APA ethnic groups and their community patterns with respect to socioeconomic and other demographic qualities. More generally, my research interest is racial economic inequality and to be able to examine this issue on the micro-level requires that detailed data be available on the smallest geo level - PUMAs - possible.

**Stephen Tordella, Decision Demographics, Demography (Nonacademic)**

I do labor force research that depends upon having access to exact occupation.

**Walter Carroll, Bridgewater State College, Sociology (Faculty)**

Grouped age, income, and occupational data would seriously compromise my research on historical family change and the processes of immigrant adjustment to the United States.

**Confidential Response**

Occupation, age, income, and race are key variables in controlling for net diffusion of population across geographical space. Less detail on these variables would mean that one could control less well for these demographic differences, which potentially increases the bias of the analysis. In particular, standard regression approaches (in spatial analysis) are susceptible to this problem. Semiparametric approaches, such as matching on propensity scores, which is a key approach to get better results than in regression, crucially rely on these background variables to reduce bias.

**Arthur Wilke, Auburn University, Sociology (Faculty)**

Grouped data precludes developing what may turn out to be more meaningful classifications of people based on a number of indicators.

**John Hall, Mathematica Policy Research, Other Nonacademic**

I use PUMS in designing samples. It is essential in those cases where the design must consider variation at the sub-state level.

**Susan Mayer, University of Chicago, Sociology (Faculty)**

Reducing PUMS income detail would prevent the comparison of census data with income data from other data sets thus eliminating one of the most important checks on the quality of data used in estimating poverty rates and income inequality. Reducing geographical detail would eliminate the ability to do some research on inequality and poverty and research on the effects of economic and racial segregation because for some purposes the Census is the only data set large enough for reliable estimates.

**Amy Goodin, UNM Institute for Public Policy, Geography (Researcher)**

I conduct policy assessments of hazardous facilities as they relate to CERCLA, RCRA, and property value impacts, thus a variety of indicators at sufficient detail is required to assure informed decisions are made with the data I collect (i.e. representative samples, etc.).

**Edward Ratledge, University of Delaware, Demography (Faculty)**

In doing policy research for the Delaware General Assembly and the Executive in Delaware I frequently use the PUMS to answer questions that cannot be answered by tables. It is also our primary cross-check with the CPS. Most recently, I used the info on households with school age kids and age of housing to develop an estimate for impact fees. Those estimates materially changed what was going to be used in the absence of the information.

**Jeff Grogger, UCLA, Economics (Faculty)**

Several years ago I proposed that twin births could be used as an instrument to estimate the effects of unplanned fertility. The results, published in the American Economic Review, were quite successful. Without detailed age data (at least age in years and quarter of birth), this strategy will be unavailable to chart changes in the socioeconomic consequences of unplanned childbearing. We would lose a tremendous resource for analyzing the consequences of this important social problem.

**Jeff Claassen, Fort Worth Star-Telegram, Media (Nonacademic)**

PUMS data allows is essential for preparing crosstabs on categories that are not in the Census 2000 standard release. For journalistic purposes, there are many trends to examine that can only be done through a PUMS analysis. That includes the demographics of long-haul commuters and other crosstabs.

**Susanna Loeb, Stanford University, Economics (Faculty)**

Geographic detail is important for my research for matching people to labor markets and government services (child care or schools for examples). Occupational detail is important so that I can identify teachers (elementary, secondary etc) and child care worker as well as identify alternative occupations for workers who choose these fields.

**Confidential Response**

We use the data to develop quota samples and make estimates of people impacted by state and local level programs. Our work would become almost impossible to defend w/o the PUMS detail.

**Russell Davis, Louisiana State University, Sociology (Graduate Student)**

I study rural labor markets. Without detailed occupation codes, income, and geographic details, it becomes impossible to identify key demographic groups. With out the detail, all you have is a bunch of people and areas that all look the same. Large population aggregation often times eliminates the place specific detail of rural areas more so than metro.

**Mark Yeisley, Florida State University, Sociology (Faculty)**

I specialize in inequality research as it relates to crime and crime policy. I require geographically-specific data regarding income, specific occupation, education, race, and other socioeconomic variables relevant to inequality.

**Confidential Response**

I live in, and do some of my research in, a sparsely populated state in which age and other demographic characteristics are crucial to analyses of the potential or lack thereof of various areas of the state. At a PUMA size of 100,000 it is possible to use Census data as part of such analyses: at anything more we cannot do analysis on anything but a state-wide level.

**Confidential Response**

Detailed occupations are used to identify individuals in science occupations. Ancestry is very important in categorizing immigrants and immigrant assimilation.

**Andrew Beveridge, Queens College and Grad Center CUNY, Sociology (Faculty)**

Since 1993 I have had a contract with the New York Times. We have used the PUMS data, often from the IPUMS project, to do a variety of analyses. To lose the sort of detail contemplated in terms of PUMA's (NYC is divided in 54 PUMA's that basically correspond to Community Planning Districts), ancestry, occupation, foreign born place of birth, income, age, etc. would be a disaster. We have had at least six analyses make it to the front page of the Times and one to the front page of the Week In Review. In each case, very detailed PUMS categories were used. For example, for a story on the earnings of male and female college graduates we used single year ages to look for those who had just graduated from college. For a project on income of blacks and whites in Queens we carefully set-up the households in such a way that included PUMA's, as well as age and detailed household characteristics. For a project on earning by occupation for blacks and whites we used the 1950 detailed occupation codes from 1950, 1970 and 1990. Such an analysis would be impossible to replicate for 2000, if these restrictions go into effect.

Currently, I am working on a paper comparing immigrants to New York from 1910 to those here in 1990. The sort of restrictions planned would preclude an analysis using 2000 data.

**Lu Chou, Data and Program Library Service, University of Wisconsin Madison, Other Academic (Support)**

I am a data librarian and help graduate students, researchers and faculty members on my campus by providing PUMS data to them. I know PUMS are used on our campus for policy research and academic research. Aging and poverty researches are big on our campus. As a service provider I would not want to see the detail level of PUMS be reduced and my users lose some very crucial data for their researches.

**Michael Sullivan, Freeman, Sullivan & Co., Sociology (Nonacademic)**

I specialize in the analysis of the statistical representativeness of jury venires sampled from the general population. The PUMS data allow estimation of the fraction of the population falling within given race/ethnic groupings, over 18 who are citizens with different levels of language capability within given geographical locations.

I can't imagine trying to do this sort of analysis without this level of detail.

**Diane Cowper, Dept of Veterans Affairs, Demography (Researcher)**

My research focuses on the migration patterns of retired veterans (and, subsequently, the impact of that migration on the VA health care resources at their destination communities). Life-cycle is important in the types of moves older people make -- early retirees are usually amenity migrants, while the old-old tend to be assistance migrants. It is important to know the types of migrants an area is attracting (or, losing) to plan for adequate services for those individuals. Therefore, age data are extremely important, as are other sociodemographic variables, including (but not limited to): race, gender, income, education, marital status, living arrangements, etc.

**William Bridges, University of Illinois at Chicago, Sociology (Faculty)**

The best way of tracking information on specific JOBS in the economy is by a crosstab of detailed occupation by detailed industry (without regard to geography). This is essential for understanding segregation by race and gender. Any tinkering with the current categories would be a disaster.

**John Blandford, University of Chicago, Economics (Post-doc)**

I research the labor effects of sexual orientation, specifically how sexual orientation impacts earnings and occupational outcomes. Because of the ability (beginning in 1990) to identify same-sex, co-habiting, unmarried partners, the PUMS provides one of the few reliable sources of information on lesbian and gay workers. The other sources for data are the General Social Survey and the National Health and Social Life Survey, both of which I also use extensively. These too are invaluable but suffer from two limitations for the purposes of my research. Firstly, they provide only categorical income data and, secondly, the sample of the target population is rather limited. In contrast, the PUMS provides a full complement of earnings and supporting data on nearly 10,000 full-time workers who identify as same-sex partners. The 1990 PUMS has proven an invaluable complement to other sources of employment data on this population. In particular, occupational clustering patterns significantly determine differential earnings outcomes for gay and lesbian workers, but these clusters are highly nuanced and an accurate estimate of their impact requires occupational data down to the three-digit level (505 titles). Studies on this topic that rely on GSS/NHSLS data measure large wage differentials attributable to sexual orientation. Because of the limited sample size, however, they are unable to generate a meaningful measure of the impact of occupational clustering patterns as a determinant of the earnings differentials. The outcome has been biased estimates of the impact of sexual orientation in labor-market outcomes. It has been precisely the availability of the PUMS, with its full complement of nuanced employment-related data, that has allowed me to conduct research in this area. Indeed, I eagerly anticipate the release of the 2000 PUMS which, if it provides complete data compatible with those of 1990, will provide the first opportunity to see how the economic status of this population has changed over the course of a decade. Needless to say, it would be a horrible loss to many researchers were the Census Bureau to release 2000 PUMS data in a form that would make cross-year comparisons impossible.

**Michael Haines, Colgate University, Economics (Faculty)**

I will have to revise Historical Statistics of the United States using the 2000 PUMS. Since I believe that the published material will be very limited, I think that a very detailed PUMS will be indispensable to anything with the census that requires any geographic, age, race, occupational, income, marital status, nativity, or other detail.

**Charlotte Steeh, Georgia State University, Sociology (Faculty)**

Loss of detail deprives researchers of the flexibility to create the groups that are most relevant to their hypotheses. In addition, predefined groups rob analysts of the powerful set of statistical methods that rely on interval and ratio level measurement.

**Christopher Lyttle, Rush Primary Care Institute, Sociology (Researcher)**

Census data, including PUMS, are important to us in our research into the sociological covariates of asthma. Sometimes we use these data to provide the sociological context of medical events extracted from other data sets. In such analyses we do not match individuals, so it is important to have area as specific as possible. For some of our work, counties are the minimum level of analysis, for other projects we work with Chicago community areas.

**Charles Brown, University of Michigan, Economics (Faculty)**

For research in labor economics, much of the point of PUMS is the detail. If I want to estimate relationships using only variables that have many respondents, I can use CPS. PUMS is useful for those who want to study relatively small groups (e.g., individual national origin groups) or groups that become small because of cross-tabulation (e.g., older black workers). The direction of the proposals would endanger both types of studies.

**Ada Haynes, Tennessee Tech University, Sociology (Faculty)**

Since my analysis is a longitudinal study of non-metropolitan areas any changes would be devastating.

**Confidential Response**

Q#6: Occupational categories can always be collapsed by the researcher, later, but if it's collapsed in the data set, then any refined analysis would be impossible. Q#7: I use both SEI and manufacturing codes, as well as historical comparability. Q#: PUMAs that also allow for classification by region, state, etc. are important as well.

**Confidential Response**

Comparison of census demographic data to sample survey demographic data.

**Nadine Misiaszek, San Diego State University, Other Academic (Support)**

Pre-grouped data is a limitation. I need to be able to group the data in the ways that the survey requires. For example, a study on child care needs to be able to get at children to age 12 (not age 18); a study on binge drinking needs to pull out data for unique years within the age span of 18 to 30.

**Joyce Jacobsen, Wesleyan University, Economics (Faculty)**

One of the main things I do is to analyze workers by occupations within industries, including calculating segregation indices and comparing them across time and space. Having a lot of categories, and historically comparable categories for occupation and industry, is very important for me.

**Carl Mason, UC Berkeley, Demography**

The most important use to which I put IPUMS right now is in teaching an undergraduate course on the History of Immigration. In that course \*undergraduates\* use IPUMS data to investigate labor market mobility, intermarriage patterns and residential segregation. Since the US is presently in the midst of a wave of immigration that in so many ways is comparable to the one that took place in the late 19th and early 20th centuries, the comparability of data between that earlier time and the present is extremely valuable. To lose detail on ancestry, occupation, race or income in the 2000 census would mean that the clock would stop in 1990. For undergraduates, this will soon be ancient history.

**Robert Morgan, Houston VAMC/Baylor College of Medicine, Other Academic (Faculty)**

I do a lot of work on health care utilization and outcomes of care among the elderly and poor (e.g., Medicare, Medicaid, and uninsured). In particular, my current work involves differences associated with geographic and race/ethnic variability among the elderly. Thus, I need to be able to compare data across years (especially recent years, e.g., since 1980), across geographic regions (at least down to PUMAs of 100,000), and across race/ethnic categories; and I need to be able to match data for specific geographic areas with aggregated data obtained from other sources. The finer-grained the data available to me through PUMS are, the more flexibility I have in developing and answering my research questions.

**Jon Lorence, University of Houston, Sociology (Faculty)**

My research has focused on occupations and income inequality. Any reduction in the level of specificity of such information would make it impossible to conduct any research that would be useful in this area. Failing to provide detailed PUMS information would gut social science research in the United States.

**Confidential Response**

A less detailed PUMS with harm greatly my current and future research, because I am focusing on comparing the trends in self-employment among Hispanic groups, especially Puerto Ricans, Cubans, Mexicans..

**Confidential Response**

I study the characteristics of immigrants, and mainly their earnings pattern. The reduction in detail on age, income and country of origin would affect badly the analysis. It is very difficult to work with variables that are defined using groups, that is especially true when we talk about numbers (age, income) and may not be such a very big problem when you talk about categories (occupation etc.) - though depending on the study a range not broad enough for the categories may render the variable itself useless.

**Karen Woodrow-Lafield, Mississippi State University, Demography (Faculty)**

I would like to use this area to give more detail about my response to item 9. I am reluctant to see any delay to release of public-use microdata files from the 2000 Census, even if a less detailed dataset were initially released. A fuller microdata file release could be made available to users who had purchased the earlier files. Another possibility is that more detailed microdata files could be available through the various Census Research Data Centers, or through the NCHS Research Data Center.

My research specialization in immigration means I am very interested in highly detailed data as to country of origin, race and ethnicity, timing of immigration, and various measures of socioeconomic assimilation within the United States and specific geographic areas. The PUMS files are already limited for several purposes for which current data sources are needed to study recently immigrated origin groups, some of which are highly concentrated within certain metropolitan areas. Asian immigrants are more heterogeneous than our studies may examine.

**Elizabeth Mustaine, University of Central Florida, Sociology**

I specialize in crime-related topics. If I did not have detailed information about the size of geographic locations, I would not be able to do my research. Also, categories like age, race, occupation, and income are always important indicators of crime rates. With out specificity in these areas, criminological research would, essentially come to a stand still. At best, we would simply be unable to make comparisons between research done on this data and research done previously. Consistency is also important. The only way to provide a better understanding of a social issue, or to make more efficient and effective policy is to continue to study an issue over time, using similar/the same measures. My view is that changing the way variables are categorized, and losing the specificity of variables will be a huge loss to the social sciences and, thus, to policy makers.

**Leif Jensen, Penn State University, Sociology**

I have done work on poverty among elders, and grouping age cats, esp at oldest ages would be extremely problematic. I also do work on poverty and income, and grouping income categories would greatly reduce the flexibility to experiment with alternative poverty definitions, and would greatly reduce ones ability to examine inequality. I also am a rural sociologist and even as is the PUMS is woeful in providing the geographic detail needed to understand the circumstances of rural populations in the US. It would be devastating to lose the geographic detail on rural/urban, metro/nonmetro, size of metro, state, etc., etc.

**Confidential Response**

PUMA size increase would preclude the ability to run custom crosstabulations for New York City neighborhoods (community board districts are for the most part equivalent to current PUMAs) and some smaller counties in the New York region.

**Catherine Weinberger, University of California Santa Barbara, Economics (Researcher)**

Careful research based on detailed information is vital to the design and implementation of reasonable government policies. When I study differences in earnings between technically trained immigrants from different countries and U.S. natives of different ethnicities, detailed information about country of origin, ethnicity and earnings is essential. When I compare the labor market conditions faced by men and women in different fields of engineering, very detailed information on occupation and age is essential. When I studied the relationship between mathematics education and later labor market outcomes, it was valuable to have detailed information on occupation and earnings that could be matched across Census cohorts.

**Rebecca Dennison, University of Minnesota, Other Academic (Researcher)**

unknown

**Masao Suzuki, Public Policy Institute of California, Economics (Researcher)**

My research is on economic achievement of racial and immigrant groups, so I would be most concerned with any loss of detail about race, ancestry, and immigration. Since my work is mainly historical, comparability with past IPUMS/census records is critical.

**James Walker, University of Wisconsin, Economics (Faculty)**

The PUMS plays a major role in analyzing individual determinants of migration. Geographical locations need to be measured with as much detail as possible to correctly characterize the source and destination locations. That is, it is important to know whether an individual moved to New York City versus upstate New York, or for that matter that the destination was Houston and not Dallas. Aggregation to census region, or state is too broad to be of anything other than the most basic descriptive use. Summary by rural/urban is also too broad. Reduction in the geographical level reported in the 2000 PUMS will have a serious deleterious effect on migration research.

**Confidential Response**

Some of my research is on the marriage patterns of older women (especially related to the Social Security re-marriage penalty). I need to calculate marriage rates by age for women ages 55-75, and this would be impossible with age-grouped data.

Detailed geography data are also important for two reasons. First, the Census is the only data set researchers have that is large enough to be representative at the state level. Second, identifiers for PUMAs make it possible to adjust for differences in local cost-of-living, which is important in any study concerning the economic status of the elderly.

**Mark Fossett, Texas A&M University, Sociology (Faculty)**

My work often focuses on comparisons across metropolitan areas and/or tracking patterns in metropolitan areas over time. Loss of geographic detail would make this work much harder.

As it presently stands, my wish is to get easier access to even greater levels of geographic detail (e.g., census tracts within metropolitan areas).

**Aimee Chin, MIT, Economics**

I have been doing some research on immigration, and it would be devastating if I would no longer have details on country of ancestry and year of immigration.

Also, in all my research, I prefer using as fine a geographic category as possible. The state level is too large, especially because large states tend to be very heterogeneous. When examining the effect of policies, sometimes one needs to be able to distinguish NYC from Buffalo, for example.

**Frank Lenk, Mid-America Regional Council, Economics (Nonacademic)**

Our subject of study is the Kansas City metropolitan area. Geographic detail must at least maintain large metropolitan areas.

**Halliman Winsborough, ICPSR and University of Wisconsin, Madison, Demography (Faculty)**

Truncation of the age coding would obviate research on the oldest old, the most important policy age category in this period. Reducing geographic coding will nip in the bud the fruits of GIS.

**Confidential Response**

We are particularly interested in PUMS because it allows us to better understand age/race/income dynamics, and age/migration dynamics.

**Francine Blau, Cornell University, Economics (Faculty)**

The detailed occupational categories in the PUMS have been used extensively by scholars (including me) to study changes over time in the extent of segregation of women by occupation. A recent study I completed examining determinants of marriage rates used the variation in labor market conditions across MSAs to study their effects on marriage. Data at the PUMA level was necessary to ensure comparability across census years due to changes in MSA boundaries. Important differences were found by race making this data also very important. The study focused on younger individual and controlled in a detailed way by age.

**Diane Lauderdale, University of Chicago, Demography (Faculty)**

My interests are in cohort effects on aging; eliminate age detail were preclude most of my use of census data. I also need state-level socioeconomic data for the elderly.

**Michael Conaway, Institute for Social Science Research, University of Alabama, Demography (Researcher)**

My interest in the PUMS data lies in using it to characterize small geographic areas of the state and as a ground truth for correcting for the inevitable selection biases within those areas. The geographic breakdown is already too large to be useful to me in many cases.

**Harold H. Kassarian, UCLA and Cal State University, Northridge, Other Academic (Faculty)**

Reliability and validity of research results would be lessened, along with credibility of the findings.

**James T. Ault, III, Creighton University, Sociology (Faculty)**

We are exploring spatial analysis of demographic attributes (both contemporary and historic), so detail is essential to enable us to match data from disparate sources. Ethnicity, age, and occupational details increase the likelihood of reasonable and logical comparisons.

**Elsie Harper-Anderson, UC Berkeley, Other Academic (Graduate)**

Data which has details on race is limited. Especially when studying a specific region or city. This is one of the few datasets which provided enough detailed to both break down into the region in question and the race variables needed while maintaining enough cases to be statistically valid. Without it, my last three research projects would not have been possible.

**Leann Tigges, University of Wisconsin-Madison, Sociology (Faculty)**

I think of my research area as involving the sociology of work and labor markets. Thus, it is crucial to me to have all of the social and economic characteristics of individuals and as much detail as possible on the labor market area in which they work. We need to understand better the relationships between human capital, occupations and industry, family and personal characteristics, and the community or local economy that matches workers with their jobs. PUMS has provided rich data to understand these relationships across time and space. Restricting data on these dimensions would lead to great difficulty understanding the dimensions of inequality (especially those associated with race, gender, and age).

**Confidential Response**

Reduction or simplification of race, age, and ancestry data would greatly reduce my ability to collect data on intermarriage and other structural variables concerning immigrant assimilation after 1965.

**Robert Fairlie, University of California, Santa Cruz, Economics (Faculty)**

1. I would not have been able to conduct my previous research on the impact of immigrants on the U.S. education system and labor markets without the geographic detail. 2. I would not have been able to conduct my previous research on the returns to education without the income, age, occupation, and industry detail. 3. I would not have been able to conduct my substantial research on racial, ethnic, and immigrant inequalities in the U.S. without the income, race, Hispanic origin, and ancestry detail.

**Confidential Response**

I study various aspects of the "Digital Divide" as they relate to all age groups and socio-economic status for small areas. I rely on PUMS data to teach my students in research methods and government information courses.

**Charles B. Nam, Florida State University, Demography (Researcher)**

I am a developer of the Nam-Powers socioeconomic scale of occupations and provide a decennial update of the scores, which are used by a broad research public. The scale depends on a detailed occupation classification (although comparability with the occupational classification in earlier censuses is not important), as well as a fair amount of detail on education and income (determinants of the score for each detailed occupation). Without the detail for those three variables, it would not be possible to create the scale (nor would it be possible for competing scales, e.g., Duncan SEI, to be constructed).

**Helen Marrow, Princeton University, Sociology (Undergraduate)**

I do work on immigration, and so reductions in race/ethnicity, ancestry, national origin group, and/or birthplace categories seriously precludes my ability to evaluate foreigners and minorities according to census data.

**Judith Rowe, Princeton University (retired), Economics (Support)**

Significant policy research has been done using the PUMS in addition to academic research. It would be a great loss to lose the comparability that we have had for so many years.

**Confidential Response**

I do research on migration and work. I need the most detailed data I can find on where people live now and where they came from. I'm interested in the effects of work on migration, so I need detailed information here as well. I'm especially interested in the self-employed, too. My interest in migration and self-employment are not necessarily always linked but I might want to look into it. Another important variable would be age. With the changes in age at retirement, I would really like more detailed information on those over 65, more detailed than the categories mentioned earlier.

**Scott Czepiel, SUNY - Stony Brook, Sociology (Graduate)**

Occupation codes are absolutely requisite for work in social stratification. Broad groupings are sometimes helpful but more often than not, researchers must recode the 505 list themselves for very specific reasons. Eg., % of workers in a given occupation who are women, or who earn higher than the average median income of all workers. Such stats are indispensable and would be irrevocably lost if detail is lost in the occ codes.

**JoAnn Dionne, University of Michigan Library, Other Academic (Faculty)**

I am a data librarian and, while I don't do research myself, have over 20 years experience in helping academic researchers. I've answered as a combination of most of the users I've helped over the years. You may want to disregard my answers to 4.

It is critical that historical comparability be maintained. Age, income, race and occupation have been among the most important variables in \_any\_ of the analyses that I've provided support for.

**David Brownstone, University of California, Irvine, Economics (Faculty)**

I do research on transportation demand, and PUMS is the only source of joint distributions of socioeconomic variables. It is critical to have this available at the highest level of geographic desegregation.

**Patricia Becker, APB Associates/Southeast Michigan Census Council, Demography (Nonacademic)**

My #1 focus in PUMS is on using the PUMAs to identify special demographic groups within a geographic area of interest. I can live with somewhat less detail and will be considering all those options in my role in advising the bureau. However, I will fight against cutting back on ancestry or occupation detail because to do so really spoils the utility of PUMS in identifying demographic subgroups. Race will be analyzed from here to Sunday in the main publications and I doubt that we need all 63 categories at the PUMA level. Income can be slightly grouped (rounded to 100 or 1000) and possibly topcoded. I see no point to making any changes on age.

**Richard O'Reilly, Los Angeles Times, Media (Nonacademic)**

I have used PUMS extensively during the 1990s as background research for a number of demographic, socio-economic, and political issue news articles my newspaper has published. For instance, it was critical in a series of articles we published on proposed welfare-to-work reforms, showing how the persons most likely impacted would be the working poor.

The data also has provided extensive help in profiling various sub-groups in the diverse culture of Southern California, fleshing out opinions and observations with real numbers.

The Los Angeles Times Poll also has made extensive use of PUMS data in planning public opinion polling of non-traditional population subgroups. The PUMS data provided geographic information and became the basis for weighting the responses.

**Cheryl Elman, The University of Akron, Sociology (Faculty)**

My research pertains to aging and living arrangements and aging and retirement. It is important to my research that census data provide information about who people are (whether young, old, urban, rural, married or not), about what they do in the domains of family and work (census questions pertaining to family formation, economic activity), as well as about the institutional settings of their behaviors (indicators pertaining to occupation and industry at the least). Also important are census questions that ascertain level of education. This type of research identifies the types of older persons who remain economically active, not only by gender, ethnicity, and prior educational background, but also by the kinds of skills non-retirees have (whether they are managers or laborers), the types of families that they live in, and the types of industries they are employed in. With regard to aging and the family, census data allow us examine patterns and trends in the living arrangements of the elderly (or younger persons). I am saddened that the amount of information available to researchers is being threatened. The identification of patterns and trends in work and living arrangements (in the short and long-term) can be very helpful to state and national-level policymakers, for the purposes of planning and the targeting of subgroups for services or incentives. Also, the identification of which types of mature workers remain economically active --or which types of families incorporate older kin- should have great significance for potential employers and/or state-level policy-makers because such information may help us predict the shape of the future labor force or need for long term care. Macro-economic forces over the past two decades, such as firm-level downsizing and job skill changes associated with a globalizing economy, have altered the opportunity structure of work, particularly with respect to the quantity and the quality of jobs available to older workers. These forces may somewhat limit individuals' abilities to improve their employment position. These forces, however, are being counterbalanced by change in the supply of workers. These competing trends INCREASE the need to have variables pertaining to occupation, industry, economic activity, education, and age if we are to learn how to best cope with these economic changes.

**Confidential Response**

Grouping all individuals over age 65 would be disastrous for aging-related research using PUMS data. For aging-related research, most of the interesting questions involve people over age 65 and it is important to be able to differentiate between 90 year-olds and 65 year-olds. With life expectancies in the high 70s-80s now, the proportion of people over age 65 is sufficiently large that identification based on age is extremely unlikely. It is also critical to retain ages for analysis across years -- a lot of research involves considering how 70-year olds have changed over time and grouped age data will eliminate our ability to track such changes.

**Tracy Dietz, University of Central Florida, Sociology (Faculty)**

I am a social gerontologist who focuses primarily upon socioeconomic and racial/ethnic minority status in my research. Although I am interested in and conduct research on a number of areas within gerontology, recently, for example, I have written manuscripts that deal with mortality among older adults by ethnic group status. It is imperative that I be able to

differentiate between people of different ages (ie 65 to 66, even). This is particularly important when we begin to look at age effects by ethnic group membership because such things as the mortality crossover effect is believed to occur at different ages for different ethnic groups (even, for example, between different Hispanic populations). So, given that the census represents our nation's most comprehensive survey of the population, it is very important that as much detail as possible be retained so that reliable, valid estimates and reports can be made. Any limits of detail only provide for less generalizability of any findings that result from research using the data and will certainly limit our ability to continue to make important strides in understanding our nation's population and the issues impacting specific communities and populations. Please, recognize that researchers should be given the opportunity to make important decisions about collapsing categories as it fits with their theoretical and empirical underpinnings because once the data become less detailed, we cannot go back.

**Susan Thistle, Northwestern University, Sociology (Faculty)**

I look at women's changing relationship to household and paid work, and differences in such changes by race/ethnicity and cohort. The proposed reductions in data would end my ability to carry out such research, and would do so for many other scholars. The recent changes in women's lives are of unprecedented and profound importance. To lose data, especially for the census year 2000, which marks the end of one century and beginning of another, as well as the opening of a new millennium, would be a sad comment upon our society. Understanding of the recent and radical changes taking place in the lives of women and their families is essential if we are to address the new social problems as well as potentials such changes are creating.

**Donald J. Bogue, University of Chicago, Demography (Faculty)**

The need to introduce age cohorts into demographic analysis requires that age groups be as refined as possible. The occupational categories used in 1990 were about as minimal as acceptable. Income distributions need detail in order to meet the needs of a variety of studies, but also to adjust income distributions between censuses for comparability. Race and ethnicity are powerful variables in differentiating the population (especially for change over the 1990-2000 decade), and the detail should be as great as for 1990, and as comparable as possible.

**John Blodgett, University of Missouri (OSDA), Sociology (Support)**

I am a programmer who primarily assists other researchers in accessing the PUMS files. So it is difficult to comment on any one specific area of research - I wind up dabbling in many areas. PUMS is a very important tool in letting us provide users with answers to the questions that are not available on the summary files. Frequently, we are asked to generate data for smaller areas such as counties with populations between 100,000 and 250,000 which we can normally do with the current PUMA geography. Raising the size limit to 250,000 would significantly reduce the utility of the files for many of our users.

**Confidential Response**

I specialize both in aging research and in family occupational research. Both types of analyses require detailed information that is only available in the PUMS. Grouping data may minimize issues of confidentiality, but it will preclude most analyses that researchers want to do. For example, with aging research, it is necessary to know ages at least in five year intervals. Anything grouped more broadly than this (especially all people over age 65!) would result in the inability to look at changes in living arrangements over time for the elderly.

**Confidential Response**

Detailed information on ancestry is very important for the research for immigration.

**Confidential Response**

I do a lot of research on racial composition and racial change over time for cities across the United States. I am interested in looking beyond crude categories of white-non-white and focusing on previously ignored groups such as Asians and Hispanics (within the literatures I am familiar with). I also look at intersections of race and class so SES measures are crucial.

**Tim Wadsworth, University of Washington, Sociology (Graduate)**

Much of my research looks at how the economic context of areas affects the employment and criminal behavior of individuals. Without having access to information concerning the economic context I could not do this research.

**Marie Mora, New Mexico State University, Economics (Faculty)**

A lot of my research focuses on the economics of language. I HEAVILY utilize the PUMS, particularly the questions regarding language, English proficiency, education, ethnicity & race, income, age, occupation, geography, immigration status & time since arrival, and gender. My empirical work usually sheds light on policy recommendations.

**Confidential Response**

For our research we need detailed information American ethnic subgroups at the county level.

**Confidential Response**

My work requires extensive occupational data, and losing the ability to construct my own class categories would negatively impact my work.

**Confidential Response**

Losing comparability would make the PUMS data useless for my research area. If the occupational data are not made comparable to the best data now available, any differences in findings concerning the determinants and effects of occupation would be easily attributable to differences in method. The same would be true of race, if not more so.

**Rodolfo Gutierrez, University of Minnesota, History (Graduate)**

My principal investigation is oriented in the analysis of hispanic population, in a comparative perspective. It is indispensable to count with enough information to compare, problem already present in the oldest censuses. We should expect at least that the newest allows a major comparability among other aspects, particularly when the ethnic composition is now one of the most important issues in the political and economic arenas. The knowledge of the process that derives to the actual situation, is indispensable in a deep sense.

**Danielle Longerbone, Indiana University, Geography (Graduate)**

I would not be able to conduct studies containing temporal trends on the depression era population because the level of detail necessary in the age category would be eliminated. This, however, is not the most devastating effect the reduction would have on my research. A greater problem for my research would occur from the reduction of the level of geographic detail. I would not be able to conduct any research regarding local (intrastate) migratory patterns.

**Stephen Snyder, University of Maryland, College Park, Economics (Graduate)**

We know that income and longevity are correlated, but does income really increase longevity or is it just that high income people are more careful with their health? This is an especially important question when Congress is considering changes to social security, and corporations are substantially changing their pension plans. Some very significant changes have been made in benefits that affect very specific population groups. The 1977 Social Security Amendments, for example, give different incomes to individuals with identical work histories, based on their calendar year of birth. I am comparing the mortality rates of those born in late 1977 with those born in early 1978. This would be completely impossible if I could not identify quarter of birth!

**Robert Duran, Charlton Research Company, Industry**

Grouped demographic data would limit analytical capabilities.

**Glenn Steigbigel, University of Chicago, Economics (Graduate)**

Specialize in regulation on specific work. Need current employment categories

**Hayward Derrick Horton, Dept. of Sociology, SUNY-Albany, Sociology (Faculty)**

A reduction in the age, race, and occupation data would be catastrophic to my research. I am currently comparing race and class over time. The availability of a detailed occupation variable is absolutely necessary to have comparability with prior years.

**Jeremy Arkes, Rand Corporation, Economics (Nonacademic)**

My research focuses on the effects of education on labor market and health outcomes. I need to know when they were born, so I know what years they were in high school. The 1990 Census caused problems over the 1980 Census because it didn't include the quarter of birth. To eliminate the actual age would make this type of analysis impossible.

**Susan Wierzbicki, University of Washington, Sociology (Graduate)**

I look at immigrants' settlement patterns within metro areas and already must make tradeoffs in the precision of geography vs. ethnicity. Reduction in data on specific ethnic groups, their employment and their residences would compromise understanding on how, for instance, refugee populations are handling the transition to American life as opposed to groups immigrating for primarily economic reasons.

**Devah Pager, University of Wisconsin, Sociology (Graduate)**

Some of my work has focused on occupational segregation (by race) and within occupational wage inequality. Because there is such a great amount of heterogeneity within broad occupational categories, they are virtually useless for determining the sources of inequality and for disentangling the multiple factors underlying racial inequality in the labor market.

**Abbott Ferriss, Emory University, Sociology (Faculty)**

In constructing indices of the quality of life, results are needed by state. The indicators show trends in the QOL. Trends by demographic characteristics, such as race, age, sex, etc., are the objective. In addition, some studies focus upon counties within a state. County data, consequently, are required. Data on occupation, income, education, etc., are combined to create an index of the QOL.

**Ross Stolzenberg, University of Chicago, Demography (Faculty)**

Any analyses of the effects of schooling on occupation and earnings, or the effects of occupation and earnings on health and well-being, would be severely compromised by reductions in the quality of data on age, occupation, and geographic location.

**Michael Dennis, UC BERKELEY, Economics (Graduate)**

My research details the interactions between socio-economic status and mortality as well as socioeconomic status and fertility. I find the IPUMS to be a good resource for these policy-relevant projects.

I also assign a term project using IPUMS data for undergraduate students. Many of these students are interested in how people of their ethnicity or socio-economic status are doing. It provides a wonderful and personalized way for students to be introduced into social science, statistics and data analysis. Erosion of the ability to identify people like themselves would probably decrease their interest in the project.

Regarding occupation, something which would enable one to distinguish industry AND occupation (so we can tell the autoworkers at Ford from the secretaries for Ford from the secretaries for Microsoft) seems to be the important thing--more important than a large level of detail in occupation per se.

**Naomi Hsu, UCLA, Sociology (Undergraduate)**

I do research on ethnic identification, that is, on how people choose their ethnicity. If there is a reduction of PUMS race/ancestry detail, I would no longer be able to distinguish between, say, those who identify as being 'Taiwanese' and those who identify as being 'Chinese' because both would be generalized under the group 'Chinese'. Consequently, it would be impossible to identify those factors that are associated with identification one way or the other.

**Marco van Leeuwen, NEHA/IISG, Sociology (Faculty)**

To study the degree to which the social structure and the social fluidity of the USA is unique or to what degree it conforms to that of other Western regions, and to trace the causes and historical paths, one absolutely needs detailed occupational data that are as detailed as possible.

**Confidential Response**

A major advantage of having PUMS data is the opportunity to examine detailed age distributions. Grouping the age data (as, for example, into 65+) would render the data worthless.

**Confidential Response**

We are planning a research project with the title: "Migration to North-America, Internal Migration, and Demographic Structures in the Late Period of the Austrian Empire." The head of the project is Josef Ehmer. So for us the information on geographic details will be very helpful. I only used the ipums-data once for a short overview. So at the moment I can not say more.

**Jan Mutchler, University of Massachusetts Boston, Demography (Faculty)**

I am a specialist in aging and minorities. Research in these areas has clearly demonstrated the importance of detailed comparisons by age, and the diversity of detailed racial and ethnic groups. Losing these aspects of PUMS detail would be a major step backward in our understanding of differences throughout the later life course. The assumptions that are often made through grouping decisions (e.g., that 65-year-olds are more like 69-year-olds than they are like 64-year-olds) are often refuted by research using detailed age classifications--loss of the detail means we cannot assess these potential differences. As well, loss of the geographic detail would mean that meaningful contextual analyses would be almost impossible, at a time when the importance and techniques for these analyses are finally being established.

**Amy Guptill, Cornell University, Sociology (Graduate)**

For community development research, detailed local information is very important for understanding which kinds of development strategies succeed or fail in different places, and who benefits and loses from those changes. Without that individual level data on age, race, ancestry, income, and occupation, it is impossible to know how different social groups are impacted by development policies. The PUMS is the single best source of comprehensive high-quality data for this research.

**Mark Melder, Louisiana State University, Sociology (Graduate)**

As most of my work involves LMA data and the attending industrial categories as well as racial disparity in income, the reduction of any detail for these categories would be very detrimental to my work and the work of my colleagues.

**Martha Hill, ISR University of Michigan, Economics (Researcher)**

As for possible cutbacks in detail, there are important international projects that would be seriously harmed. The Population Activities Unit of the United Nations Economic Commission for Europe is launching a project to facilitate international comparisons of quality of life for the different generations (children/prime age/elderly but with concern as well about age differentiation within those groups and at the seams where they overlap) and genders. Many European countries tend to collect even more complete environmental/contextual data in their censuses than does the US, and this currently restrains international comparisons involving the US. If further cutbacks are made in the US census data, the possible research will be limited even more. International comparisons offer vast potential for understanding the role of cultural differences in achieving quality life, and having internationally comparable data gives a handle on the state of the world within which the US must operate. As for timing, if adequate detail is not designed into the data, then no matter when it is publicly available it will fall short of what is needed. Better to design in the needed detail even if it delays the release of the data.

**William Hart, Saint Louis University, Other Academic (Faculty)**

The fastest growing segment of our population is in the aged. Percentage increases increase as you define older and older categories. Lumping together everyone into a category of 65+ would significantly inhibit important research into the problems faced by the older groups. For example, I am researching the effects of retirement on demand elasticities in broad categories of expenditure. Given the wide range of age of retirement, lumping into a category of 65+ would significantly inhibit my own research if not make some comparisons impossible.

**Dirk Early, Southwestern University, Economics (Faculty)**

Research on housing issues requires detailed information regarding characteristics of tenants. Any loss of detail will reduce my confidence in the results.

**Amy Pienta, Wayne State University, Sociology (Faculty)**

PUMS data have been used recently to compare active life expectancy across major racial/ethnicity groups in the U.S -- a key question in aging research. Important questions about population health and aging can be addressed with PUMS data only if detailed information is available for all ages after 65. If age data were grouped after age 65, this work would be measurably less precise.

**Confidential Response**

In defining target populations for Federal, State and Private Grants.

**Confidential Response**

Changes in income and occupational detail would affect our ability to analyze many trends, especially regional and racial variations in occupational hiring and income.

**William Dow, UNC-Chapel Hill, Economics (Faculty)**

Example: My research on economic determinants of mortality at different ages would be severely compromised if I did not have detailed population denominators for age groups above 65. Losing detailed geographic identifiers would also make it impossible to study how mortality rates are related to local economic and environmental conditions.

**Mary McCall, St. Mary's College of California, Other Academic (Faculty)**

Clearly, with grouped data of "over 65" there would be no way to do detailed analyses of young-old, old, and old-old, not to mention the growing group of "oldest-old" over 100 years of age. The same would apply for racial/ethnic groupings, especially with the influx of various Asian older populations.

**Kathryn Hornberger, HealthFirst Management Group, Industry (Nonacademic)**

I am a healthcare consultant. Because my clients range from those needing help with policy research (towards legislative lobbying) to those needing help with service area development, my use of population data is crucial. The more specific the data, the more accurate my findings, often down to zipcode and census tract. When working with State agencies, for example, I have needed clear delineation between the elderly and frail elderly in regard to assessing program development, i.e. home health, rural clinic development. With the rapidly changing demographics among age 50+, it seems especially crucial in coming years to have the data segmented with detailed sets. This will become even more important as we study alternatives and innovative programs in longterm care, assisted living as well as subacute care. The delivery of healthcare is becoming more complicated, not less so - and the data needs to become more specific, not more generalized. It is rather

incredible with our level of technical sophistication to consider that data categories may be consolidated rather than expanded.

**Charles Rambeck, St. John's University, Economics (Faculty)**

I do research on migration. Geographical detail is obviously important. Age and, to a lesser extent, occupation are important variables in my analysis. Grouped age data would force me to use statistical techniques (dummy variables on age categories) that give less satisfactory results.

**Wilbert Van der Klaauw, UNC-Chapel Hill, Economics (Faculty)**

Many government programs (social security, Medicare, SSDI) have eligibility criteria that depend on a person's actual age (ie 62,65 etc). By grouping ages in broader intervals very important information is lost that is essential for evaluating the impact of such programs on behavior (retirement etc).

**R.S. Oropesa, Penn State University, Sociology (Faculty)**

Although it is hard to comment without information on specific proposals, efforts to collapse existing classifications in several areas would seriously compromise future efforts to document patterns of ethnic/racial incorporation and the processes that underlie the incorporation of specific ethnic/racial groups into American society. The geographic detail is also important because it is a common practice link responses of individuals in surveys to information from the PUMS about the areas in which they live to understand how socioeconomic contexts influence behaviors and attitudes. The loss of geographic detail would seriously hamper the ability to understand how individuals are affected by the contexts within which they live.

**Janet Wilmoth, Purdue University, Sociology (Faculty)**

I have used the PUMS data to study older immigrants. A strength of the 1990 data is that the race, ethnicity, ancestry, and place of birth items provide sufficient detail to construct a variable that allows within group comparisons. Reducing the detail of the race and ancestry items would severely limit the usefulness of the 2000 PUMS. This would be unfortunate since the PUMS is one of the few data sources that contains enough cases to make extensive within group comparisons.

Another strength of the PUMS data is the detailed information that is provided for age. I have used this item extensively in my research. For example, in my research on older immigrants I have used the age item to construct a measure of age at immigration. It would be impossible to construct this variable if the age item were grouped. Measuring age in years is critical to researchers like myself who are interested in the older population. Grouping the age categories would result a loss of information (and variance) that is essential to our research. It would limit our understanding of the rapidly growing older population.

**Alicia Adsera, University of Illinois at Chicago, Economics (Faculty)**

Any reduction in geographic detail would affect my research in housing markets and spatial economics. Any reduction in occupation detail would affect my ability to understand differences in productivity of similar workers across geographic areas.

**Carl Backman, Buffalo State College, Sociology (Faculty)**

I have been involved in evaluating jury selection procedures. It is essential to have as much geographic detail as possible. It is important to have age detail that allows identifying those old enough to vote (18+), those old enough to drive (16+), and the old old (say 80+) who might reasonably be excluded from jury duty. Migration is a critical issue for evaluating lists maintained by the courts, so migration detail (already weak, given a 5 year delay) needs to be maintained (particularly same house, same county). Poverty status is also valuable.

**Shiro Horiuchi, Rockefeller University, Demography (Faculty)**

In order to understanding demographic, social, health, and other aspects of aging processes and in order to serve demands of the elderly population better, we need to investigate AGE-RELATED CHANGES of various characteristics in the elderly population. People 65 years old and those 95 years old have completely different profiles, characteristics, problems, and needs. To understand age-related changes of behaviors and characteristics, the AGE IN SINGLE YEAR [65, 66, 67 etc.] is essential.

**John McHenry, Demographic Data For Decision-Making, Inc., Demography (Nonacademic)**

My firm's local government and business clients require the best possible local area (i.e. sub-county) demographic data in order to make the best possible local area business and policy decisions. My firm is embarking on an SBIR-sponsored line of research to produce local area demographic estimates and, later, local area demographic projections. Not surprisingly, our preliminary research indicates that lower-level geo-demographic inputs produce higher quality demographic outputs. To restrict PUMS and other census outputs to higher levels of geography and/or higher levels of demographic aggregation will

cripple my research and like research efforts. Regional health-care providers, banks, local businesses, school districts, utilities, local governments and all their direct and indirect clients will suffer if PUMS databases are no longer readily available.

**Shelley Lapkoff, Lapkoff & Gobalet Demographic Research, Inc., Demography (Nonacademic)**

I am a consultant to many school districts. My company provides school enrollment forecasts for effective planning. Perhaps surprising, understanding the senior population is very important to enrollment forecasts in mature communities, because housing turnover is an important variable predicting migration and increases in the school-aged population. Also, which housing turns over is important - seniors who move out of their homes create the biggest impact on school enrollments. We use PUMS to study relationships between age and housing turnover generally, which we can then apply to specific school districts. Please do not eliminate ANY age detail from the PUMS data - this is critical (65 year olds act a lot differently than 85 year olds!) Also, we hope no change will occur on providing characteristics of the housing.

**Joseph Persky, University of Illinois at Chicago, Economics (Faculty)**

I first used PUMS data in 1967 as a graduate student and have been using them ever since. Most critical to my current work are the puma and powpuma income data within metropolitan areas. These data allow us to research the distributional implications of different economic development scenarios. The same data are key to analyses of suburban sprawl.

As to historical comparability, I recently supervised a thesis on interregional income convergence since 1950. The IPUMS allowed the PH.D. candidate to carry out detailed comparisons of incomes adjusted for education and occupation over time. I can't imagine another source of data that could have played this role.

**Jeanne L. Anderson, Jeanne Anderson Research, Industry (Nonacademic)**

Difficult to specify. See suggestion below for issuing several versions of the PUMS file.

**William Yancey, Temple University, Sociology (Faculty)**

It appears that you may be planning on combining occupational and industrial categories. My interest in the consequences of industrial transformation could not be conducted if this were done.

Given variation in industrial and occupational profiles between urban areas it is also critical that detailed geography is maintained.

**Rachel Boaz, Graduate School, City University of New York, Economics (Faculty)**

Aging research: Do NOT compress age-classes; the aged population is a very heterogeneous with substantial differences by health status, income, marital status, living arrangements, and race. Income: Narrow income classes are needed for meaningful analysis. The open-ended class should be at the 99th percentile.

**Debra Erdley, Tribune-Review, Media (Nonacademic)**

As a newsreporter covering issues in one of the nation's oldest communities, any reduction in specificity with regard to age would hurt me tremendously in trying to make comparisons over the years. When you live in an aging community that has faced economic hardships, the boosters always try to paint a picture that is slightly rosier than reality. Good solid data is a tremendous reality check. It also helps community leaders make informed decisions.

**Melissa Binder Binder, University of New Mexico, Economics (Faculty)**

My recent research involves the effect of local labor market conditions on social and economic outcomes such as fertility, school enrollment and poverty. The identification of localities over time, as well as detailed individual information regarding race, age, income and occupation is vital to understanding the effect of labor markets on particular population groups. I believe that this type of research is crucial in making informed policy decisions.

**Jeff Burr, UMASS Boston, Demography (Faculty)**

The kinds of grouping being contemplated would make historical (recent) research nearly impossible. My aging research focusing on race groups would be finished as far as national data are concerned. AS you know, the census PUMS offer the only source of individual data for research on certain kinds of race/ethnic groups. It is yet another great irony that the federal government and politicians are trying to fix something that is not broken. Where are the examples of any person being specifically identified with these data?

**Steve Pischke, MIT, Economics (Faculty)**

Since one of the main advantages of the PUMS is it's large sample size, it allows analysis of very detailed subgroups, which would be compromised by aggregation. A specific example in economics is a very influential paper on the returns to schooling by Angrist and Krueger using the QUARTER OF BIRTH variable available until 1980. Even if the proposed

aggregation is not particularly harmful to a specific project now, it might always compromise the ability to implement good research ideas in the future.

**Mary Grace Kovar, NORC, Demography (Researcher)**

People age 65 and older exhibit more diversity than any other age, and are the subject of more national policy (except in times of war) than any other age. While the population pyramid narrows considerably, it is a 40-year age span and is much like having a 25-64 year age group, which no one considers.

Death rates are much higher at older ages. Higher for men than women - same as earlier ages Higher for white men than black men - reversal of earlier ages Hospitalization rates are much higher. Medical visit rates are not. Nursing home residence rates are much higher. There is a radical change in living arrangements and in marital status - women are widowed and live alone.

**Confidential Response**

Age, income, and race/ethnicity are crucial factors in the work that I do on low-income communities and populations. For example, how many 3 and 4 year-olds living in poverty (or near) poverty are enrolled in preschool programs. An aggregated less than 5 category would not provide this type of detail. Similarly, the prevalence of disabilities and the need for supportive services vary greatly by age for older persons. An aggregated 65+ category would not allow this crucial analysis. Income levels and race/ethnic factors are similarly needed to target programs and services to persons who are most in need.

**Confidential Response**

PUMS data have been used by my doctoral students in their dissertation research, and have been a major source of insights into demographic characteristics of people in the PUMS regions that have informed the development of primary survey instruments. Any reduction in detail would have major negative impacts on basic research of this type.

**Myron Floyd, Texas A&M University, Other Academic (Faculty)**

My specialty area is race and ethnic studies. My research examines how race/ethnic impacts natural resource policy and management. Reduction of PUMS limits development of my GIS databases.

**Peter Philips, University of Utah, Economics**

I research the construction industry. Occupational and industry detail are essential to understand the dynamics of this industry. The income, employment and mobility patterns vary considerably comparing electricians to carpenters to roofers to laborers. These are very distinct labor and commercial markets. They cannot be condensed meaningfully.

Race and age are key variables in predicting outcomes. Construction workers in outside work, operating engineers, laborers wear out much faster than inside workers, plumbers electricians (as opposed to linemen). Restricting age information would hide these important differences.

There are key policy questions that will not be well informed if these information are suppressed. examples:

Should retirement age for social security vary by occupation? Should prevailing wage regulations for construction continue? (The Davis Bacon Act and 31 state regulation plus many municipal regulations of construction. The census plays a key role in the debate regarding these laws.

I also use the census pums as the key data set for forensic research on the effects in cases involving wrongful termination, wrongful death and personal injury cases. No other data set provides the occupational detail at a state level that allows for predicting the effects of these events on income. The justice system relies on these data for fair determination of damages in thousands of cases each year.

**Jeff Tayman, San Diego Association of Governments, Government (Nonacademic)**

Our agency uses PUMs for a number of activities, including the preparation of detailed demographic and economic estimates that support multi-billion dollar capital facility projects, the development of local housing elements, plans and grant applications, and for infrastructure and health care planning and analysis.

**Ronald Lee, University of California at Berkeley, Demography (Faculty)**

I work on aging, and the loss of age detail above age 65 would be devastating; it would mean that PUMS would no longer be of use for research in this area. Loss of information on source of income, eg public assistance or Social Security, labor earnings, etc would also be very damaging.

**Charles Rynerson, San Diego Association of Governments, Government (Nonacademic)**

We do estimates and forecasts of population (by age, sex, ethnicity, and military/civilian status), labor force participation (by age, sex, ethnicity, and military/civilian status), employment by occupation and industry, and household income. We use PUMS to develop basic assumptions and inputs for all of our models. The race and single year age detail is particularly important, because we need to crosstabulate race and age by military status, labor force, and migration history (both the 5 year mover and the born abroad questions). Migration history is important to us at all age levels. Maintaining the occupational categories is important, because it would allow flexibility to group them to be compatible with occupation groups from BLS' Occupational Employment Survey. The geographic detail is less important, because most of our use for PUMS is at the county (almost 3 million people) level, and we are crosstabulating too many variables to get significant results from a sample of an area with 100,000 people.

**Virginia Appl, Houston-Galveston Area Council, Government (Nonacademic)**

I use PUMS data to determine "synthesized households" for use in Urban Land Use Modeling ... less detailed means broader assumptions regarding population characteristics which is not good when you are trying to produce estimates at the block level!

**Mark Hill, Pennsylvania State University, Demography (Faculty)**

Grouped age data would severely limit the usefulness of the PUMS data. I have developed a survivorship method ideally suited for the IPUMS data that requires the same cohort members be identified in various PUMS files. With out age in single years, this method could not be used appropriately. Furthermore, with the aging of the population the key demographic issue in the US, grouping data at the oldest ages poses a key threat to research on aging.

**Haydar Kurban, University of Illinois at Chicago, Economics (Researcher)**

It will be almost impossible to have a representative sample in different income, age, race and geographic classification.

**Jenifer Bratter, University of Texas at Austin, Sociology (Graduate)**

I research family dynamics, particularly marriage and cohabitation. Currently I am examining geographic variability of interracial marriage for 1990. Detailed race and ancestry data is crucial to identify the ethnicity of whose marrying whom, when the broad categories are too general. I also greatly appreciate the geographic detail the census provides. Currently, I can match the detail applied in other studies which increases the comparability of my results with other work.

**Bill O'Hare, Casey Foundation, Demography (Nonacademic)**

Loss of geographic detail would be very difficult for me. The only reason the PUMS are more useful than CPS is because PUMS provides data for smaller geographic areas. If that strength of the PUMS is lost, it becomes relatively useless.

**Jennifer Garner, BBC Research & Consulting, Industry**

PUMS data is a critical input to consolidated housing plans and other assessments of regional demographics, particularly on the PUMA scale. We often analyze PUMAs to segment demographic markets within a small geographic region in order to assist public sector clients in prioritizing spending.

**Jeanne Gobalet, Lapkoff & Gobalet Demographic Research, Inc., Demography (Nonacademic)**

We examine small geographical areas and reduction of geographic detail would greatly reduce our ability to provide the answers our clients need. The more detailed the information, the better, as far as meeting needs of our clients, who include public schools, public health agencies, and those assessing child care availability. We have also used PUMS data to assess whether housing discrimination exists.

**Yoshinori Kamo, Louisiana State University, Sociology (Faculty)**

Chinese/Japanese elderly's living arrangement (JMF 1996): Age cut off at 65 made it virtually impossible to do research like that. Married women's labor force participation affected by the husband's income, own education, etc. 1950 data had to be dropped since education was not available. 1960 data categorized income (rather than dollar value), which makes the husband's income measure quite primitive. I don't want to see it again for 2000 data, since income is critical in sociology.

**Mark Mather, Population Reference Bureau, Demography (Nonacademic)**

I have often used the 1990 PUMS to create customized tabulations for states, that can then be combined with national-level data from the Current Population Survey to arrive at post-census state estimates for various sub-groups of the population. Combining age groups or race/ethnicity data would be harmful if these variables could no longer be used in combination with other federal data sources, like the CPS.

For me, the main advantage of PUMS is the availability of data for customized sub-groups of the population -- for example, data on persons in a specific age group that aren't published in summary tables. If these data are grouped, then the PUMS data become less useful.

**Tim Smeeding, Maxwell School, Syracuse University, Economics (Faculty)**

PUMS is a once a decade occurrence; there is considerable interest in detailed age and location material on income, family status, age etc. During the 1990's the interest was at the level of the school district; during the next decade the interest will be in aging populations. Please do not skimp on the PUMS

**Julie Hudson, The Urban Institute, Economics**

I study behavioral responses to welfare and other policies at the national, state and local levels. It is very important for me to have detailed information on a person's geographical location of residence. In some cases, state level is enough, but in others it is not. I also must break down my groups of interest by age, race, sex and income levels. Income is also an important outcome variable in some of my studies.

**James Holland Jones, Harvard University, Demography (Post-doc)**

What can I say? Making projections about mortality and fertility changes with respect to income and geographic mobility simply can not be done without data on age structure, income (and, to a lesser extent, occupation) and geographical location.

**Sharon Ju, Houston-Galveston Area Council, Economics (Post-doc)**

Employment details

**Harry Holzer, Urban Institute/Georgetown University, Economics (Faculty)**

I and my coauthors use race/ethnicity data and geographic information, especially at the census tract level from the STF3 files.

**Roy Treadway, Illinois State University, Sociology (Faculty)**

For my research detailed geographical units are more important than detailed economic and social characteristics since we more often are involved with general profiles for smaller (100,000 population or so) areas.

**Charles Purvis, Metropolitan Transportation Commission, Government (Nonacademic)**

1. PUMS is needed in preparing county-level data on socio-economic characteristics by market segment, for use in travel demand forecasting models in use by regional and county-level government agencies. For example, PUMS is used to calculate mean household income, mean household size, and mean workers per households for households stratified by household income quartile by auto ownership level (0,1,2+ vehicles per household). Only PUMS provides this critical information.

2. PUMS is used in analyzing characteristics of commuters by means of transportation, including age, sex, occupation, earnings, race/ethnicity, ancestry, year of immigration to US, and industry.

**John Straub, University of Wisconsin-Madison, Economics (Graduate)**

The lack of geographic detail in the 1990 PUMS data poses many challenges to researchers wishing to use the data. Mapping between PUMAs and other geographic entities can be quite complicated. This is a reasonable price to pay to ensure confidentiality, however. I'm not sure how hard it is to ensure confidentiality with PUMA-level geocoding, but I do know that more coarsely geocoded data would be useless for many purposes.

**Julie Lee, National Bureau of Economic Research, Economics (Researcher)**

I do research in health economics. Very often health-related data sources lack social economic variables, such as income, education, occupation, etc. which influence choices people make regarding health. PUMS is a good source of such variables. Age, race and geographical details are especially important because they allow me to merge social economic information with individual level health-related data; finer the age-race-sex-geographical cells are, more accurate the merging would be.

**Jim Hinterlong, Washington University, Other Academic (Graduate)**

I use 5 year age-cohorts to examine patterns of engagement in paid employment, while limiting my sample to those aged 60 and over. Aggregation beyond the current level (which is "lumpy" enough) would preclude me from using this data for future research.

**Confidential Response**

We use seed matrices derived from PUMS to establish joint distributions of householder age, household size and income, workers, and vehicles. We would like to maintain as much specificity as possible within these categories.

**Simon Choi, Southern California Association of Governments, Government (Nonacademic)**

My research generally might not need a detailed categorization of some items (see below). However, the more information we have, the better we can analyze. This will help us to devise a better planning or public policy.

1. Cohort component model run: I use age and race/ethnic specific birth rate and migration rate from PUMS to be applied to cohort component model. Generally 18 age groups (0-4, 5-9,...) and 4 different race/ethnic groups (Non-hispanic white, non-hispanic black, non-hispanic asian & others, hispanic)
2. Household projection model run: I use age and race/ethnic specific headship rate from PUMS to do household projection.
3. Travel behavior of commuters by race/ethnic group and income level.
4. Household overcrowding by age group, race/ethnic group, income, and occupation.

**Lisa Catanzarite, U.C. San Diego, Sociology (Faculty)**

My research field is stratification, with a primary focus on labor market inequality by race/ethnicity, immigration status and gender. Occupational segregation is central to my research. Any reduction of detailed occupational categories would be disastrous to my work, as well as that of others in my field. Information on ancestry, race, and place of birth is critical to research pertaining to immigration, race, and ethnicity. Both my work and that of others would be adversely affected by loss of detail on these variables. Much of the research on immigration has been conducted at the level of metropolitan areas or PUMAs. Such detail is crucial for work in this field. I also conduct research on other aspects of stratification, including poverty. Both earnings and income data are essential to this work.

**Francine Atwell, Hawaii SBDC Network Business Research Library (under the Research Corporation of the University of Hawawii), Economics (Researcher)**

I work with small business people. They often want to know the number of likely customers for their products/services in their neighborhood, and want to get what often is a cross-tab of 2 or more characteristics. I can't get that from the usual STFs.

**Deb Greene, City of Berkeley Health Dept., Other Academic (Researcher)**

My major focus is racial disparities in health outcomes. I am currently working on broader historical models that take into account income, occupation and am analyzing trends over time. Geographically, I am making city level (Pop >100,000) models. Any reduction in geographic detail or racial/ethnic detail would prevent much of my research.

**Susana Adamo, University of Texas at Austin - Population Research Center, Demography (Graduate)**

My current works deals mostly with non-metropolitan populations. The reduction of geographical detail (I mean, less detail than in 1990 census) may, eventually, reduce the possibility of meaningful analysis other than at national level. In addition, grouped age-data (for example, in 5-years groups) limits the analysis of life transitions where a difference of one or two years may be potentially important, like age at marriage, or entering and exiting from the labor force.

**Karen Lamphere, San Diego Association of Governments, Government (Nonacademic)**

One of the most valuable features of PUMS, obviously, is that users can determine their own groupings of categories. Although the files are large and cumbersome, I would encourage the Bureau to limit the grouping of variables as much as possible.

**Shawn McBrien, University of Michigan, School of Public Health, Other Academic (Graduate)**

Reduction in detail of data would significantly hamper accurate research in the area of social epidemiology. Accurate research requires the greatest accuracy at the smallest scale possible.

**Confidential Response**

The new format of the census race question, which will allow multiple responses, will involve the potential reclassification of millions of Americans. In order to separate the artifactual changes introduced by question format from real changes in the socio-economic status of various minority groups, it will be extremely important to have as much detail as possible on which people are marking more than one race.

**Edward Miguel, Harvard University, Economics (Graduate)**

I am interested in intergenerational mobility, especially among ethnic and national minority groups. Reduction of PUMS in future censuses would make this research extremely difficult, if there were less detailed data on geographic mobility, income and occupational outcomes, and national origin.

**Tara Gibbs, University of Minnesota, Other Academic (Graduate)**

My field is applied linguistics

I would like to see MORE detail, not less. 2000 job titles would be better.

I need summaries of data available by neighborhood and small town and rural area, but this isn't currently available. I'd like to be able to search by primary language and get maps of any areas with 5 speakers or 1000 speakers, etc.

I would like a data set that is broad released soon and a more detailed set released later.

Historical compatibility would be essential, if historically they had actually asked the questions that would have provided the information I need to know. It would be nice if certain questions were asked and then kept the same. Specifically, I need to know first and second language information and whether someone is a first or second generation immigrant or later. I need to know if the person speaks no English, very little English, some English, conversationally fluent or is perfectly fluent in English. And I want to know this about the other languages that this person speaks. This is only surmisable from the national origin data using statistical procedures over time, but it is grossly inaccurate to use this, however, there is nothing else available, so this is what we do in this field.

Thus our estimates on what level of interpreter, L.E.P. and other services needed, in what languages, and where, are grossly inaccurate, as is our ability to measure levels of language maintenance, language shift, and bilingualism which would highlight the effects and conditions of various policies, as well as whether a given language group continues to need services.

**Alexander Bucur, USC/UCLA, Sociology (Researcher)**

Changes in the American society cannot be measured or extrapolated from other surveys unless historical compatibility is achieved. Since we have such good data for 1980 and 1990, and we have experienced such change in our society it seems that we must maintain compatibility to be able to analyze changes. Even more, all other surveys that compare or balance themselves against the US Census will be undermined since they will not be able to properly analyze changes at the societal level.

**Cecilia Conrad, Pomona College, Economics (Faculty)**

I am investigating historical trends in the economic status of teen mothers by race. It is critical to distinguish between young teens and older teens. Each year of age can matter a great deal.

**Kenneth Sylvester, University of Victoria, History (Post-doc)**

My research takes me into the exploration of degrees of self-employment in the farm economy, and the shape of regional non-farm economies.

I need to have fairly detailed breakdowns to explore inter-regional and international comparisons.

**Confidential Response**

I'm most concerned with reduction in geographic level. If it was only for areas 250,000 or greater, then we would be hard pressed to look at cross tabulations for our region.

**Marie Cornwall, Brigham Young University, Sociology (Faculty)**

The pums data have been most useful in my research on changing gender roles using the 1940 to 1990 data sets. Most data available used 1960 and beyond, but the pums data allowed us to examine the 40s and 50s and see a more consistent trend line.

**Karen Norberg, National Bureau of Economic Research, Other Academic (Faculty)**

I specialize in epidemiological & economic research concerning children and youth, especially pertaining to high risk behavior, household resource allocation, maternal employment, and factors affecting fertility and teen births. I make use of the finest level of detail about age groups & geographic variation possible, & these are the highest priority areas, for me, in preserving geographic detail. I compare local area estimates from the PUMS with US vital statistics data, & much of the value of the PUMS would be eroded, for my work, if children's ages or geography were further collapsed.

**Confidential Response**

I study the changing nature of childhood in the United States. A reduction in detail would not allow me to bring my studies into the present, and would reduce historical comparability.

**Dennis Yee, Metro (Portland, OR), Government (Nonacademic)**

I and others use the PUMS data as the initial seed matrix for creating current and future year estimates of populations for transportation and land use modeling and forecasting. Household, income and age data are extremely important for us as these variables represent key components into our simulation models. In addition, geographic detail is very important to us. Reduction in population size of PUMA's could damage our modeling and forecasting precision.

**Confidential Response**

I am a labor economist, so I am very interested in income and occupation data. However, given differences in income, etc. between metropolitan areas, even small ones, I think having PUMA information is important. Personally, I'd like to have county data (or county groups for very rural counties).

**Richard West, Social Policy Research Associates, Economics (Nonacademic)**

We do a lot of analysis for the Department of Labor in which we calculate demographic and economic characteristics of local workforce investment areas. Geographical detail at the county level is highly desirable and indeed necessary for this analysis.

**William Sundstrom, Santa Clara University, Economics (Faculty)**

I have recently been examining historical changes in occupations by race. One dimension of this is to examine how these changes varied according the characteristics of local markets. Thus I combine detailed information on occupation and location of residence. A reduction in detail would reduce the available variation for analysis considerably. For example, knowing the city is much more informative than merely knowing the state.

**Jon Stiles, UC Berkeley, Sociology (Researcher)**

One of the many strengths offered by the time scale and sample sizes available in the IPUMS is the ability to construct birth cohorts which can be sliced several ways at the analyst's discretion over time. This virtually requires single years of age to accommodate different analyst's intentions.

**Laura Leete, Willamette University, Economics (Researcher)**

I work on the issue of how capable local labor markets are of absorbing welfare recipients, among other things. Thus, having maximum detail on occupational classifications by sub-metro area is crucial in being able to characterize the nature of local labor markets. More generally, reducing the precision of any of the variables mentioned here reduces the precision with which we are able to characterize and understand poor, welfare-dependent and at-risk populations.

**Rose Nguyen, University of Minnesota - Census Projects, Economics (Support)**

If I specialize in occupation research, then the market analysis for the unemployment rate and employment rate is not statistically clear for any occupations.

**Confidential Response**

PUMS has been one of our sole means of gaining greater understanding of the racial, ethnic, and linguistic diversity in our area. Some of the collapsing of categories already has presented problems in using these data for health assessment. For example, tribal affiliation for American Indians/Alaska Natives was collapsed to the largest tribes nationally. This made information on local tribe affiliation non-existent for our health assessment purposes. We would like greater detail in areas such as this.

**Confidential Response**

I explore theoretical issues of representing data uncertainty and data fusion. Reducing detail and/or increasing the size of PUMS regions will significantly reduce the ability for researchers to model urban areas.

**Jong Won Min, UCLA, Other Academic (Graduate)**

With a rapid increase in old-old population (85+), updated information on detailed age group is crucial to understanding their health care and long-term care needs.

**Robert O'Connor, Penn State University, Government (Faculty)**

My research has used PUMS data to examine the ability of government programs to involve different racial, ethnic, and income groupings. Grouping data beyond that done for the 1990s PUMS would weaken this research.

**Maggi Rose, Open University, UK, Economics (Undergraduate)**

I examined the effect of home computer ownership on occupation and income. I required detailed economic, geographic and occupational data to complete this. To remove other influences I also required sex, ancestry and race so that I was sure I was comparing like individuals whose only difference was their computer ownership. To have been unable to identify the similarities and differences between individuals would have rendered the study useless.

**Confidential Response**

My area of research is primarily a comparison of aging and working-aged populations over a number of topics. It is crucial that lumping of persons aged 65 and over does not occur.

**Confidential Response**

The population estimates and projections produced by the PA State Data Center rely on the single year age data. The model ages the population factors in birth and death data from our Department of Health and applies migration rates. A different model would need to be used if only grouped age data was produced. In addition, the PA State Data Center has studied the income of elderly citizens (detailed income categories for those age 60 and older, both unmarried persons and married couples with age detail for both partners) for the PA Department of Aging. The reduction in the geographic, income or age detail would effect this study as well.

**Dawn Hubbs, University of North Carolina at Charlotte, Other Academic (Researcher)**

I work in a library and therefore help other researchers. Many faculty at our institution do research on family relationships, and need specific age and race data. We also have several faculty doing research on the aging population, and need to compare 65 yr olds to 70 yr olds to 75 yr olds, etc.

**Francine Stephenson, SDC, NC Office of State Planning, Government (Nonacademic)**

Government policy makers need information on VERY SPECIFIC age groups in local areas cross-tabulated by a variety of characteristics. Not having the age detail means that policy decisions would be based on guess.

**David Wong, George Mason University, Geography (Faculty)**

PUMAs are already quite coarse in geographical scale. Many phenomena have significantly variation with a county and PUMAs are county or groups of counties, which fail to reflect the spatial differentiation within county. Grouping PUMAs together will be much worse. The results will be of no significant geographical implication and useful for policy formulation related to differences within regions.

**Michael Stover, Hillsborough County Planning Commission, Government (Nonacademic)**

No comment.

**Confidential Response**

The PUMS allows our office to virtually do any analysis that is not available from the standard products. This file has been essential for researching the impact of policy and proposed legislative bills. It's use has escalated in the last couple years (NOT reduced) -- even as the data has aged.

**Michael Spar, Weldon Cooper Center, University of Virginia, Demography (Researcher)**

I doubt that changes in the PUM will affect me directly. However, as a member of the state data center network for Virginia, we receive a high volume of information requests. Some of these can only be answered by reference to PUM data. If those data are compromised in any way - either by changing the minimum population size for designating a PUM area, or by reducing the level of data detail, then fewer data users will find the answers to their questions in the PUMs data. For some data users, the ability to "create" their own tables is the one and only saving grace of PUMS. Reductions in the number or quality of those home-grown tables will simply reduce the usefulness of the Census to some data users.

**Leonard Gaines, Empire State Development (NYS SDC LEAD), Government (Nonacademic)**

As a data provider/intermediary we frequently get requests for specialized tabulations (specific age ranges, etc.) for program evaluation purposes. Using different age groupings would bring the entire analysis into question. For example, a user might be looking at a potential program aimed at 16 and 17 year olds not in school. Using any other age group would not be valid. The next user might NEED 17 to 20 year olds. It would be impossible to meet the needs of both users with a single grouped age variable.

**William Even, Miami University, Economics (Faculty)**

It would be impossible to examine how aging affects the level of income among the aged, as well as how widowhood at different ages affects income.

**Duane Marble, The Ohio State University, Geography (Faculty)**

The current level of geographic aggregation is already far too high to permit valid inferences to be made about many aspects of behavior. E.g., changes in the journey to work.

**Lisa Nelson, Case Western Reserve University, Sociology (Support)**

In general, I use the PUMS data to assist community organizations and some governmental agencies with their work -- whether it be proposal writing, program planning or needs assessment. The detail of PUMS is necessary to meet some of their specific questions and needs. Communities need this type of information in order to understand and work toward improving the social and economic conditions in their areas. While the STF3 and STF1 may meet many of the information needs -- it does not meet all of them.

**Peter Granda, University of Michigan - ICPSR, History (Support)**

While I am not an active researcher and user of PUMS data, as an archivist I interact with numerous researchers who have used the 1980 and 1990 PUMS files. My strong impression is that these individuals want as much if not more detail in the 2000 PUMS files as appeared in either the 1980 or 1990 files. ICPSR expects a very heavy demand for the 2000 PUMS files.

**Richard Rathge, North Dakota State University, Demography (Faculty)**

Often we are asked by congressional staff to assist in understanding the consequences of changes in legislation. The last several times dealt with legislation regarding income. We had to use detailed income breakdowns to offer insight into the consequences of different thresholds by detailed age. Thus age and income are necessary in detailed form.

**Susan Alibozek, Montgomery County Dept. of Planning & Development, Government**

Our county has a high elderly population (24% over the age of 60 in the 1990 census), and as a result many many programs affecting the elderly rely on age population. Additionally, we have a high hispanic population which also relies on data for that specific category therefore reduction of information on the age and race category would harm the data requested by many organizations who contact our office for information.

**Ronald Cossman, University of Central Arkansas, Geography (Graduate)**

Both my dissertation and professional research agenda are focused on the relationship between economic structure and wage inequality in the United States over time. This research is critical to both understanding the processes that cause changes in inequality, as well as guiding economic development professionals (from EDA to the city level), in policies and programs that will mitigate, not cause, increased inequality. The PUMS is the only currently available data set to examine the migration of workers between occupational and industrial sectors of our economy.

**Lisa Dillon, Institute of Canadian Studies, University of Ottawa, History (Faculty)**

My research on the aged has demonstrated strong distinctions among older persons on the basis of age; the living arrangements, income and meaning of old age itself varies as older persons grow even older. Grouped age data would prevent me from tracing fine distinctions among the elderly on the basis of age. This will be particularly true in the case of the 2000 Census, which will record the largest ever elderly population--and largest ever population of old old persons--in the history of the United States. I would prefer that the age variable be available in single years so that I can fashion my own age groups as suits my research needs--including researching the declarations and experiences of centenarians. The other changes being considered for the 2000 Census will also have a grave impact on research. For example, a fine level of geographic detail is necessary for health research. With improvements in Geographic Information Systems technology, scholars can now research correlations between disease and environment by mapping at the level of small communities. By erasing fine-grained geographic distinctions, the 2000 Census will considerably hinder our efforts to understand links between disease and place.

**Gastón Repetto, UIC, UNLU, UCC, Economics (Researcher)**

I have used IPUMS in my research regarding immigrant assimilation, ethnicity, how different social groups invest in different forms of human capital, and the rewards they get in the labor market and in home production across time. Therefore, all the information concerning: 1) labor market individual's characteristics and outcomes (income, occupational choice, tenure, OJT, etc.), 2) sociodemographic data (age, ancestry, household composition, race, language spoken at home, etc.) and 3) data specific to immigrants are a must for my line of research.

**Steven Ehrlich, US Dept of Housing & Urban Development, Economics (Nonacademic)**

At the very least, central city status should be identified for large and medium sized metro areas.

**Confidential Response**

I study age adjusted rates of disease using race, geography and place of birth as markers for increased or reduced likelihood of infection. Since groups from certain countries with certain background are much more likely to contract certain infections it is essential to have this information for health program planning.

**Diane Hite, Mississippi State University, Economics (Faculty)**

I work in urban, regional and environmental economics, and analyses on race and the environment, for example, would be very adversely affected if the geography level were to be broadened.

**Michael McDonald, SUNT--Binghamton, Other Academic (Faculty)**

My area is race and politics. Without detailed information on race, by small geographical units, much of what I have been doing for 15-20 years can't be done for 2000. Because my research involves voting--turnout and candidate choice--not having detail information on age (turnout), income and occupation (vote choice controls when looking at race) most inferences would have large clouds of doubt hanging over them.

**Craig Olson, University of Wisconsin-Madison, Economics (Faculty)**

I'm doing research now which focuses on the supply and demand for technology workers and the impact of higher education on the creation of high skilled jobs. Detailed data over time by geographic area and occupation are essential for estimating the relationship between education and high skilled jobs.

**Beverly Allen, Iowa State University, Sociology (Faculty)**

I am using the occ score variable over the 1850-1990 period. This is crucial to my current research. Any changes, other than improving comparability over time will hinder planned research for the next few years.

**Williams Olatubi, Louisiana State University, Economics (Researcher)**

In my work, being able to map earnings by industry to occupational categories is very important. The more the detail, the more plausible the results of my analysis and the better the inferences.

**Confidential Response**

In the library, we provide very specific census data to researchers, whether they are students, faculty, staff, or members of the public. Topics run the gamut--physical and social sciences, humanities, health, fine arts... No longer providing this service would be a great loss.

**William O'Dell, University of Florida, Other Academic (Faculty)**

A reduction in detail and a more highly aggregated geography will substantially impair our research which is in housing policy. Our Center works very closely with state and local governments and non-profit organizations. PUMS data has been crucial to these organizations in policy-making, resource allocation decisions and strategic planning. Recently, for example, the state's housing finance agency was contemplating an expansion of elderly housing subsidies (affecting the allocation of millions of dollars annually) and needed information on relative cost burdens across several age sub-categories within the group - PUMS data was absolutely essential to the development of that information. For a number of organizations that we consult with a loss of geographic detail in PUMS would be a significant impediment to informed resource allocation decisions.

**Roy Williams, Massachusetts Institute for Social & Economic Research/Umass/Amherst (State Data Center), Government (Researcher)**

As a State Data Center we get many requests for tabulations of data for very small geographic areas. PUMAs of 100,000 give us some flexibility to construct geographic areas that approximate users' areas of interest even though a tabulation based on a single such PUMA would rarely have statistical validity.

**Ren Farley, University of Michigan, Demography (Researcher)**

My work involves describing racial differences nationally and at the PUMA level. It is necessary to have detailed information about race, occupation, earnings and industry. I also work on immigration. Immigrants are now clustering in specific occupations within specific industries in specific geographic locations. Only the census can tell us much about the assimilation process for most immigrant streams.

**Kathy Kaufman, Columbia University, Sociology (Graduate)**

It is difficult to project how one's future research may be impaired by the lack of data on particular variables because one cannot anticipate in advance what issues will be pertinent in the future. But once a new problem emerges, we can count on the necessity to go back and examine the historical record in order to place it in proper perspective. Even a slight increase in the detail of data exponentially increases the information's utility, because it vastly expands the flexibility with which the data can be applied to the examination of unforeseen questions.

**George Sweeting, New York City Independent Budget Office, Economics (Nonacademic)**

Our research is largely concerned with changes in tax burdens (both income and property) over time. We do both cross-sectional distributions as well as micro-simulation modeling. Detail on components of income, occupation, age, tenure status, and wealth factors are critical to this work.

**Jim D'Amato, Spartanburg County, SC, Other Nonacademic**

As a transportation planner for a Metropolitan Planning Organization, I need specific population and population derived variables for 700 separate zones for my county. This data is crucial for the proper working of our travel demand modeling program. Other variables (by zone) include numbers of employees; population, labor force, registered vehicles, and occupied dwelling units by zone of residence, school attendance zone, and retail sales.

**Paul Lachance, Department of History, University of Ottawa, History (Faculty)**

Since my research is primarily on the 19th century, the 2000 PUMS is less important for me than for colleagues who specialize in the next century. Nevertheless, any problem needs to be addressed in comparative perspective, whether forwards or backwards in time. It is also important to think ahead to the census as a source for historians of the future. What we can observe from work on U.S. censuses is that the more detailed they have become over time, the more valuable they are. It is especially important, in variables such as occupation, to provide as extensive a breakdown as possible; since how historians will regroup that data into categories can be expected to change in the future in unpredictable ways, just as it has in the past. As a rough rule of thumb, I would suggest maintaining as a minimum the amount of detail provided in the IPUMS samples for the latest censuses.

**Earl Ruiter, Cambridge Systematics, Inc., Other Nonacademic (Nonacademic)**

The trend in transportation research is toward lifestyle classifications. Loss of information on income, age and occupation will limit the usefulness of the PUMS data for these types of studies.

**Confidential Response**

My work is focused on racial inequality in the US. Reduction of PUMS detail may preclude me from on-going analysis of ways racial inequality changes, intensifies and diminishes over time.

**Louis Kyriakoudes, University of Southern Mississippi, History (Faculty)**

I study internal migration and economic change. To limit the geographic, occupational, income, racial or education data would greatly impair my research

**Janice Madden, University of Pennsylvania, Economics (Faculty)**

PUMAs of 100,000 have allowed us to examine the effects of race, household structure and education on job and residential locations. The individual data in PUMS are unique in that they allow us to examine the simultaneity of these location decisions. Given the criticality of spatial mismatch to overall economic performance and to equity, it is important that we understand its dimensions and its causes and effects, both for science and for policy.

Detailed occupations and small PUMAs are also critical to measuring the effects of, and assuring the enforcement of, affirmative action and anti-discrimination policies. While many firms can use aggregated occupations, the new economy is relying on much finer matches between skills and jobs. To measure compliance with non-discrimination norms, it is important to have detailed occupational data on workers in the vicinity of the firm.

**David Hollingsworth, Nunez Community College, Other Academic (Researcher)**

using the regions sets for prediction of both academic and occupational training needs is crucial in planning the ed services, marketing strats, and overplanogram for resource allocation here...

**Confidential Response**

I use PUMS data in my traffic research. A reduction in geographic detail and detail on income and other factors which influence travel patterns would seriously impact the accuracy of my forecasts.

**Willis Goudy, Iowa State University, Sociology (Faculty)**

Gerontological studies need detail for more than a 65-plus category, primarily because of the gross differences between those in various subgroups above that age. At a minimum, 65-74, 75-84, and 85-plus age categories are needed. Housing, health, and many other factors vary substantially across these age groups. Both basic and applied research would be hindered without information by more than one age category for those 65 or older.

**Debbie Pitts, Missouri State Library, Demography (Nonacademic)**

I am the coordinator for the State Census Data Center Program in Missouri. My network of data users will be greatly affected by a reduction in PUMS data. One of the members of our coordinating group uses PUMS data extensively to support the work of local planners and regional planning commissions in Missouri.

**Patricia Beeson, University of Pittsburgh, Economics (Faculty)**

I am currently using the pums data in a research project examining differences in wage structure and wage inequality across metropolitan areas. For this we need to have at least metropolitan areas identified. Since we are looking at changes over time we need historical comparability. This is very difficult as is, given changes the census has made over the years in reported substate geography. 1960 data cannot be used for substate analysis, and it would be catastrophic for researchers examining local areas to also lose 2000. I hope the Census will maintain some substate geography. While I prefer MSAs, the PUMAs would be okay. Also, since we are looking at wage structures and wage inequality, we need the detail on individual characteristics that might affect wages (age, occupation, race).

**Don Burrell, OH-KY-IN Regional Council of Governments, Demography (Nonacademic)**

OKI is a transportation planning agency (Metropolitan Planning Organization). For our use, the occupation, housing and household characteristics, and vehicle ownership data are important. The purpose of PUMS data is to enable development of special cross tabulations for specific purposes that would not have wide use. Therefore the more subject detail, the better. The compromise is in the geographic detail for confidentiality. Hopefully the 100k limits will continue to be sufficient to insure this.

**Confidential Response**

We have been conducting housing needs analyses for state, county, and local governments for almost two decades. Grouped age data will not give us the level of detail we need to conduct the household formation projections on which these analyses are based.

**Confidential Response**

Our primary use of data is for economic development and planning. However, all the data is important when you are trying to entice more companies, or job training programs, or social issues. Then bring planning into your work. How can you effectively work towards economic development and planning when you don't know what you have? I believe that you can't have enough various types of data when making decisions that will affect peoples lives.

Response time has always been less than what a person wants but on a wish list, I'd like an overview of data, then I'd be willing to wait for detail.

Please keep answers confidential if you can.

**Sara McLafferty, Hunter College, Geography (Faculty)**

I am very concerned about the loss of detail in all areas, but for my research, geographic detail is crucial. To understanding commuting patterns, geographical barriers to employment and questions of spatial mismatch, we MUST have data for relatively "small" areas (i.e. populations of 100,000 as in the past) Otherwise it is impossible to describe and analyze the movements of people from residential areas to worksites.

**Edward Schafer, San Diego Council of Governments, Government (Nonacademic)**

A significant aspect of my work deals with industrial clusters. These are defined from detailed industry classification. The loss of this detail would severely restrict my ability to produce comparable information over time.

**Jack Solock, Center for Demography and Ecology and Center for Demography of Health and Aging, University of Wisconsin-Madison, Demography (Support)**

I am a data librarian who serves over 100 faculty, students, and researchers in two demography centers. Many of these researchers use PUMS data extensively and would be severely affected in their research by the proposed reduction in PUMS detail. In particular researchers in aging in the Center for Demography of Health and Aging would be hampered by grouped age data.

**Don Larrick, Office of Strategic Research (the State Data Center in Ohio), Government (Nonacademic)**

Leave the age variable as it was in 1990. Our research and the work we do for clients in-and-outside of state government requires that we be able to create age groups as desired.

**Erzo Luttmer, University of Chicago, Economics (Faculty)**

In our research we identify the role of network effects in welfare participation decisions among non-native English speakers using the fraction of speakers of the same language in a PUMA as a proxy for the size of the network. Loss of geographical detail (larger PUMAs), loss of detailed ethnicity/language information or loss of detailed income information would be detrimental to this line of research. In other research, I look at the effects of behaviors and racial composition of welfare recipients in one's area on support for welfare spending. This research too relies heavily on PUMS data where detailed information at the PUMA level is critical.

**Richard Sherwood, Maine State Planning Office, Sociology (Nonacademic)**

I am a policy advisor to the Governor of Maine and his cabinet. I have probably used the PUMS files two or three hundred times over the past twenty years to address issues raised by the cabinet or the legislature. It is essential in this sort of setting to be able to undertake customized tabulations and cross-tabulations. Sometimes, I am asked about a policy pertaining to a narrow age cohort, e.g., ages 2 to 3 or ages 13 to 15. Sometimes, I am asked to focus on a specific industry or occupation or income level. For example, I have used the PUMS file several times to analyze residential property tax burdens for households in which the number of occupants, the householder's age and the household's income all met criteria specified by a legislative committee or by the State Tax Assessor.

**Robert Burchell, Center for Urban Policy Research, Rutgers U., Other Academic (Faculty)**

We use the PUMS data (CD's) for various projects, although much of my work is related to determining need for affordable housing. In order to use the PUMS we need age data, but grouped by 5 year groups would be work. Reducing geographic detail would be disastrous. Reductions in detail regarding income, beyond grouping to ranges of \$100 or maybe \$250, would also be disastrous. We already have contracts with two agencies, the New Jersey Council on Affordable Housing and a large county in another state, which rely heavily on the availability of PUMS as it was done in 1990. In the past we have used PUMS with a number of other governmental groups to look at housing affordability. We have also used the PUMS for determining the characteristics of households in new housing and for various journey to work studies.

**Stephen Wilson, SRF Consulting Group, Industry (Nonacademic)**

Working and commuting patterns of the 65-69 age group very significantly for the 70+ . As the workforce ages this becomes a critical issue for the travel behavior business.

**David Kruse, Alamo Area Council of Governments, Government**

We need PUMS data for small area forecasting of population and employment in support of transportation planning, especially travel demand forecasting. I have also used PUMS data for social services planning.

**Joyce Benson, Maine State Planning Office, Government (Nonacademic)**

State level policy research/analysis has relied heavily on PUMS to obtain the level of detail to gauge impacts of various programs and policies. I am often asked questions such as what is the social and economic profile of persons who work as farm laborers by a legislature or state dept. considering the development of programs or services to the farming industry. This kind of profile can only be developed by having detailed information on age, household structure, housing conditions, income, etc. as found in PUMS. I have relied on the data from the 1990 pums file hundreds of times in the past decade. Historical consistency is also critical to my work.

**David Elesh, Temple University, Sociology (Faculty)**

I am an urban sociologist who continually works on issues of occupational and industrial change. The available database--particularly with respect to industrial change--are already very limited. In a period of rapid industrial change and concomitant urban change, the proposal would seriously damage researchers' ability to examine the causes and consequences of change.

**John Garcia, University of Arizona, Government (Faculty)**

On the race data, reduction of the newer combinations of racial categories would undermine the breadth of the expanding racial make-up of the U.S. also it would be more difficult to conduct "historical" analysis over past censuses.

**Sarah Wilhelm, University of Utah, Economics (Post-doc)**

Discovering the effect of antidiscrimination laws on occupational segregation by race would be impossible.

**Jerome Deichert, University of Nebraska at Omaha, Other Academic (Faculty)**

We have had a number of questions from our Congressional delegation concerning the impact of various policies on Social Security recipients in our state. To do this we needed detailed income by detailed age by type of income so we could run different scenarios by different age and income cutoffs.

**Joseph Willey, Binghamton University, Other Academic (Support)**

Academic support, so focus is others research needs. Many geographers, sociologists and economist relying on PUMS. Most notable needs are in the field of labor economics. Without good quality, COMPARABLE, Pums data cross-temporal comparisons are incredibly difficult. In this, as in all cases, I am pleading with the Census Bureau to actually stick with consistent classifications for a while and not tinker with definitions and indicators every ten years.

**Claudia Goldin, Harvard University and the NBER, Economics (Faculty)**

I would like historical continuity in the PUMS.

**Tara Watson, National Bureau of Economic Research, Economics (Graduate)**

I do research that requires detailed race and income information. Grouped income bins make it impossible to accurately determine changes in income across areas or demographic groups. Suppression of racial information would also make it impossible to do my research.

**Betty Dodson, university of Nevada, Reno, Other Academic (Faculty)**

I live in the State of Nevada. We have had a high in-migration of elders. Geographic information and age data are essential for a state that is experiencing phenomenal percentage growth in its population base. In fact, any and all detailed information is important as we lack a sophisticated infrastructure at the state level for collecting useful information that will support grant applications and help to identify areas of need.

**Jeffrey Kamakahi, College of St. Benedict / St. John's University, Sociology (Faculty)**

I use the data in my statistics and demography courses as well as for background on other research. Since data can always be truncated into larger categories, the greater the detail in the data the more useful it is for researchers and research questions. Aggregation of income data, for instance, might very well mask curvilinearity and would have an effect upon what statistical procedures the researcher would be able to perform in the analysis.

**Walter Treichel, Wilder Research Center, Demography (Nonacademic)**

PUMS offers the only way in which sub-populations not included in Census crosstabs can be studied. I find geography particularly important. Minnesota has 30 PUMAs currently and degrading that to about 15-18 would make it almost impossible to draw distinctions between city, first ring, second ring, and rural residents. This level of detail is critical for us in conducting research and program evaluation.

**Matthew LaPenta, Federal Reserve Bank of Boston, Economics (Nonacademic)**

My research is concentrated on migration patterns by educational attainment, age, race, occupation, and income. Comparability across years is essential for me to look at migration trends over time.

**Linda Giannarelli, The Urban Institute, Other Nonacademic (Nonacademic)**

In the past, the PUMS data have been used as input to microsimulation models of tax and transfer programs. The PUMS have never had the detail of the CPS, but they do offer the possibility of state-specific analysis. Limiting the amount of detail on income could make it impossible to use the PUMS for microsimulation in the future.

**Keith Lawton, Metro, (MPO in Portland, Oregon), Government (Nonacademic)**

Note: We do limited "research", But we use the PUMs/PUMAS extensively in the development of synthetic populations by small area (Block Group) for application in the microsimulation of personal travel in models of urban travel demand. (A planning application). In this arena both income and age are related to demand for mobility, and ages above 65 show decreasing demand for travel that we need to understand when forecasting a future with an aging population.

**Matthew Ides, UCR, History (Graduate)**

My last research with IPUMS was an examination of Asian immigrant groups to the US, measuring their social mobility, occupations and geographical locations. I analyzed this data in order to explore the impact of different immigration restrictions. If this categories were collapsed, IPUMS would not have been useful to me. Grouped age data or any other groupings diminish the specificity of social research. The quality of PUMS is that its multi-variate data allows the researcher to refute theories based on broad based aggregates, as well as allowing the research to explore aspects of American society in

greater detail. It is much easier for an individual to pay 25\$ for a web service to find out someone's personal information, why and how could someone use IPUMS for this purpose?

**Jeffrey Reitz, University of Toronto, Sociology (Faculty)**

My research examines the economic success of immigrants as this varies across urban areas, looking at specific racial origins groups, and examining the effect of changing occupational characteristics. The analysis requires information needed to construct conventional human capital models (e.g., education, age, gender). Reductions in the detail of racial data would severely reduce my ability to consider and/or eliminate race as a factor in urban area differences. Reductions in the detail occupational data would make it difficult or impossible to examine how labour market change is affecting the success of immigrants.

**Pamela Klein, Penn State University, Sociology (Graduate)**

I specialize in military sociology so any reduction in occupation and income would have negative effects on how easily military personnel are identified. As it is, I have a very difficult time determining if someone who identifies themselves as a military personnel are enlisted or in the Reserves or ROTC, etc. If there was a reduction in detail concerning age, occupational status, and income, I could no longer even attempt to analyze the level of poverty in the current military ranks, nor the differences in poverty levels among veterans of different periods/conflicts.

**William Grant, Old Dominion University, Government (Graduate)**

GIS analysis of multiyear census data would be seriously impacted resulting in loss of urban analysis and a significant reduction in the effectiveness of this emerging analysis tool.

**Patrick Costinett, Parsons Brinckerhoff, Industry (Nonacademic)**

PUMS has been and is becoming even more important in the development of behavioral travel demand forecasting models in which households and members are classified by multiple characteristics such as income, occupation, industry, age, household size, and vehicle ownership. Similar information is collected in special surveys but these involve small sample sizes and high cost per sample. PUMS is a critical basis for survey expansion, analyzing sample bias and providing more detailed cross-classifications of households and individuals for model building. Maintenance of socioeconomic detail is more important than greater geographic detail but race and ancestry are generally not of interest in travel demand modeling.

**Michael Gillett, Oregon DOT, Government**

We use the data for travel estimation and forecasting. The detail that is needed for our work is geographic detail. The household demographic data is often aggregated to three or four income groups, age groups, HH size groups, etc.

**John Pitkin, Analysis and Forecasting, Inc., Demography (Nonacademic)**

I study housing consumption by birth and immigration cohorts. Loss of age detail for the elderly, or at any age, would limit my flexibility to compare specific cohorts between censuses. This is fundamental to my research. I am especially concerned that with the expected shift to ACS data, I will need more detailed PUMS data to calibrate PUMS-to-ACS changes for non-standard age groups. The ability to look at PUMS data for smaller areas (100,000-plus) has been very important for intercensal comparability, e.g. comparing metro areas with constant boundaries. The larger the "building blocks" the worse the intertemporal comparability.

**Irene Browne, Emory University, Sociology (Faculty)**

I study race/ethnic differences in labor market outcomes among women. In this area, there are some important trends that have important policy implications. For example, by 1979, the earnings of Black and White women were almost equal. In the 1980s and 1990s, Black-White wage inequality INCREASED. Why are we moving towards more racial inequality? This is particularly important as a policy issue given that many women are supporting their families alone. The detailed occupational and industry categories, income, labor market status and geographic boundaries in the PUMS data are INDISPENSABLE for my research. I create aggregate measures at the MSA level to track changes in economic opportunities. For instance, in my most recent publication, I show that changes in employment opportunities within PARTICULAR TYPES OF SERVICE INDUSTRIES (retail services) benefited White women but not Black women. (Social Forces, March 2000) I COULD NOT HAVE PERFORMED THESE ANALYSES WITHOUT PUMS DATA THAT WAS COMPARABLE ACROSS YEARS. The 2000 PUMS will be even more important, as it is the ONLY dataset that captures the characteristics of relatively small but rapidly growing race/ethnic groups, such as many Asian groups. Only the PUMS, with its 1990 detail, allows for the possibility of creating population and labor market measures that can help determine how changes in labor pools within local labor markets affect women's wages and employment.

I am also intending to look at labor market trends among Asian-American women. This group can ONLY be studied nationally with decennial census data (the CPS is inadequate). The geographic information is important, as different areas of

the U.S. have very different economies. The occupation/industry/income detail are important because we cannot determine how Asian women are faring in the labor market without knowing precisely which occupations and industries in which they are employed and what they are earning.

**Frank W. Sweet, American Military University, History (Graduate)**

My research focuses on the history of U.S. black/white intermarriage; how it varied over time, and across states. Without race, marital status, and geography, the census would be worthless.

**Mark Hardt, Montana State University-Billings, Sociology (Faculty)**

I do time series research. A major problem I often encounter is a difficulty, if not an inability, to measure variables over time. Reducing the number of categories will have a serious impact on the ability to conduct historical research and to analyze trends.

**Jonah Gelbach, University of Maryland at College Park, Economics (Faculty)**

Much of my research uses age structure to identify effects of public expenditure programs. Moreover, my research frequently involves using variation across states in public policies. Without these variables I might as well not have the PUMS at all.

**Blaine Billings, Boise State University, Sociology (Undergraduate)**

Life span longevity (higher mortality age) Higher age requirements on retirement Social Security Bankruptcy scheduled for 2035 (determined by US Department of Social Security) Between the ages of 65 and over covers a minimum of 4 extremely significant socioeconomic categories. 1) Still working 2) Economically stable on retirement 3) Economically assisted by government subsidy 4) Total public subsidy for medical care

**Melissa Levitt, San Francisco State University, Other Academic (Faculty)**

My research areas are cities, and I primarily concentrate on race-based or ethnic based shifts in population, income brackets and occupational changes. Any modifications that would limit comparability over time would severely damage my ability to conduct thorough research.

**Andrew Harvey, Saint Mary's University, Economics (Faculty)**

My interests are in regional research, occupational change and the aging population. With the labour market as fluid as it is aggregated data along any of these dimensions would cripple much of the analysis than should be done.

**James Gregory, University of Washington, History (Faculty)**

I study the effects of migration and need both geographic detail and the full range of demographic detail. To scale back either would be a tragedy with both immediate and long-term implications. I pity the policy analysts, not to mention the historians, who will have to look back on the 20th century as the golden age of social research.

**Benjamin Ben-Baruch, StarWorks, Inc., Industry (Nonacademic)**

The most catastrophic affect that decisions of this kind could have on market research would be to compromise confidentiality and to cause people to lose faith in the ability of the Census Bureau and/or of researchers to protect privacy, maintain confidentiality, and protect the integrity of research.

Maintaining public confidence that privacy is being protected is -- in the long run -- more important than better data.

**Bill Godair, Old Dominion University, Other Academic (Graduate)**

With reference to question 6: Need to broaden occupational categories to reflect emerging technical occupation, however at the same time, there is a need to maintain comparability across census years.

If geographic detail is reduced then the ability to track changes and compare changes in educational achievement, industry and occupational changes, between different geographic area would be lost. Since I am primarily interested in changes at the central city and MSA level of detail I would be unable to conduct my research.

**Confidential Response**

My research tracks men's work involvement relative to their family status from 1940 on. Although I have only been tracking behavior thus far, the next stage of my analysis will be to analyze my data according to various demographic variables. At some stage of my research, I will want to use fairly specific racial and occupational categories. If these variables are made more general, it might hurt my research.

One of the great things about census data is the amount of specific data with such a large sample size. There is nothing else that even comes close. I think it would be a mistake to reduce the level of detail available to researchers.

**Anne Jachim, Moraine Valley Community College, Other Academic (Faculty)**

Our world and technology are changing at an astronomical rate. Those of us involved in research for either academia or business depend upon the research results of PUMS as a viable resource.

**Louise Gunderson, University of Virginia, Other Academic (Graduate)**

I am working on the problem of crime prediction. Detailed information is vital to this work.

**Confidential Response**

The type of research I am currently working on requires detailed information at the level of the census-tract. When this is not available, I have been forced to use metropolitan areas as the unit of analysis. I require social and demographic detail on residents of neighbourhoods over time (1900-2000) and therefore seriously rely on consistency in data compatibility across decennial census years. Race, socio-economic status, and age are integral components of my research, and to limit the categories available would jeopardize analyses over time. Longitudinal research is extremely fruitful, relative to cross-sectional research, in determining time-order and macro-level trends. To alter the type of data released for the 2000 census would, in essence, impose severe limitations to longitudinal research and may even reduce the utility of cross-sectional research by limiting the richness of descriptive data.

**Deborah Ward, Columbia University, Government (Faculty)**

The most important loss is in comparison over time. We should never go backwards in data collection, only forwards.

**Kenneth Sunderland, University of California - Riverside, Economics (Graduate)**

I use nonparametric methods to extract and test information on income and earnings distributions across virtually all the categories you named above. In order to do this, I need datasets that are both large and detailed so that I can break them down in a variety of different ways and still have sample sizes large enough for reasonable nonparametric estimations. Without sufficient detail, comparisons across categories and subcategories become difficult or impossible. In particular, I have developed a method of quickly generating distributions for age on a per year basis that reveals considerable detail in the evolution of income/earnings with increasing age that is not available with parametric methods. This would become almost impossible with aggregated data.

**Wendy Gordon, Plattsburgh SUNY, History (Faculty)**

Because I research the nineteenth and early 20th centuries reduction of detail in the 2000 census would not immediately jeopardize my research. However, loss of detail in any variable endangers the ability of future historians to analyze data by methods and for reasons that we cannot imagine today.

**Carrie Yang Costello, University of Wisconsin-Milwaukee, Sociology (Faculty)**

I do research on professionalization, for which detailed occupational categorical data is necessary. I also do research on identity, for which ancestral and racial data are vital.

**Jill Marlette, Geography (Nonacademic)**

I use income, age, occupation and other variables to create maps. The more detail I have, the more informative my maps are.

**Charles Starnes, Oregon State University, Sociology (Faculty)**

Income distribution data have always been so crudely truncated at higher income levels as to severely damage research on topics where income inequality is a relevant consideration. We need more information on distribution at higher income levels, not less!

**Confidential Response**

I am a graduate student at Queens College/CUNY and using the IPUMS for my thesis. Any reduction of any sort regarding the data quality would definitely prove to be harmful to me, to other researchers, and to the general citizen who would like to "see" how the society they live looks like.

**Paul Voss, University of Wisconsin-Madison, Demography (Faculty)**

I mostly do applied demographic work. When the STF's are not detailed enough, one must turn to the PUMS data. The geographic detail is already too large. If it were increased to something larger than areas with minimum 100,000 population, the PUMS would lose much of its usefulness to me.

**Jason Fields, U.S. Census Bureau, Demography (Nonacademic)**

While I am certainly biased, and have access to internal files for any runs off the 2000 census that I will need to make, I have used IPUMS data to supplement a bureau report and hope to continue using it. The ability to make time series comparisons is invaluable and should be retained in as much detail as possible. Examination of detailed living arrangements will certainly require the detailed age data beyond age 65, for both aging and child related research.

**Leslie McCall, Rutgers University, Sociology (Faculty)**

On the specific question of how to deal with occupations (question #7), I think a major issue is that certain detailed occupational categories are very "timely" ones and these cannot be determined in advance (e.g., securities analyst, or, in terms of industry, personnel supply industry workers).

On the more general question of the broader impact of less detail, it would be truly catastrophic since my research is on earnings inequality in \*PUMAs\*, the smallest geography available in the PUMS and the only one that includes non-urban areas. States are too large to use as labor markets and urban areas are not representative of the country as a whole or of the differences between segments within large metro areas (i.e., between central city and suburbs). Since more and more of our theories and policies emphasize the importance of place and locality, and the U.S. has an extremely large and geographically diverse population, it seems that the Census should be moving toward more geographical detail rather than less. Another issue is income. The U.S. has a very poor documentation system when it comes to wealth. Given the large accumulations of wealth over the 1990s, it seems especially important for the 2000 Census to be able to accurately reflect the structure of income (if not wealth), which can't be done if there is top-coding. This has always been a major weakness of the census.

**Dunbar Brooks, Baltimore Metropolitan Council, Government (Faculty)**

We are an MPO and we use these data for a variety of purposes in transportation planning and to assist local governments in developing policies and/or building a case to apply for federal and state funding based on the detailed crosstabulations that we can make. The smaller the geographic detail the better. PUMS data is only data that allows you complete freedom to develop critical, but non-published data tables that may fit special circumstances. We have used PUMS to assist state and local governments in developing child care programs. We have used PUMs to assess certain conditions among the disabled. I have been making PUMs tabulations that aid both government and private industry for a quarter of a century. It would a travesty to now dump this program or to severely diminish it.

**Julie Whittaker, Rutgers University, Economics (Faculty)**

Wage distribution by sex and race. As a labor economist I cannot think of what could be worse for my research than these proposed changes. Geography, age, race, sex, occupation...all of these are necessary to understand labor markets and the fundamental changes that occur.

**Confidential Response**

We used the data to track the educational attainment of teachers, looking for trends and focusing on age, race, and gender. Losing access to this information may hinder the research that influences the teacher training programs across the United States.

**David Olive, Southern Illinois University, Other Academic (Faculty)**

It could be useful to make year by year comparisons. For example, having all of the variables from the 1800's in the 2000 data set would allow comparisons between historic and current data.

**Confidential Response**

I have worked on some research on Italian Americans and if there was limited information available on ancestry this research would be severely compromised.

**Dr. Gene George, Butler County Community College, Demography (Support)**

I do demographic research to support strategic planning at a community college in south-central Kansas. It is critical that we know as much as we can about the six-county region we serve. We especially need to understand the long-term dynamics of population growth and decline and the evolution of an urban society in what was once a rural area. Analyzing the change in income, occupation, and race is vital to our understanding of what's going on in the region. And since we rely on demography to identify more specific qualitative research (focus groups, depth interviews), a loss of demographic detail will diminish the effectiveness of our social research into emerging trends. Thus, the college's capacity to make effective plans for its future will be damaged.

**Confidential Response**

We use the PUMS data to identify sample bias resulting from non-response. We also rely on PUMS data in the development of our sampling plans, to identify our sampling universe. If reduction is necessary, the sooner we are aware of the new categories, the easier our post-processing will be (i.e., we can design instruments to collect data in line with the new categories). If reduction is necessary, we request that the reduction take place along current categories so that comparisons can be done at some level rather than not at all.

**Ken Prandy, University of Cardiff, UK, Sociology (Faculty)**

My research, which is part of an international comparative study using census data from a large number of countries, involves an analysis of the occupations of married (or cohabiting) partners in order to develop a graphical representation of social space. The first dimension of this space is one of generalized social advantage/disadvantage, so that the co-ordinates provide a measure of social hierarchy.

**Mark Killingsworth, Rutgers University and NBER, Economics (Faculty)**

On #9 above, it might be possible to release a less detailed dataset immediately, and LATER release a more detailed dataset - the two strategies aren't mutually exclusive.

In answer to this question, analyses of earnings need detailed occupation data, and an enormous amount of detail would be lost if the detailed occupation categories were collapsed. Note that this has significant implications for equal pay policies.

Likewise, analyses of earnings and of hiring need detailed data on residence and commuting patterns. The STF files don't allow one to do any meaningful analysis controlling for covariates like education -- you have to take the info. in exactly the form the STF tables give it to you. The alternative, which is far superior, is to use the PUMS data, where you can look at reasonably detailed info. on residence/commuting patterns and you can also control for education, age and the like. Note that here too there are some significant implications for public policy -- for example, "utilization analysis" for EEO purposes often requires info. on the geographic distribution, within rather small areas, of persons with specific characteristics (education, age, occupation, etc.) by race and sex. A lot of this would be lost if the geographic detail were condensed.

**Hannah McKinney, Kalamazoo College, Economics (Faculty)**

The key issues in US urban policy today are race and class. We need to untangle the interactions between income, education, age, race, and class to try to understand the extent of spatial and labor market discrimination in the market. This data is indispensable for this purpose.

**Michael Stoil, Conwal Incorporated, Industry (Nonacademic)**

Health care delivery research by ethnicity is not available from SSI/Medicare records because SSI Idents lists only White, Black, and Other. As the population diversifies and we become more concerned about the interaction of gender, age, insurance and health care access, and ethnicity on health status, there is no substitute on a nationwide basis for PUMS detail. We simply will not be able to accurately track these variables if detail on race, nativity, age, and ethnicity are not available.

**Harriet Duleep, The Urban Institute, Economics (Researcher)**

I do research on the socioeconomic determinants of mortality (including occupational experience, income, where one lives), as well as the economic assimilation of immigrants and their effect on the employment opportunities of the native born. Obviously, having very detailed information on all of the variables discussed above is imperative for the quality of this research, and its ability to speak to important policy issues.

**Confidential Response**

I specialize in research in American economic history that make VERY heavy use of IPUMS data -- for example, data on occupation, educational attainment, home ownership (the latest paper, covering 1900 to 1990), migration, income, and many other variables. Such studies are essential for putting current policy problems in their long-term historical context; without the necessary detail, such studies are difficult or impossible to conduct.

**Mary Kritz, Cornell University, Demography (Researcher)**

My research focuses on demographic and socio-economic analysis of the foreign-born population in the USA. I have carried out a number of research projects using 1980 and 1990 PUMS files. Most of my analyses involve disaggregation of the foreign-born population by origin and geographic residence in the USA. Several of the research projects that I have done using the 1990 files would be precluded if a higher level of aggregation is utilized and the research that I would like to do has already been constricted because of sample size constraints.

One of my projects has focused on the internal migration patterns of the 20 largest foreign-born groups in the USA, using PUMS files. In 1980, a similar analysis could only be done for 10 foreign born groups. (Some peculiarities in how the Census

Bureau classified internal moves also constrained analysis possibilities in 1980.) Internal migration of the foreign-born into non-metropolitan areas is a topic that was already precluded in 1980 and 1990 because of restricted sample size.

Another one of my projects focuses on living arrangements of elderly foreign born from the 10 largest origins. On this topic also, the number of persons aged 60 and over from different countries of origin is already very small. I did some explorations with the 1990 files to determine whether there were sufficient cases to permit comparisons between immigrants and native-born persons of the same ancestry. Using the ancestry classification (i.e. 2nd generation and higher), I discovered that there were only a sufficient number of cases [elderly 60+] from the two largest countries of U.S. immigration (Mexico and the Philippines [N=812 from the combined PUMS 1/1000 and PUMS-O]). If I aggregated Chinese from different origins, not an advisable strategy, I would also be able to obtain a sufficient number of persons of Chinese ancestry for national-level analysis [N=2,507]. However, for the 4th largest U.S. immigrant group, Dominicans, the number of elderly of Dominican ancestry drops to 119 cases. As these figures indicate, the number of cases for detailed subgroup analysis is already too small for certain types of analyses based on the 1990 PUMS.

A third illustration of a research project that I did on which I quickly "ran out of cases" involved comparisons of women's labor force participation for Hispanic foreign-born women from the 5 largest groups (Mexico, Dominican Republic, El Salvador, Cuba, and Colombia). I wanted to compare how their levels of labor force participation differed by metropolitan labor markets. After checking sample sizes, I discovered that I only had a sufficient number of cases in 3 metro areas -- New York, Miami, and Los Angeles.

I have been looking forward to replicating and extending analyses done on the 1980 and 1990 files, based on the expectation that the sample sizes would be larger for the 2000 PUMS given the continued and large volume of U.S. immigration during the 1990s. Apparently that may not be the case. As the U.S. population becomes more diversified, it is important to do subgroup analyses of multi-racial/ethnic populations.

#### **Douglas Gurak, Cornell University, Demography (Faculty)**

Other countries, e.g. Canada, already group information for place of birth and those data for national origin are not very useful. Having as detailed information as possible for origin, occupation, location, age, and other characteristics such as income is essential if we are to be able to learn anything about what is really going on in important areas such as: 1) occupational success of minorities, immigrants, or other groups; 2) migratory patterns, causes and consequences of the general population and specific sub groups such as immigrants and the elderly; 3) fertility trends and correlates for the general population and specific ethnic groups. A lack of detail in any of these and similar areas lends itself to findings that are misleading because they are averages of more complex underlying patterns. The greater the detail, the better our ability to identify such complexity.

#### **Matthew DeZee, University of Maryland (at College Park), Economics (Undergraduate)**

The reduction of detail would be catastrophic to future economic research.

A general grouping of age or income would strongly deteriorate any research in labor economics.

#### **Warren Barclay, New Jersey State Department of Personnel, Government (Nonacademic)**

1. With employees retiring later, past 65, it is necessary to have occupational data by the "above 65" years to determine availability for affirmative action programs.

2. Having less occupational detail will hinder establishing availability percentages for affirmative action programs.

#### **Gary Portenier, Milwaukee County Department on Aging, Government (Nonacademic)**

Our use of PUMS data has been limited in the past, but hope to use this information in conjunction with mapping software to improve program planning.

#### **Rowan Wolf, Portland Community College, Sociology (Faculty)**

Reduction in PUMS data would impact my research and that of other sociologists dramatically. First, reducing the detail decreases and may invalidate historical comparisons of the population as a whole, and of specific populations. Second, sociological and work force research indicates there are transitions in how inequality works in the labor force which is a trend towards smaller units - rather than industry-level units. Losing the detail masks the ability to see these changes. Third, we have rapidly changing occupations and race/sex stratification within fields (i.e. clerical, decreasing detail makes it impossible to see these changes).

**Bobray Bordelon, Princeton University Library, Economics (Faculty)**

Princeton University uses the PUMS data for Economics, Sociology, Demography, and Politics. I see a broad spectrum of students and faculty.

**Eileen McConnell, University of Notre Dame, Sociology (Graduate)**

My research focuses on Mexican immigrants in the Midwest, thus, less detailed information about ancestry would mean that I could not determine whether individuals are foreign born, whether they are children of immigrants, etc. Additionally, states and even MSAs vary widely in proportion of the foreign born etc--thus less detailed geographic areas means that estimates of local/area characteristics are less precise. Similarly, grouped occupational categories means that it would be difficult to compare detailed occupations within the manufacturing/industrial occupational sector across regions of the country--for example, percentages involved in food processing may be similar in CA and NE (they are not--but just as an example)--but employees in CA are involved in canning while in NE they are in meat processing and packing. With grouped occupational data--that might be impossible to differentiate.

**Ken Blum, Center for Labor & Community Research, Other Nonacademic (Nonacademic)**

I have done research using specific occupations. If this detail was eliminated, it would destroy this type of research. We serve poor communities.

**Elaine Murakami, US Department of Transportation, Government (Nonacademic)**

The transportation community pays for the Census Transportation Planning Package through a cooperative project via AASHTO. However, the CTPP is only part of the Census data used by the transportation profession. PUMS is also a critical Census data source for transportation planners.

We are primarily concerned about the size of the PUMAs. We do not want to see the PUMAs be any larger than 100,000 population, and would, in fact, prefer that the size be decreased as much as possible. When the geography of the PUMA becomes too large, the PUMS becomes less useful for transportation planning. For example, transit use is directly tied to accessibility to transit buses and rail systems. If the PUMA includes areas with both high and low transit accessibility, we are less able to use the data for regression or other analyses on travel mode choice. Similar arguments can be made for other census variables such as departure time to work, vehicle availability, etc., and related models and simulations.

There is a trend in the transportation community toward microsimulation of travel behavior to address congestion and air quality issues. That is, characteristics of individuals and characteristics of their household are used to develop activities and trips which are applied to the transportation network to forecast such things as travel volume and travel speeds, which are then transformed into estimates and forecasts of congestion and air quality.

We would rather give up detail on response categories than give up geographic detail. That is, we would be willing to live with some grouping or collapsing of categories for age, income, and occupation. Because of environmental justice concerns, we DO need race and Hispanic origin data, but we could probably live with a limited number of race categories in the PUMS, since the detailed counts will be available at the block level from PL-94-171.

Submitted by CTPP Working Group

**Confidential Response**

Often need the PUMS data to assist in comparing other suveys on income by age, and region, race etc. We never know exactly and would have to request special tabulations if not available.

**Patricia Fernandez-Kelly, Princeton University/Department of Sociology and Office of Population Research, Sociology (Faculty)**

I conduct ethnographic research in urban settings. Because my work emphasizes behaviors found at the local level, I often depend on detailed Census information concerning SES, occupation, ancestry, and other demographic characteristics. Reduction in the level of detail on these and other characteristics would have a very harmful effect on my research.

**Lorraine Porcellini, ISR, Temple University, Other Nonacademic (Nonacademic)**

To determine sample sizes for a study, I use PUMS data to calculate the proportion of the population I can expect for certain subcategories, most often age, sex, race, Spanish origin, marital status, income, and education. I use these data to determine the number of households I must sample to attain the required subgroup sample sizes. I often need to know the breakdowns by state, region, MSA, or PUMA.

**Bruce Western, Princeton University, Sociology (Faculty)**

I have used the PUMS to study racial inequality in employment, and earnings and employment among immigrants. The PUMS is an invaluable data source for studying narrowly defined demographic groups, such as young minority males with

low levels of education. No other source offers the necessary quantity of data for studying such small groups over a relatively long time span. Such research is only possible because the PUMS provides detailed information about age, region, and schooling. If only aggregated codes were reported, the value of the PUMS would be significantly diminished. Rival data sources, like the Current Population Survey, provide relatively large sample sizes for coarsely defined groups. In my view, eliminating detailed demographic codes in the PUMS would largely undermine the utility of the Census for social research.

**Jim Reilly, NJ Office of State Planning, Government (Nonacademic)**

If reductions are unavoidable, they must be made in a manner that allows comparability with earlier data.

**Confidential Response**

Regional modeling requires consistent data at the county level for my research, therefore the loss of detail in the dataset would make future updates difficult if they weren't in some way compatible.

**Confidential Response**

Need maximum level of crosstabs possible for detailed data on age, income, migration, and housing.

**Confidential Response**

We are doing transportation planning consulting to all level of governments. Accurate and more up-to-date data are essential to an accurate planning. Age, income, type of job, geographic location of both home and other activity location are essential to the kind of work we do.

**Brian Cushing, West Virginia University, Economics (Faculty)**

Reducing geographic detail would significantly (maybe severely) limit my ability to study metropolitan migration and the usefulness of my findings. I already lose the smaller metropolitan areas and have limited ability to distinguish central cities from suburban areas. Loss of geographic detail may eliminate most of my research on rural poverty and migration since the PUMAS would be too large to be useful - would have to lump together too many distinct areas. Any loss of detail on income could wreak havoc in research on income distribution, income inequality, and poverty. The decennial census is the only data source that provides the combination of large sample size, detailed characteristics, and geographic detail to allow detailed stratification by personal characteristics and type of place.

**Eberts Randall, W.E. Upjohn Institute for Employment Research, Economics (Nonacademic)**

One area of research is economic development and the dynamics of local labor market. The PUMS offers the most detailed information about the workforce in smaller metropolitan areas. Without these data, we and others could not provide the analysis that policy makers need and want.

**Confidential Response**

In Hawaii, we have a great diversity of ethnic groups. There is a great need for the cross-tabulation of data on population and housing characteristics by specific detailed race category. Frequently, these requests involve three-way or more cross-tabulation of data as well and/or using one or more selection criteria. We have had many requests over the years where the PUMS data is the only one available that is able to satisfy specific data needs.

**Roger Avery, Brown University, Population Studies & Training Center, Demography (Faculty)**

We are working on the new disability questions. Only the census has the huge n that is necessary to study differentials in some prevalence of disability, which now is shown is significantly more detail than before. Detail in age is crucial to our research. Those 65-69 are very different from the 85+, those 5-9 are very different from those 15-19.

**Confidential Response**

Detailed-level data, by geography, is the foundation in creating the proper sample design in each of the research projects my firm undertakes. By reducing or condensing the level of detail would be detrimental in how we are able to properly sample the market under study (a majority of which are of interest to local transportation planning agencies around the country).

**Heidi Wong, Asian & Pacific Islander American Health Forum, Other Nonacademic (Nonacademic)**

Our primary level of analysis is at the individual Asian and Pacific Islander ethnic groups (e.g. Chinese, Vietnamese, Hmong, Chamorro, Samoan, etc.). PUMS is the only source from which we can do specialized analyses at this level (e.g. county by Korean by age by income by public assistance). Should PUMS detail be reduced, our ability to impact policy and decision makers around issues concerning the Asian American and Pacific Islander communities. There are many data obstacles our community faces, but in 1990, PUMS data was the one source that we could use to support our advocacy work (e.g. around health care reform, immigration reform, etc.).

**Confidential Response**

Occupational and SES detail: as the economy changes, the only way we will be able to meaningfully analyze the distribution of work and wealth in our society is to retain as much detail as possible. Currently many different theories and concepts apply to research in this area. Decreasing measures available will preclude us from being able to test existing and new theories. I would say the same about the other variables under discussion as well. At present, we have a dearth of available measures in which to test more and more complex theories. We need more not less. This is especially true for being able to measure the effectiveness of State policies on different groups in society. Please, do not limit yet further our already very limited tools in which to study some of the most pressing social, economic and political questions of our day.

**Cynthia Rogers, University of Oklahoma, Economics (Faculty)**

PUMS is essential for doing substate level analysis. The geographic detail is essential for making regional research more relevant for practical policy analysis.

**Robert Gibbs, Economic Research Service, USDA, Economics (Nonacademic)**

Much of my work requires knowledge of the industry and occupational mix of labor market areas (substate metro/nonmetro). This information is indispensable for estimating local skill levels (computed with detailed industry/occupation) and employment opportunity. For instance, estimates of the supply of jobs available to welfare recipients requires attaching average skill and educational information to detailed occupations in a local area. Broad occupational categories are useless for this analysis, since the variation in skill and education are so great within each broad grouping. Similarly, I have computed wage premia for workers in local industry clusters, again requiring detailed industry and geography in order to identify clusters (these must be done at a local labor market level for specific industries).

**Judith Stallmann, Economics (Faculty)**

I have done research on the elderly. Grouping everyone over 65 does not allow us to look at what happens as people age.

I have done work on rural labor. Geographic detail is essential for rural labor work.

I have done work on the relationship between educational achievement, dropout rates, and the local economic structure. Occupational data is essential for this work. Industry data is not as useful because it includes everyone from the CEO to the janitor.

Occupations are changing rapidly so that the more detail gives an idea of the changes that are taking place and have taken place in occupations. I know that many times researchers use industry data when they really want to use occupational data, but occupational data are only available every ten years, whereas industry data are available annually. We need more emphasis on occupational data.

**Confidential Response**

The geographic detail is essential!

**Jim Ross, CSUB, Sociology (Faculty)**

Reduction of PUMS detail would prevent much of my use of census data in my research. The detail is a necessary for my work.

**David Pearson, Texas Transportation Institute, Other Academic (Researcher)**

Reduction in detail of household information (e.g., size, income, etc.) would virtually render the data for our research as unusable and worthless. When the 1990 census only used household size categories of 1 to 4+, it created some problems in using the data for estimation purposes and in effect, we had to use estimates in some areas.

**Confidential Response**

My research focuses on urban area (ward/census tract level and up) economic base change and demography across time. Any reduction of industrial classifications or in age, race, income, education, geography, would preclude detailed analysis of across time change in economic base through industrial shift and resulting demographic migration patterns. Given the economic restructuring currently experienced in this country, this research is vital to understanding the services needs of urban areas, the employment pressures on migration etc. My research agenda would be seriously curtailed. While my focus is academic research, much of it has strong policy implications and is developed as such.

**Susan Hendricks, KJS Associates, Industry (Nonacademic)**

We look at travel behavior and location choice based on geographic location and person and household characteristics, including as age, income, and occupation.

**Michael Lahr, Rutgers University, Center for Urban Policy Research, Economics (Faculty)**

I do economic research in geographic detail. Mentioning a set number of categories, like 25, is simply not enough. Indeed, 25 may be sufficient but it would have to be something like that displayed in Appendix II of McGranahan, David A. and Linda Gelfi. 1991. "The Education Crisis and Rural Stagnation in the 1980's," Education and Rural Economic Development: Strategies for the 1990's, Agriculture and Rural Economy Division, Economic Research Service, U.S. Department of Agriculture, ERS Staff Report No. AGES 9153, pp. 40-92.

**Confidential Response**

I analyze the income and poverty status of persons, families, and households in rural or nonmetropolitan areas. If the 2000 PUMS does not contain detailed information on sources of income, poverty status, and the employment and occupational status of persons by place of residence, it will be of no use in my research.

**Jon Moen, University of Mississippi, Economics (Faculty)**

A loss of age detail will produce truncation and censoring problems for most empirical work, rendering research into retirement useless.

**Lionel Beaulieu, Southern Rural Development Center, Sociology (Faculty)**

My work focuses primarily on human capital and workforce development issues. The PUMS data allow us to examine these critical issues across various spatial areas and labor market areas. The loss of specificity in the PUMS data would seriously compromise our ability to understand the unique dynamics associated with urban/rural areas in various LMAs.

**Tom Trozzolo, University of Notre Dame, Sociology (Researcher)**

Any attempt at comparability between the "youngest old" (65-74), "middle old" (75-84), and "oldest old" (85+) becomes impossible without birth year information remaining as is.

**Confidential Response**

Since I assist an area agency on aging to utilize major population data on trends for providing services needed by the growing Aging cohorts, any major deletions would be critical. For example, we cannot accept a lump sum category like ages 65 & over. We MUST have more definite groupings to meet different trends that we know have been in progress over the last two censuses. Age groupings, economic groupings, health status and other changes within counties are REQUIRED to meet both service needs and planning requirements.

**Marie Bousfield, City of Chicago, Government (Nonacademic)**

To create a forecast of population and employment as required by ISTEA and the Clean Air Act, we used PUMAS as basic units of geography. The model we use (DRAM-EMPAL created by Professor Putnam) requires input that is not available from the Bureau's prepared tables. The Chicago City Council has approved several programs for middle-income homeowners and has developed guidelines for scattered sites for public housing. These require custom-made tables generated from the PUMS file. Q9 reformulated: Do you want the electric chair or the guillotine.

**Rose Martin, New Jersey Department Of Health, Government (Nonacademic)**

The population over the age of 65 is growing rapidly and is projected to continue this growth well into this century, in view of the increasing life expectancy. In terms of health indicators, we have found major differences between the "young" elderly (e.g., 65-84 years of age) and the "older" elderly (85+ years). If we do not have estimates of the population beyond age 65 broken down into ten or twenty year age groups, we can not assess health status in these groups or plan health programs to improve the quality of life for this population.

**Pamela Perlich, Bureau of Economic and Business Research, Economics (Researcher)**

In our economic and demographic modeling, age detail in the Census is our only cross-sectional data that can be cross classified with other characteristics. It is the most basic input to our modeling work.

**Confidential Response**

I do housing research and all of the suggested variables are important.

**Confidential Response**

My research on changing occupational sex composition across time would be absolutely impossible without occupation data at the detailed occupation level. I have a dataset of occupational characteristics aggregated from the 1970, 1980, and 1990 PUMS datasets. Updating it further with 2000 data would be impossible with any loss of detail.

**Arthur Wittman, Broward County School Board, Government (Nonacademic)**

I am the demographer for a large school district and ideally I would like even a smaller level of geography than 100,000. So making it 250,000 or even higher is a move in the wrong direction, in my opinion.

The most need piece is 0-4 years with cross tabs - ideally at the Census Tract level - but no more aggregated than 100,000 population.

**Confidential Response**

Teaching requires detailed information about historical times so students can understand what types of occupations were available and how they might have been gendered. For example, in Family Problems and Social Change, I lecture on what percentage of jobs women and men occupied and so forth. All demographic information is pertinent.

**Donald Hirasuna, House Research Department, Economics (Nonacademic)**

We are asked a variety of questions regarding different demographic groups in different regions. For example, I have been asked about the distribution of occupations across geographic regions. Another example is to determine the number of single parents by real income categories for different regions across the state.

The geographic, age and income detail is also indispensable for monitoring policy issues such as access to jobs in central cities. Without the current detail of the PUMS data, much research cannot be properly conducted.

**Alexandra Hill, University of Alaska Anchorage, Economics (Researcher)**

Geography: Currently Alaska has 2 PUMAs in Anchorage (101 and 102), which I always combine; PUMA 200 that can best be categorized as "other non-bush" and PUMA 300 that is very rural. If we were to lose PUMA 300 (which is the smallest in population) our ability to do any urban/rural comparisons or to look at rural Alaska populations would be eliminated. Combining 200 and 300 would cause Fairbanks and Juneau to overwhelm the rural data.

Race: separating Eskimo, Aleut, Athabaskan, etc is often the only way we have of using PUMS data to look at problems in one or another area of rural Alaska. Western Alaska has a very different economy and culture than northern or southeast Alaska. These racial designations allow us to look at how differences in these regional cultures (differences that often persist in urban dwellers) affect policy questions.

**Aileen Duldulao, Other Nonacademic (Graduate)**

I'm involved in projects that document different characteristics of the Asian and Pacific Islander community in Los Angeles. In Los Angeles County alone there are over 25 distinct Asian and Pacific Islander identities with over 40 languages spoken. In my research, which focuses on domestic violence, I already find a dearth of quantitative research that uses the category of "Asian and Pacific Islander" and instead lumps this community into "Other." Even when research or data does separate Asians and Pacific Islanders into a distinct category, this does not capture the diversity of this racial group. Wide variances exist within this community between ethnic groups. For example, Japanese have the highest per capita income of all ethnic groups in Los Angeles county, while Cambodians and Laotians have the lowest per capita income in the county. To group these two groups, who have drastically different languages, histories, and socioeconomic statuses, in one monolithic group is a disservice to these communities, particularly communities in need of services. Greater detail in census data would allow social service providers like myself to make more informed decisions regarding what communities are in greatest need for our services. Without this detail, especially breakdowns in race and geographic location, many communities and families will go without services they so desperately need.

**James Kinard, Government (Post-doc)**

In social policy determination, critique, analysis, etc. details are extremely important for gaining insight into the quantitative and qualitative aspects of populations with different characteristics, needs, and so forth. The "lumpen" achieved during data analysis almost invariably hides important population groups. Less detailed data hide such populations even further. Perhaps more important is the fact that the unavailability of less detailed data will leave researchers, policy makers, advocates, etc. without the data needed to verify general conclusions and/or assumptions. Many populations would be excluded from the benefits of policy-making processes because they are just not accounted for.

For instance; there is a serious problem in service-delivery systems because people with multiple disabilities practically have no where to go. Persons with multiple disabilities are forced into disability-specific programs where they don't get adequate treatments. If they have vision disabilities, they must go into programs designed for the blind; if they have emotional disabilities, they must go mental health programs designed for the mentally ill; if they have mobility disabilities, they must go to programs designed for people with mobility disabilities; and so on. This population is forced to identify such resources on their own and try to determine which ones are more appropriate for their different disabilities. They receive no professional help in this time-consuming and frustrating process.

The patient wastes invaluable time needed to receive the services that could optimize his/her rehabilitation; while programs like Medicaid, Vocational Rehabilitation, and others, make inefficient use of their resources by haphazardly providing services to this population, or by not addressing all their needs properly.

More detailed data is needed to know about those populations that suffer paralysis and blindness, emotional problems, and other disabilities. The same applies to the poor that need shelter AND medical help AND employment AND food AND economic aid, etc.

In conclusion, more detailed information is needed to know what is going out there in the real world and to give analysts what they need to design efficient and humane programs.

**Hans D. Laux, Geography Department, University of Bonn, Germany, Geography (Faculty)**

My research as a social and population geographer is mainly concentrated on the aspects of spatial and social integration of racial and ethnic minorities in US American cities. Detailed information about ethnic and racial categories as well as about demographic and socio-economic characteristics is indispensable. Moreover, I need information on a very detailed geographic level. And last but not least, intercensal comparability is requested.

**Confidential Response**

Most of my work focuses on inner-city, economic development. Statistics on age and income are critical to assessing the community dynamics in a given neighborhood. In addition, these data are necessary to identify "qualifying recipients" for various federal programs that fund inner-city economic development, especially HUD's Community Development Block Grant program -- funded agencies need to show that the majority of their clientele is "low or moderate income" (a number which is geographically specific . . . something like 75% of median household income for low income, 50% for moderate income). Geography is also critical as regions (NE, South, Mid-West, West, etc) vary dramatically in their income and age statistics, but even within these regions there is great variance in statistics (e.g., Atlanta, Charlotte, and Memphis could have very different age populations). Thus, the smaller the region for PUMS data, the more specific and appropriate the analysis is from a policy perspective.

**Confidential Response**

I study the determination of wages and the impact of discrimination on wages and occupational attainment. Reducing the detail in the age, occupation, income, race, and geographic variables (and in the ancestry variables to some extent) would greatly reduce the precision of studies of this kind. It would make it very difficult to compare findings across time, which is essential in analyzing the impact of policy and the need for further policy development.

**Confidential Response**

I am particularly concerned about any reduction in geographic detail as well as grouping of any other data. Greater geographic detail is vital to understanding the spatial patterns and behavioral motivations of everyday human activity. My own research looks at intermetropolitan migration by race, gender and the influence of labor market dynamics. You lose so much detailed information of market structures at the state level - labor markets operate at the metropolitan scale. Distinguishing by race, gender and age are paramount since each sub-category displays unique patterns and motivations. Loss of data detail would be a step backwards in the sort of research I conduct.

**Richard Ogburn, South Florida Regional Planning Council, Government (Nonacademic)**

The extraordinary age, race and ethnic diversity in South Florida, and the pace of change, make it critical to be able to carry out socio-economic analysis at the greatest level of subject and geographic detail consistent with the preservation of confidentiality. Much analysis depends on PUMS data because published tabulations do not cover the full range of need.

**Marianne Hill, Mississippi Center for Policy Research, Economics (Nonacademic)**

Maybe 1/3 of women on welfare here in MS live with someone else, based on PUMS data: approximated as females under 40 with own child present not head of household and receiving public assistance. I also look at occupational segregation by race and sex--black and white female childcare workers are one concern, but generally want to know if the Duncan index has improved or not. Need detail on ages of children under 18, and each year from 1 to 6 for childcare. Also projecting occupational demand, detail on occupations is critical, though I welcome the replacement of some old occupations with new.

**Carolyn Lauer, Michigan Information Center/DMB, Government**

PUMS, as they have been released in the past, have not provided the level of detail we require to provide service to our clients

**Marilyn Cavell, Virginia Tech, Other Academic (Faculty)**

We use PUMS data to create more detailed demographic estimates (generally at the county level) than are available through STF tables.

**Alicia Sasser, Harvard University, Economics (Graduate)**

As a labor economist, my central focus is analyzing data on income. However, in order to analyze issues such as the gender gap one needs to be able to control for age, race, ethnicity, occupation, geographic region in order to determine whether income differences stem from differences in the characteristics of men and women in the labor force or from differences in the returns that men and women earn for having these characteristics - being a certain age, race or ethnicity, working in certain occupations or regions. Without such detail one cannot know whether gender differences in income could be alleviated by encouraging women to enter certain occupations or by continuing a policy of affirmative action.

**George Hough, Portland State University, Demography (Faculty)**

With the new Disability Questions, age is of utmost importance. I utilize this data to provide information necessary for policy planning for a number of aging and disability service agencies.

**Deborah Garvey, Santa Clara University, Economics (Faculty)**

The PUMS is the only nationally representative database for demographic and economic policy research on immigrant incorporation and assimilation into the United States, one of my current research areas. The PUMS is the single most important data source for my work. Significant reduction in detail would render it much less useful, and possibly completely useless, for my work.

Loss of geographic detail would make it impossible to provide estimates of short-run and lifecycle fiscal impacts of immigrants on state and local governments. It would be more difficult to track migration patterns and identify the jurisdiction where an individual lives. Loss of income detail beyond the current topcoding would also make fiscal impact estimates less precise and potentially very misleading, due to an inability to calculate tax remittances and government benefit receipt with reasonable precision. Age groupings beyond the current classification will make it difficult to track changes in economic behavior (e.g., labor force participation or government transfer receipt) over the life course.

Of course, all these problems would be compounded with the loss of comparability to previous PUMS data sets.

**Glenn Fuguitt, University of Wisconsin-Madison, Sociology (Faculty)**

I make a lot of comparisons between metropolitan and nonmetropolitan areas for smaller subregions of the country. Its important to me that these distinctions be maintained. Also, I consider elderly migration, and wish to be able to divide the 65 and over category into at least two groups.

**Philip Wuest, Clark County, Washington, Government (Graduate)**

I deal with transportation and development models. Location and specific characteristics associated with that location are everything. Without location and detailed information assigned to that location, it is extremely difficult to build useful and meaningful public policy analysis models/tools.

**Sue Wong, National Economic Development & Law Center, Other Nonacademic (Nonacademic)**

We are a national intermediary that works with local nonprofits to develop workforce development strategies to link low-income residents to high growth, high wage occupations. We need specific workforce profiles for specific targeted occupations and to demographic information of our target population (e.g. low-income youth) to inform the development of an effective workforce strategy. Factors such as age, race, occupations, and geography and their current breakdowns (1990) allow us to compare and contrast the target population's demographics with the existing workforce to determine how far apart or similar these two labor forces are which, in turn, will determine if the employment linkage strategy should target the occupation in question. Furthermore, the information will be used to develop the specific workforce strategies to include components such as basic remediation, technical skills training, diversity issues and/or transportation modes, for example.

**Janette Kawachi, University of California, Santa Barbara, Sociology (Graduate)**

My research focuses generally upon issues of labor and economy in both a domestic and international context. I focus particularly upon issues dealing with inequality and urban poverty. Much of my work would be impeded if they were to reduce data collection on variables like industry, occupation, income, wage, race, age, sex, housing, education, socioeconomic status, geography, etc. I depend upon IPUMS data as a reliable and comprehensive source of information since census aggregation techniques for city or state level data is extremely rough and untrustworthy.

**Diane Muehl, University of Illinois at Urbana-Champaign, Sociology (Graduate)**

I am doing research in the area of older immigrants and need maximal details of age, income, race, ancestry, etc. As my teenager just told me when I explained your questionnaire to him: BAD!! BAD!! BAD!! I concur.

**Lori Hunter, University of Colorado at Boulder (as of summer, 2000), Sociology (Faculty)**

My research entails linkages between geographic context and demographic processes, therefore the loss of geographic detail in the PUMS would render these data unsuitable for this type of research.

**Judy Olson, Michigan State University, Geography (Faculty)**

I have been working with a student interested in ancestry, the interest having developed as a result of some work in another of my classes in which less detailed data were used for mapping. It is early to say where our efforts will go upon completion of the current project.

**Stephanie Kellner, National Economic Development and Law Center, Other Nonacademic (Nonacademic)**

My organization works to revitalize low-income communities. We use PUMS data to develop a demographic profile of specific communities. We require a fairly small level of geographic detail (i.e. counties and smaller) so we use PUMA data. If PUMS data is grouped in larger geographic areas, the data becomes much less useful. Also, as part of our work, we develop occupational profiles of the workforce in our target area (for example, if we are looking at nursing assistants, we develop crosstabs to show the median age of nursing assistants in our target area, race breakdown, the number of years of schooling, how they get to work, etc.).

**John Foster-Bey, The Urban Institute, Other Nonacademic (Nonacademic)**

I do a good deal of research examining the differences across metropolitan areas by race, age, gender, industry and occupation.

**Confidential Response**

Reducing the number and quality of variables available would destroy the possibility of testing previous researchers' work to determine if their results are still valid for more current data.

**Confidential Response**

I need to identify children age 2 years in my work. I need counts of children by various characteristics for many small geographic areas.

**William Mead, Gloucester County New Jersey Planning Division, Demography (Nonacademic)**

Primarily in the allocation of financial resources necessary for support services affecting TANF recipients, i.e., Child Care Planning. It would also affect analyses regarding the distribution of resources to meet labor force maximization objectives within age and income categories.

**Kenneth Mitchell, University of Minnesota, History (Graduate)**

I would be lost without PUMS data (hopelessly lost...)

**Jeremy Atack, Vanderbilt University, Economics (Faculty)**

My work is primarily historical and I think that the Census needs also to consider the permanent long-term impact of these decisions on research when issues of confidentiality have become irrelevant. This decision has long-term consequences.

**Kate Kelly, University of Central Arkansas, Other Academic**

My research will be dealing with children, those of day care age, and those in each grade level age. To be most useful, it is necessary to know how many children are in what grade or if they are in day care or care for at home.

I also intend to work with the aging demographics not only for today, but also for 30 years from now when the baby boomer generation begins to die.

## **Responses to Question 12: If you have any thoughts or suggestions on issues of privacy or confidentiality in the PUMS, please elaborate.**

### **Confidential Response**

I certainly have never tried to identify individuals using the PUMS, but I can't even conceive of how it could be done. Maximizing the response rate is certainly an important goal; and ensuring confidentiality probably helps reach it. But to collect a lot of data that no one can use ... that seems like a waste.

### **Alair MacLean, University of Wisconsin - Madison, Sociology (Graduate Student)**

I can't really imagine how people would use the PUMS to get access to information about specific people. There are so many more real threats to privacy embedded in the proliferation of online databases accessible through the web.

### **David Autor, Massachusetts Institute of Technology, Economics (Faculty)**

I would be happy to sign a confidentiality form that declares under penalty of law that I would not use PUMS data to identify individuals. Though in fact I don't believe the data could be used for this purpose, I would be happy to put in writing that I would not attempt to do so even if I could. Assuming completing such paperwork would not turn into a major time delay, perhaps other researchers would agree to this stipulation if it would facilitate the Census Bureau in providing greater data detail.

### **Confidential Response**

There is no information in the PUMS that I can conceive of anyone wanting for inappropriate reasons which could not be found more easily from credit reporting agencies and other administrative sources. Because of its size and comprehensive nature, the Census data allows researchers to approach many questions which are otherwise inaccessible. Its scope should be increased, not limited.

### **Dowell Myers, University of Southern California, Other Academic (Faculty)**

The added emphasis on confidentiality is overblown. Someone would have to work VERY hard to reveal odd cases that stand out in the data. Given that only 1 in 20 cases will be released, census data are a much less effective means of discovering private information about specific individuals than recourse to privately maintained databases that even include exact names and addresses.

### **Confidential Response**

Given that there has never been a breach of confidentiality privilege in the use of PUMS, why is it necessary to make any changes?

### **Jesse Rothstein, University of California, Berkeley, Economics (Graduate Student)**

I'm not aware of \_any\_ instance in which PUMS data has been used to violate someone's confidentiality. If one were interested in gathering information on a particular individual, more detailed information could be obtained from commercially available marketing databases, at far lower expense and difficulty. Concerns about the confidentiality of PUMS data are vastly overstated.

### **Steve Doig, Arizona State University, Media (Faculty)**

This move to limit detail sounds like a knee-jerk reaction to the pseudo-fuss last month over "invasive" long-form questions. You might ask if anyone has produced any evidence that anyone was harmed by the 1990 PUMS.

### **Charles Haynes, The University of Texas at Austin, Demography (Graduate Student)**

It seems to me that the sheer volume of data in the PUMS provides substantial protection for respondents, except in the smallest levels (census tract, block).

### **Confidential Response**

Provide a more detailed set of Census data available to academic researchers willing to follow data security/confidentiality guidelines established by the government. I work with such data on a daily basis. I agree that a more detailed dataset should be kept out of the hands of market researchers.

### **Dan Devroye, Harvard University, Economics (Graduate Student)**

I know of no instance where individual privacy was compromised by researchers using large data sets. Indeed, for larger samples it seems even less likely that the data can be used to determine an individual's identity.

In any case, these data are the primary source of knowledge in social science and in designing public policy. Whatever the costs of releasing detailed information, the benefits are far greater.

**Matthew Sobek, University of California, Riverside, History (Researcher)**

Should it be necessary, I think top-coding is less destructive in most cases than combining values or categories. Researchers require flexibility with the categories, since one cannot feasibly predict the range of applications to which they might want to put these data.

**Confidential Response**

I can't see how (let alone why) a researcher could identify a particular person based on the PUMS variables. But IF this were possible, the identified person would have to be the only person with a particular constellation of variables (e.g., occupation, race, ethnicity, age, AND sex category) in a particular detailed geographic location. Why not make data on detailed geographic location subject to screening (i.e., it is only released to researchers who have submitted a request and passed some minimum eligibility requirements) and leave the rest public? Or, alternatively, one could program the extraction programs so that they only allow a certain number of potentially identifying variables to be extracted in one file (a rule also subject to exceptions based on a screening process).

**Richard Campbell, University of Illinois at Chicago, Sociology (Faculty)**

The issue is disclosure by whom? If the data were restricted to the research community, then we should expect that local IRB's and professional codes of ethics (plus simple common sense) would preclude malicious attempts at identification. Accidental disclosure has always struck me as a non-issue. The real issue is malicious disclosure by non-professionals. Since the data are, effectively, available to anyone, it is conceivable that someone might seek to actively identify a subject, either to demonstrate that it can be done or in order to target a specific individual. I don't doubt that a reasonably astute high school student who really wanted to do so could ID someone. Only a few people with unusual characteristics would be vulnerable to identification, but it could happen. So, a way has to be found to allow legitimate users to access the data. Census enclaves are not the answer. I would prefer some system of registering users after they signed a confidentiality pledge and/or established a legitimate need for the data.

**April Eaton, University of Washington, Sociology (Graduate Student)**

I have never even heard of an academic or other researcher using or wanting to use PUMS data to identify specific individuals. I think this concern is ridiculous.

**Peter Meyer, Northwestern University, Economics (Graduate Student)**

It does not seem costly to me to keep some precise geographical identifiers confidential for a few years after the release of the data in general, though I still hope narrow geographical identifiers are available eventually. It also does not seem costly to mask very high incomes. These seem like areas where public-use could make some compromise to achieve more convincing privacy results.

**Kim Rueben, Public Policy Institute of California, Economics (Nonacademic)**

In dealing with numerous confidential data sources and interacting with other researchers, I do not know of an instance when people tried to identify individuals.

I find that those who I work with and other researchers are the most excited to receive and eagerly fill out Census long forms. I readily filled out a copy of the American Communities survey and was disappointed when I realized it was not an ongoing survey (like the CPS). If there were serious confidentiality problems, I would think economists and other researchers who use these data and know others who use the data would be more reluctant to fill out these forms as it is their peers who would gain access to this information. Instead they are the most eager.

**Katherine Condon, Florida International University, Other Academic (Faculty)**

While I believe it is "plausible" for some unscrupulous individual to identify the characteristics of specific individuals ... I believe that most social science researchers who use this data are not unscrupulous. Maybe users of PUMS data at the very smallest geographical units should be required to have on record a signed (not just e-mailed) promise of privacy and confidentiality in their usage of the data on the same lines that HCFA requires for a researcher to use Medicare data.

**Lisel Blash, Public Research Institute/San Francisco State University, Other Academic (Researcher)**

I really do not see how you could identify individuals using PUMS data. The geographic areas are so broad, one would be hard-pressed to find an individual from the data, even if that individual were quite unique.

**Frank Braconi, Citizens Housing & Planninc Council & New York University, Economics (Faculty)**

I don't think there is a legitimate issue. There are many serious threats to Americans' privacy; this is not one of them. It would be a major setback for social science research in the Unites States.

**Mark Smith, The MEDSTAT Group, Economics (Nonacademic)**

Top-coding of income and the lack of highly specific geographic information prevents any breach of confidentiality. Reducing the level of detail on important variables will limit future research while bringing no gain in confidentiality.

**Darius Lakdawalla, RAND, Economics (Researcher)**

I suspect that the people who use PUMS use it because it is a large dataset with many variables, at least this is true for me. Anyone working with a large dataset has neither the time, the inclination, nor the computing resources to identify specific individuals.

**Joseph Doherty, UCLA School of Law, Other Academic (Researcher)**

I have never been inclined to break the confidentiality of respondents, and I don't know anyone who has. In addition, there is little incentive to violate privacy if one wanted to; the current level of aggregation makes it unlikely that any attempt to identify a respondent will be fruitful.

**Confidential Response**

In the unlikely scenario that someone knows enough about someone else to identify them through IPUMS it is highly unlikely that there is anything useful left to find out about them.

Of course, one possibility is to take a STF approach... only provide files with certain combinations of variables. To be meaningful there would have to be a common set of variables to each file, which could be used to find potential matches that could reconstruct the original PUMS file, but with multiple possible matches across files it add another level of anonymity.

**Confidential Response**

Within a population of 100,000 people, if you have the information to try to track someone down from PUMS, then you've already got all that information about the person -- you're not getting it from PUMS. You can't take a few variables that you might know about someone (age, race, sex, occupation) and correctly identify that person from the sample. It's true that if I filled out a long form and it was included in the PUMS sample, I might be able to go through them and find myself, but that's true no matter how big the sample size or how few variables are released.

**FlorenceMae Waldron, University of Minnesota, History (Graduate Student)**

This concern strikes me as utterly irrelevant and completely reactionary in its origins. To the best of my knowledge, 2000 is the first census year in which there has been so much media coverage devoted to any sort of privacy concerns related to census data. If not for said media coverage, I doubt this would even be an issue. There is absolutely no reason for individuals to be concerned about their personal privacy being violated by the release of detailed census data, because of the amount and type of census material subject to 72-year confidentiality rules. My own work with the IPUMS suggests that anyone who wanted to track down a specific individual's confidential census responses, via the 2000 PUMS, would have to undertake an extremely long, extremely difficult, extremely laborious, and ultimately unsuccessful process.

**Tom Rex, Arizona State University, Economics (Researcher)**

How can a 1% or 5% sample of areas of 100,000 people possibly violate privacy or confidentiality?

**Confidential Response**

Has there ever been a case of anyone violating privacy through PUMS? What is the plausible scenario?

**Dan Black, Syracuse University, Economics (Faculty)**

Trying to identify individuals from the PUMS -- given the enormous safeguards that are built into the product currently -- is silly. This threatens to render the PUMS useless to the research community and threatens the very existence of the long form data.

**Scott Drewianka, University of Chicago, Economics (Post-doc)**

I for one--and I am confident that this goes for everyone in my field--have no interest whatsoever in linking lines of data with real persons in the world. In fact, I have no idea whether such a thing is possible. My primary interests lie in constructing aggregates (though not always in ways that the Census Bureau foresees) and in examining broad patterns prevalent in large volumes of data.

**Rebecca Sandefur, University of Chicago, Sociology (Graduate Student)**

Some ideas:

(1) FIRST CHOICE: Ask users of the data to sign affidavits asserting the data will not be used to identify individuals. If you want to have a block against frivolous applications to use the data, require that the affidavits be notarized, and set up a password system for access to the data. This still allows open access (though with a small fee barrier) while discouraging the behavior folks seem worried about.

(2) IF NECESSARY: Archive the data in such a way that specific extracts cannot include enough information to identify individuals. For example, a study of income variation within detailed occupation categories does not need to be able to identify the block on which a respondent lives. Deciding exactly how to do this would require serious study.

**Confidential Response**

The amount of effort needed to get any useful information on individuals from the PUMS is prohibitive to make it a major concern at present. Anyone who wanted to obtain such information would find it more easily through normal investigative channels.

**J. David Hacker, California Institute of Technology, History (Faculty)**

Obviously, there are issues of privacy and confidentiality related to the release of recent census data. The current practice of removing names, local geographic information, and other details in the PUMS, however, has proven to be more than adequate measures to protect individuals' privacy. This loss to detail already comes with a significant cost to social research. Further loss will be very harmful to social research without significantly improving the level of individual privacy.

**David Clark, Marquette University, Economics (Researcher)**

The current definition of PUMA's effectively prevents the user from identifying individuals. Clearly more spatial resolution would be great, but I recognize the reasons for limiting the scope of geography to PUMA's. However, to further reduce that detail would greatly diminish the value of this data, especially for those doing research using multiple years of PUMS data.

**Patricia Kelly Hall, University of Minnesota, History (Graduate Student)**

Is it possible to identify an individual with extremely unique characteristics? Probably. Is it plausible that anyone using the PUMS would bother? Most likely not. What would be the motive? Who would take the time? Researchers who use the PUMS for academic or economic gain are unlikely to bother looking at an individual case. While we want access to individual level data, it is the ANALYSIS of these data that is of interest to us. No one has time to go looking through individual records. And malicious evil-doers have ready access to datasets with far more trouble-making potential than the PUMS.

**Confidential Response**

We can get the data only of anonymous, random sample. So it is impossible to detect the specific persons data from PUMS. I don't think PUMS put some danger on privacy of replicant.

**David Burruss, University of Kansas, Economics (Researcher)**

The only time I can imagine a real payoff from trying to identify individuals that is high enough to justify the cost and likelihood of failure would in the case of the very rich. Suppression of very high income categories is an adequate defense.

**Susan Carter, University of California, Riverside, Economics (Faculty)**

I believe that certain people in the PUMS might have their identities revealed. It is a very tiny fraction of all individuals whose records are included in the PUMS. These individuals, in turn, are a very tiny fraction of all individuals in the census. Against this cost is the cost of limiting research by restricting the detail of the categories. My own view is that the social costs of the latter outweigh the social costs of the former. I urge Census to maintain the detailed categories in the PUMS.

**Lawrence Nitz, University of Hawaii at Manoa, Political Science, Other Academic (Faculty)**

One of the factors we often forget is that data ages rapidly. In urban areas there may be a turnover of 30% annually in telephone numbers and residential addresses. Even if the turnover were only 15% to 20% within a few years many neighborhoods will have different composition. Thus, even block and block group statistics will report on people no longer there.

But how does one find a single person in the 100,000 members of a PUMA. Almost all of the social and economic categories have been compressed to some extent. Existing Census practices are reported to re-allocate information that would refer to only a very tiny number of people in a PUMA. Thus, I expect that I will not be able to identify a physician on Molokai, because there is a chance he has been replaced by a lawyer, if only to hide the identity of the few physicians on the island. (Molokai, of course, cannot be separated from the PUMA which contains Maui County and Kauai.)

**Joelle Saad-Lessler, Columbia University, Economics (Graduate Student)**

The way the data is currently set up, it is impossible to find out information on any one person. The smallest level of detail is a set of average characteristics that are available by zip code. This already makes life difficult for some researchers, but at least a certain level of confidentiality is required. However, there is no need for aggregating the data further for purposes of enhancing privacy, as privacy matters are already being taken into account. Besides, there is so much data available out there through marketing companies that could really endanger privacy, and that information contains so much more detail. I just don't see how the census is providing any more detail than that.

**Leora Friedberg, University of California, San Diego, Economics (Faculty)**

Please give us more data, not less! Good data helps make good policy.

**Patricia Oslund, University of Kansas, Economics (Researcher)**

As I understand it, some variables are already top-coded. This probably provides a lot of confidentiality already. In other cases (as I understand it), data items are already "fuzzed" by blanking out true values and imputing values instead. Maybe more fuzzing could be used for individuals in ethnic or other groups that have a small representation in a PUMA.

**Reeve Vanneman, University of Maryland, Sociology (Faculty)**

If this is inevitable, then we must do more to make census data centers more easily accessible to legitimate research.

But I cannot believe this is necessary. What imagined problems will end up destroying real research possibilities?

**Stephanie Aaronson, Columbia University, Economics (Graduate Student)**

Honestly I think the whole thing is ridiculous, particularly since earnings is top-coded. What we are talking about is identifying a person by knowing that he is a male laborer in the textiles industry who immigrated to the New York City from the Dominican Republic 15 years ago with his wife who is a housekeeper and his two kids. Please.

The only way someone might be able to be identified would be to combine this data with the very detailed geographic data which is already restricted in use. Even then, I think that the odds of being to identify individuals is pretty slim.

Actually, I think there is a lot of confusion on the topic. The concern among the public (and republican congress people) seems to be that the government has information about them. But the government will have that information no matter what is released to academics. What academics get just isn't useful for that purpose.

**Confidential Response**

Confidentiality of responses is a legitimate concern. However, bona fide researchers do not pose a serious threat: have better things to do with their time. Researchers are interested in the overall patterns, not in identifying individuals.

**Charles Wetherell, University of California, Riverside, History (Faculty)**

Individuals would be extremely hard, if not impossible, to identify using the PUMS, even in a 5% sample

**Gunnar Almgren, University of Washington, Demography (Faculty)**

While I think it may be possible to link PUMS data to individuals, the probability of error relative to the effort would preclude this being at all useful as a tool. I believe there are easier ways to get information about individuals that are legal.

**Marcus Stanley, Harvard University, Economics (Graduate Student)**

It seems to me this is a political stampede without much reason behind it. I know there have been a few people trying to whip up concern about the long form, but I don't know of any case where Census data has been used to identify individuals. One would need incredible resources to do this; it is just not a reasonable thing to try to do.

**Laurie Smith, University of Nevada, Las Vegas, Other Academic (Faculty)**

I think confidentiality has to be assured or eventually this is going to harm the data set as much as reducing the data. However, if the reduction could be made very carefully so that it occurs only in cases where there is a real threat to confidentiality, then that needs to be done.

**Albert Monroe, Harvard University, Economics (Graduate Student)**

It is absolutely essential that the Census Bureau maintain the level of geographic detail (states and metropolitan areas) used in the current data. The current use of geographic identifiers is more than enough to maintain confidentiality. Any less, and critical questions that involve differences over space in the United States may go unanswered.

**Confidential Response**

Privacy is an overworked slogan. Strip identifying data and outliers and the rest is moot.

**David Katz, Loyola University Chicago, Sociology (Graduate Student)**

I do not believe that the PUMS in any way threaten the confidentiality or privacy of Census data. The concerns seem to focus on highly implausible "what if" scenarios rather than on actual human activity.

**Carolyn Liebler, University of Wisconsin at Madison (sociology & demography), Sociology (Graduate Student)**

I don't think people misuse the PUMS data, if that's what you mean. I do think the PUMS data is useful at the individual level; I care about individuals' characteristics as opposed to characteristics of a group.

**Confidential Response**

I am increasingly concerned about efforts to limit availability of high quality data such as PUMS to a broad spectrum of social scientists and other users, and to eliminate important information out of concern about privacy violations. The chances of such violations appear trivial. Insisting on secure-access sites is not a very appealing solution because doing so greatly disadvantages persons not located near such facilities. Restricting access or eliminating important information does not serve the broader purposes collecting such data - better understanding of social conditions and problems, information about the effect of public policies on social conditions, and insights into how to improve matters.

**Mark Ellis, University of Washington, Geography (Faculty)**

I understand the concern and am appalled by the loss of privacy in the age of linked computer records and the internet. But it seems quixotic to worry about privacy in Census data when marketers and geodemographic companies combine to know more and more about individuals. I worry more about my medical records being in the hands of my employer.

**Confidential Response**

Possibly have restricted use variables, similar to the PSID and the NLSY.

**John Sandberg, The University of Michigan, Demography (Graduate Student)**

Although I don't use them as much now, I've worked extensively with PUMS data, and I sincerely believe that identification of respondents is so difficult as it stands now, and possible for only such a microscopic proportion of respondents, that confidentiality is already at the highest level possible. An analogy that springs to mind is that of the proverbial needle in the haystack; even if you made your living selling needles, the returns to searching the haystack would be so low as to make it a worthless endeavor. I also believe that the definition of the PUMA scale could be lowered substantially before any posing any threat to confidentiality. We are already at the top of an asymptotic confidentiality curve-any further effort would not increase confidentiality considerably, but would cost dearly in valuable data sacrificed in the effort

**Confidential Response**

The Census may make available to files: one for the general public, including researchers; and another more detailed files for researchers or any other person/organization which want more detailed information. Clearance should be required to use the detailed file. The clearance may be process through data organizations like yours.

**Confidential Response**

I am sure legitimate researchers will not divulge results such as to reveal the personal information of individuals. If the fear is that researchers do this unknowingly, perhaps the Department of Human Subjects at each institution should expend more efforts in disseminating rules concerning privacy.

**David Belkin, City of New York Independent Budget Office, Government (Nonacademic)**

I regard it as a bogus issue.

**Steve Holden, Johns Hopkins University, Economics (Graduate Student)**

An agreement can be set up between institutions and the bureau. This is what BLS does with the NLSY geocode data.

**Ernest Schreiber, Lancaster New Era, Media (Nonacademic)**

It is difficult to comprehend in what way the PUMS data might be employed to identify a specific individual, or what motive any researcher/journalist would have for doing so. Even if the rare case can be made -- identification of the lone multimillionaire teenager in such-and-such a region -- it is difficult to see that the harm of that potential identification would outweigh the immense benefits of maximizing PUMS data.

**Cristian deRitis, Johns Hopkins University, Economics (Graduate Student)**

In terms of delayed release of PUMS data if more detailed categories are included, I believe that the research community is generally accustomed to lagged data releases. Lags are preferable to quick releases if such releases need to be revised later on. With confidentiality measures in place and lags in release, I do not believe that privacy is being compromised.

I believe other resources such as the Current Population Survey (CPS) offer sufficient detail for many purposes requiring more timely data. The PUMS offer benefits in terms of detail and sample size that other surveys do not.

**Confidential Response**

If someone is traceable (which I do not believe) because of the PUMS than it is in parts because of many other (unnecessary) information given, such as the number of cars someone owns.

**Confidential Response**

PUMS have been around for years. Has there ever been a breach of confidentiality.

**Anupam Tyagi, University of Utah, Economics (Graduate Student)**

\*\*\* For me the confidentiality of publicly and privately collected information and violation of privacy is a real and serious concern, and so is freedom of information.

\*\*\* In this discussion about privacy a distinction should be made between \*\*privacy concerns with census data\*\* and \*\*privacy concerns with non-census data combined with census data\*\*. Census should be held responsible only about the former; that is, information that it provides. For the latter, if there is a concern, restrictions on non-census information collectors are also called for, at the same time when such restrictions on the release of census information are being considered.

Following are some suggestions:

1. Only the information in addition to that was released from the 1990 census is should be a cause of concern, since a similar informed decision on privacy and confidentiality must have been made after 1990 census. Please try and maintain historical compatibility, at least with the 1990 census.
2. You could release all the details for large areas on the IPUMS site, since this is unlikely to be a concern.
3. If there is a concern for privacy in a sample including an specific subgroup (for example, a small occupational group in a small areas), then present the researcher with a trade-off among the details, that can be chosen from to make the sample \*\*concern free\*\*. Keep a robust margin for error, since privacy is a serious concern. This should be done on a case by case basis-- using real-time decisions by the IPUMS computers. Not all researchers will have something in their sample that raises concerns for privacy.
4. In some less sensitive cases where there is still a cause for concern, a policy of confidentiality agreements should be formulated. It would be best if third parties handle this and leave the census bureau alone--- since there is still a chance that something may go wrong, people will too concerned and the census response rate next time will hit the floor.
5. Census should look for precedents about privacy violations--- what has been the extent of such violations in the past? How do private information holders compare in misuse and release of information (dot coms, financial companies, credit card companies)?

**Confidential Response**

People who are concerned about confidentiality seem to forget that marketers already know far more about them through credit agencies, shopping cards, etc. The concern about PUMS is just another opportunity for irritated people to sound off. The personal information that could be learned from PUMS is miniscule when compared to what is already available through private sources.

**Confidential Response**

Perhaps statistical methods might be used to identify relatively unusual individuals in small geographic areas (e.g., Zip codes) and cloud (alter, restrict, obscure, etc.) all available data to preclude identification of such individuals--prior to public release of any data. This may be an occasion to see how discernible person-specific data really are before raising an alarm. E.g., there may appear to be no fair way to assign the chores, but if you simply asked, you'd find that everyone wanted a different job anyway.

**Gunnar Thorvaldsen, University of Tromsø, History (Faculty)**

Even if it is a hassle, one possible compromise is to construct one public sample that groups people, and another with full details but only given to researchers who sign a form about confidentiality.

**Confidential Response**

I work on the Longitudinal Study, which is a 1% sample of the population in England and Wales with people being selected from each census by 4 particular birth dates and we have to also take into account confidentiality of individuals. The privacy and confidentiality laws are frustrating however I do believe that they are there for a reason. However since PUMS data has previously had data to a higher level of subject and geographic detail I believe that it would be a shame to not be able to continue in this way. I also feel that researchers will suffer and analyses will be harder which will result in a lack of interest therefore making the census not as fruitful.

**Barbara Reskin, Harvard, Sociology (Faculty)**

The census already ensures privacy through several mechanisms, such as top coding, and not releasing data on some levels. The UK uses a different strategy--of being more selective in to whom the data are released. Although reducing access would be a disaster for graduate training and undergraduate instruction (making experiences with census data unavailable to many people), I would prefer reducing access to data to not making detailed data available to anyone for analysis.

**Louis Chauvel, Institut d'Etudes Politiques de Paris, Sociology (Faculty)**

No sociologist is about to find specific individuals with so little information. And even It could be possible, what is the interest of finding 1% of the pop. and not the 99 other percents?

**Carl Schmertmann, Florida State University, Demography (Faculty)**

Selective imputation or 'blurring' of data can (and already does?) make PUMS data unreliable for identifying individuals, without compromising its overall statistical integrity.

**William Collins, Vanderbilt University, Economics (Faculty)**

I cannot imagine that more stringent restrictions are necessarily or desirable.

**Monte Aaker, Minnesota Housing Finance Agency, Government**

This is silly. No one can be identified!

**Susan De Vos, University of Wisconsin-Madison, Sociology (Researcher)**

If PUMS data form a sample, how can one trace a specific individual even if one wanted to?

Might it be possible to enable a researcher to customize a data set such that one variable of little interest were aggregated while the other ones remained more detailed?

**Irene Bloemraad, Harvard University, Sociology (Graduate Student)**

As mentioned above, if less detail is given in the PUMS, there must be a mechanism in place to allow researchers (American citizens or not) access to the detailed data for specific projects. I would imagine that this would require travel to Washington DC, the establishment of a statistical center for outside users, and a significant investment in equipment and personnel. The beauty of Public Use Microfile data is that individuals can analyze the data in their own institution, reducing costs to the government and Census Bureau.

**Caroline Hoxby, Harvard University, Economics (Faculty)**

Identifying individuals through the PUMS would be impossible. Researchers who use the PUMS also have no interest in identifying individuals. Someone who did want to find an individual (e.g. a detective) would certainly not use the PUMS--there are data easily accessible on the web that would be far more useful.

**Allen Glicksman, Polisher Research Institute of the Philadelphia Geriatric Center, Sociology (Researcher)**

At this point it would be very difficult to identify any individual in the data set. To be honest, I am not even sure how one would go about the process even if there was a desire to do so.

**Myron Gutmann, University of Texas at Austin, History (Faculty)**

It is important to preserve the confidentiality of Americans in the PUMS, but the preservation of confidentiality does not require more stringent measures than are already employed. The same is already so small, and the American population so large, that in the context of regions of 100,000 persons or more it is unthinkable that anyone could be positively identified based on the variables used in the 1990 PUMS and the size of 1990 PUMAs.

**Confidential Response**

Considering the large group of people included in a PUMA and the sampling technique I do not think confidentiality would be easily compromised. The PUMS are the single most used raw data source at least at our Census Data Center and we would like to retain the level of detail or enhance it to be better able to use the data for research purposes.

**Confidential Response**

Regarding the time frame for release of the data and confidentiality, I would suggest that less detailed data be released as soon as possible with the intent of releasing more detailed data later after confidentiality is insured.

**Ruth Turley, Harvard University, Sociology (Graduate Student)**

My response to question 11 is based on the current level of detail. It would take too much effort for even an experienced researcher to identify an individual. These days, with so much information available on the internet, I would be more concerned with privacy issues there than in the Census.

**Confidential Response**

Other surveys such as SIPP use topcoding of income to prevent potential identification rather than income grouping.

**James Moody, The Ohio State University, Sociology (Faculty)**

One solution to the problems posed by deductive disclosure would be to fuzz some of the data. One could add a small random number (with mean 0) to income and age, for example, which would have only marginal impact on research, but would mask individuals quite well.

**Tina Lund, Urbanomics, Industry**

I generally work with data for New York City--which I believe has the smallest PUMAs as far as geographic area is concerned and even then, it's practically impossible when dealing with a representation of 100,000 people, to find information out about any one individual.

**Olivia Mitchell, University of Pennsylvania, Economics (Faculty)**

If it is necessary to issue geographic detail under slightly more restrictive conditions, this could be done (see the protocol developed for the Health and Retirement Study, for instance, at [www.umich.edu/~hrswww](http://www.umich.edu/~hrswww)). But it must be available for research in order to do what is necessary for good science and good policy.

**Confidential Response**

Has privacy vis-a-vis the PUMS been a problem or is this just a backlash of privacy issues and the Internet? If it's just the Internet, then the problem comes in making sure the Census Bureau's computers are secure from hackers, not making the data practically useless. If there are/were problems with the PUMS, then those problems should be identified and addressed specifically.

**Ellen Meara, Harvard Medical School, Economics (Faculty)**

No researcher I know has any interest in uncovering information about individuals. If you are worried about unscrupulous use of the PUMS, it seems like very strict fines, criminal prosecutions for individuals or institutions who use data improperly are appropriate ways to deter this. As a potential respondent in the PUMS, I feel confident that researchers can learn much less about me from this survey than what is readily and legally available based on my purchases, internet, etc. This is not a realistic concern.

**Susan Brudvig, Other Nonacademic (Nonacademic)**

I am a PUMS user and do not see that for geographies of 100,000 persons how it is that data be used to identify individuals. Rather than restricting data availability, it may be prudent to develop policies &/or procedures which would prevent small cell reporting. Let the user community decide how that would be done ... and police themselves.

**Confidential Response**

For items such as age and income, I'd suggest adding "white noise" to the data rather than using intervals. Both methods result in a lowering of data quality, but the white noise solution would achieve a similar degree of disguising of individuals at much lower cost.

**Confidential Response**

Assume someone has an unusual occupation, is married to someone else with an unusual occupation, and you know the exact ages of their children and a few other facts about them. You might be able to identify them (or guess with a high degree of probability that it was these particular individuals) on the PUMS and find out other information that you didn't know, such as

their income. I personally have no interest in information about particular individuals, and the PUMS is a rather crude method of gathering information compared to the many sources available in the private sector that actually have a person's name attached. Privacy is a legitimate concern, however.

#### **Confidential Response**

If it has been possible to maintain privacy and confidentiality in the past, how strong can the case for the Bureau be to minimize details?

#### **Darcy Hango, Ohio State University, Sociology (Graduate Student)**

If the census is concerned about privacy and confidentiality they should talk with people at other major survey centers (e.g. NORC, or the group at North Carolina who collected the ADD Health data) to see what they do.

#### **Roland Daeumer, The Pennsylvania State University; Population Research Institute , Demography (Post-doc)**

The Bureau could create two data sets - one more restricted than the other. The less restrictive data set would require additional permissions, e.g., special agreements of how to use the data, precise description of the intended research project, institutional affiliation etc. This would help everybody: serious researchers, policy makers who rely on serious high quality research, and the general public.

#### **Confidential Response**

Increase the penalties associated with the misuse of data. Permanent loss of federal research funds and loss of a job would be sufficient penalties to deter academic researchers from misusing data. Hefty fines might work for non-academic research.

#### **Eric Millis, Federal Reserve Bank of Dallas, Economics (Support)**

I have never seen tabulations of how many people have been identified by Census forms and how they have suffered because of it. At least this much is necessary evidence to provide before so injuring the country's primary study of itself.

#### **Gregory Reeves, The Kansas City Star, Media (Nonacademic)**

What about substitution? Doesn't the Census Bureau already substitute data in cases where a respondent might be easily identified (e.g., a high-income Asian household in a poor rural area)? Using substitution in PUMS would remove any real possibility of recognizing an individual household.

#### **Scott Adams, Michigan State University, Economics (Graduate Student)**

You were probably not planning on doing this anyway, but the 1980 census (and perhaps earlier censuses) provided quarter of birth and year of birth. This, coupled with SMSA and other information, might allow identification of subjects in extreme circumstances.

#### **Confidential Response**

Why not allow for geographic information down to the tract, block group, or block level when confidentiality agreements are signed (as is done in other data sets such as PSID or NLSY)?

#### **Liana Sayer, University of Maryland, College Park, Demography (Graduate Student)**

I believe the issues of confidentiality are less salient for academic research than for other uses of Census data, such as market research for fundraising or targeting sales. Perhaps the Census bureau could provide more limited files to non-academic users. In general, it seems the Census bureau does an excellent job protecting the confidentiality of users.

#### **David Kantor, Institute for Policy Studies, Johns Hopkins University, Other Academic (Support)**

For question 11, maybe it's plausible; it's hard to say, since I've never been inclined to try it, and I suppose one would have to be maliciously inclined to do so.

#### **Confidential Response**

In an age in which individual privacy is a legitimate concern it seems that PUMS is the last place than an unscrupulous individual seeking private information would go. Given the widespread availability of sources of personal information (i.e., direct marketing firms, credit reporting agencies etc.), it seems absurd to assume that somehow individual privacy will be enhanced by restricting PUMS data.

#### **Confidential Response**

Using characteristics like age, race, ancestry, and occupation to breach confidentiality is ludicrous. If the PUMS is a sample of the microdata, and there are no names on the public use data, these characteristics have absolutely NO value to those who want to find out about specific individuals.

To reiterate: There should be a choice in question 11: "Absolutely not. Anyone who claims that it is has an agenda."

#### **Confidential Response**

It seems plausible but unlikely, owing to the amount of work required.

#### **Mark Stern, University of Pennsylvania, History (Faculty)**

It strikes me that after the data are tabulated, the Census Bureau could identify rare combination of traits (e.g. elderly African-American women steamfitters) where privacy could conceivably be compromised and then use a more surgical process to reduce the threat.

#### **Confidential Response**

Some ethnic groups may be so small (i.e. persons with Middle Eastern backgrounds) that if a researcher were studying that particular group than he/she could ascertain much information about that group of individuals. Any information published would be a breach of privacy for these groups of people. What if these are individuals that the US government is interested in gaining information? The PUMS is then freely available with geographical detail for them to find the information they need. I can understand a move to reduce detail in race categories and geographic detail for privacy concerns, although I use both in my research.

#### **Susan Losh, Florida State University, Sociology (Faculty)**

Some idiots or charlatans will probably try. Keep the geographic units large enough and truncate outliers (see above) and that should do it.

#### **Mary Washington, Lehigh University, Sociology (Faculty)**

There are many sources for gathering individual data level that are more user friendly than PUMS. For example, using an address and the online "reverse" phone book directory an individual's name, phone number as well as a map of the location their house--with directions can be acquired in less than 10 seconds. Additionally, the user is directed to public sources of personal and financial data on an individuals--including where they attended high school. And given that most PUMS releases are 1 or 5 percent samples the odds of locating a particular individuals are much lower than the perceived 100 percent that these "find a lost one" databases.

#### **Confidential Response**

The users are a select group. Abuses could be traced quickly and would not be worth the risk just for a sample

#### **James Cassell, University of North Carolina at Chapel Hill, Sociology (Graduate Student)**

Are there any \*actual\*, verified cases of specific individuals being identified in a PUMS file? I sometimes wonder if the elaborate security measures employed by the Dept. of Education, BLS, and other statistical agencies are more products of the fertile imagination of lawyers than of real, potential threats to invasion of respondent's privacy. As they say, if it ain't broke...

#### **Jeff Elsworth, Purdue University, Economics (Graduate Student)**

As researchers we are taught to always keep our data and survey responses in the strictest confidence. It is paramount to getting people to trust us enough to give us information for our research. At some point there needs to be a level of trust or all research will be suspect.

#### **Confidential Response**

Standard screens like removing/censoring outliers, or grouping into categories. After processing data, try identifying individual records, and decide what variables to group.

#### **Confidential Response**

It's hard to imagine anyone trying to 'undisclose' individual data using the PUMS. There seem to be so many other sources of information on people (e.g. through the Internet) that would seem to be easier to use, if one wanted to find individual-specific data. Besides, researchers working with the data have no incentive to try to match up observations with actual people - the individual-specific data is valuable for statistical analyses on representative individuals, not for some sort of 'case study' of particular people after they have been identified. I can understand the reluctance of the Census Bureau to risk even the appearance of disclosability, but I very much doubt that it is a practical problem for this data.

#### **Confidential Response**

I agree that privacy is of great importance. However, in the current form it is already all but impossible to infer an individual's identity. Furthermore, I am not aware of any attempts to identify individual respondents.

**Arthur Wilke, Auburn University, Sociology (Faculty)**

I haven't entertained the idea so I am unclear what might be involved in violating a person's confidentiality.

While I endorse confidentiality, it seems strange that if I incorporate I can get all kinds of confidential information. This may be the way privatization will proceed.

**Susan Mayer, University of Chicago, Sociology (Faculty)**

Given the costs associated with collecting Census data, an important goal of Congress should be to maximize its usefulness to non-government and government users. This increases its social value and increases the benefit to cost ratio. It would be a pity for the huge effort that goes into collecting census data to result in a data set with very limited usefulness because of the extremely remote possibility that a determined person could identify an individual. Given the information sources available, using PUMS data would surely be a rather silly and cumbersome way to gather information on an individual.

**Amy Goodin, UNM Institute for Public Policy, Geography (Researcher)**

The data is only traceable if someone knows an individuals particular characteristics and the data is released at the individual level. The data I have seen released is aggregate level data for small subpopulations. I thought the law required individual level data to be held for at least 72 years to protect privacy and confidentiality?!

**Edward Ratledge, University of Delaware, Demography (Faculty)**

I personally think the whole thing is an over reaction. How could someone conceivably find the linking information from another source that would reveal the identity of the individual. What would be their purpose?

**Robert Wolff, Central Connecticut State University, History (Faculty)**

The collection of a highly specific Public Use Microdata Sample does not present any significant threat to individual privacy. The confidentiality measures already in place mean that PUMS information is far more secure than, for example, a person's individual medical history.

**Jeff Claassen, Fort Worth Star-Telegram, Media (Nonacademic)**

Given that the data is a sample of 5 percent of a group of 100,000 people, it's inconceivable how an individual record could be traced back to a person.

**Susanna Loeb, Stanford University, Economics (Faculty)**

It is plausible that PUMS could be used this way if the data were available but not likely. Even the level of detail available in 1990 does not allow even sufficient detail answer many of the questions that I am interested in.

**Russell Davis, Louisiana State University, Sociology (Graduate Student)**

I have it open and running right now, and I would challenge anyone to find someone they know.

**Confidential Response**

There are actually far more intrusive and less confidential means of gathering all kinds of data all of us would prefer not be available. I suggest that those concerned about the use of this data confer with those in the credit card industry, which has access to everything I have ever done or will even contemplate doing, and sells it to anyone it chooses. Depriving researchers of the highest available quality data for such reasons sounds to me like an attempt to make sure that government is the last to know what is actually happening to the nation's population.

**Confidential Response**

Ok to quote as "A Federal Statistical Agency survey manager", but not by name.

As frightening as the response rate problems this and the very public concerns about the privacy of Census data are to both survey collectors and users, I think the way the Federal statistical community has reacted to it very counterproductive. It misses the point and actually undermines support for data collection. People are scared of information being added to a government dossier on them. The idea of some researcher figuring out their salary by some elaborate crosstab or neural net analysis would never occur to them. And if it did, it would seem more silly than frightening. What increasingly esoteric efforts to prevent some theoretical possibility of individual identification do do, however, is greatly reduce the value of the data collection to researchers and thus ultimately its value to the nation. Up until now, the restrictions on the Census PUMS has still left a fairly usable product. However, many other public use data files at Census and other Federal agencies have been striped of their utility. Agencies should fear that they are also stripping away the constituencies for their surveys when they do this.

**Andrew Beveridge, Queens College and Grad Center CUNY, Sociology (Faculty)**

At present through CUNY Research Foundation we hold licenses to two confidential data sets. One is the National Educational Longitudinal Study 1988, which includes questions about drug use, sexual behavior, pregnancy, and the like. It is given in a confidential form, no names are on the file, but zip codes are. Anyone using the data must present a security plan to the Department of Education. There is spot inspection. (In fact, we were ruled in compliance.) The other data set is files of Hospital Discharge Records for New York State, which includes an encrypted patient ID, but also patient address. There is limited ethnic/race data, but there is great detail about medical condition, including STD and HIV/AIDS status, drug use, alcohol, pregnancy, and the like.

In both cases, there is a "public use" file which does not have restricted data, and then there is a file that includes the restricted data. In both cases, the license agreement calls for each person using the data to file an affidavit and that affidavit must be filed with the authority that grants use to the data. In both cases there are criminal penalties for violating the license. I know the Census bureau has set-up a few secure facilities to do analysis with Census data, but I think going to such a licensing scheme would make sense. This way a public use file could be readily available to anyone, and researchers would have access to the more restricted file. It would make vending the data from the IPUMS project more cumbersome, but it is a far better alternative than not allowing the data to be accessed by legitimate researchers.

**Michael Sullivan, Freeman, Sullivan & Co., Sociology (Nonacademic)**

Frankly, given the mobility of the population, the level of detail in the PUMS data and the availability of alternative information sources, I think it is highly unlikely that this information would be used to invade the privacy of citizens. But it is possible that some individual (with a lot of time on their hands) might attempt (successfully) to do this for reasons none of us can fathom today. If you want to ensure that these databases are not misused on a systematic basis, establish Federal Regulations making it illegal to use the PUMS to identify individuals providing for stiff economic penalties for each instance in which a violation occurs (say \$1,000 per case).

**Diane Cowper, Dept of Veterans Affairs, Demography (Researcher)**

The level of aggregation of the PUMS data is already sufficiently high as to preclude identification of individuals.

**William Bridges, University of Illinois at Chicago, Sociology (Faculty)**

Is there any evidence that anyone has ever been harmed by release of PUMS data? Perhaps there should be a level of confidentiality in between Special Sworn Status, and what would be provided to the general public, including marketers, information snoops. The Center for Educational Statistics has such a system; it's a pain, but if the alternative is collapsed data, I would prefer it.

**Michael Haines, Colgate University, Economics (Faculty)**

I think that privacy is well protected right now. I don't think that is the real issue. I think that is a cover for just reducing resources to the production of data for public use. This is a very short sighted and false economy.

**Eric Bettinger, MIT, Economics (Faculty)**

PUMS might also start to use restricted-access datasets which require some sort of affidavit similar to National Center for Educational Statistics

**Christopher Lyttle, Rush Primary Care Institute, Sociology (Researcher)**

Please do not take the approach used with the Mortality Detail Files. In those data the county identifier is suppressed for counties with a population less than 100,000. This may be reasonable if the individual died of a rare condition, but those conditions are, by definition, rare and should not drive the overall level of aggregation. What is needed is to identify those individuals with a highly unique combination of characteristics, and to reduce the level of specificity in the data FOR THOSE INDIVIDUALS.

**Charles Brown, University of Michigan, Economics (Faculty)**

The discussions I have seen about confidentiality have been heavily weighted on the side of "is it possible to identify anyone...." and if you go down that path seriously you would end up with no micro data. While I don't over-estimate the ethics or judgment of credit checkers and others who actually want data on identifiable individuals, it's important not to over-estimate what they might gain from abusing PUMS and so their incentive to do so. Only 1/(sampling rate) of the folks they would want to learn about are on the file, at best only a few of the oddest cases among those might be identifiable. And, while the data are accurate enough to be used for statistical purposes, there is enough error in items like income and occupation that there is already an important amount of blurring in these variables.

I would personally support stronger legal sanctions against anyone using PUMS or other data files to identify individuals for credit checks or similar abuses.

**Joyce Jacobsen, Wesleyan University, Economics (Faculty)**

Could limit access to the microdata files, having persons sign a confidentiality agreement that would be kept on record with icpsr and/or other gatekeeper institutions.

**Jon Lorence, University of Houston, Sociology (Faculty)**

I know of no academic researcher who would want to know the name of any specific individuals in the PUMS. The concern about privacy is a bogus issue because the detailed information in the PUMS would still not enable one to identify specific individuals.

If the Bureau wants to maintain confidentiality, certain kinds of information could be deleted for individuals.

**Confidential Response**

I may be wrong, but as far as I know when one talks about confidentiality or privacy, one talks about names and I do not think one can access the names.

**Karen Woodrow-Lafield, Mississippi State University, Demography (Faculty)**

At this moment, I am not certain whether ethical guidelines of the American Sociological Association, Population Association of America, American Statistical Association, and others are written in sufficient depth to assure members' commitment to using PUMS and other data sources with care to protect privacy and confidentiality of respondents. With greater technology, greater vigilance in drafting these statements is needed.

We have been accustomed to acquiring microdata files easily through the Internet; we have statistical analysis packages that accommodate highly sophisticated models with many variables and categories. This ease may be somewhat antithetical to the constant attention to the nefarious uses that some might choose to make of innocently produced tabulations.

I would hope a Census-data user agreement to avoid disclosure of individual's privacy could be developed and used as a matter of course.

**Elizabeth Mustaine, University of Central Florida, Sociology**

I cannot imagine how someone could look at the PUMS data and figure out who a particular individual is. There are too many cases, too many variables, and too many individuals who will have similarities across their characteristics. It is unbelievable that anyone would or could be able to identify individuals.

**Steven Levitt, University of Chicago, Economics (Faculty)**

It seems to me that if the level of geographic detail is 100,000 person units, it is not a concern about confidentiality, except in unusual circumstances (e.g. Eskimos in Alabama). I believe it is better to offer a detailed data set and veil information in the occasional small cell sizes, then to lessen the amount of information throughout the entire data set.

**Young Cho, University of Illinois at Chicago, Sociology (Post-doc)**

I think it is more proper to make it illegal to use pums for identifying individuals rather than making the data useless.

**Leif Jensen, Penn State University, Sociology**

People on the Hill don't seem to realize that it would be impossible to identify individuals with PUMS data, and that nobody tries to use the data in this way anyway. They also don't seem to realize that PUMS data have been used to conduct research whose results support positions on ALL parts of the policy (liberal-conservative) spectrum.

**Masao Suzuki, Public Policy Institute of California, Economics (Researcher)**

I am concerned that the Census Bureau may be under political pressure, since I have read newspaper accounts of Congressmen and/or Senators questioning the need for census data. I think that there is a lot of public concern about privacy, but this needs to be directed at its true source, private-sector internet businesses, and not redirected at academic researchers who would have no interest in trying to identify individuals.

**James Walker, University of Wisconsin, Economics (Faculty)**

I think it might be useful to move the discussion from "restricting the ability of researchers to identify specific individuals" (as in question 11) to "penalties for misuse of the data." I believe few researchers, if any, will use the PUMS to identify characteristics of specific individuals. However, with modern computing equipment and the wide availability of data on the web, some people will be identifiable (e.g., wealthy blacks Wyoming). Perhaps some restrictions could be placed on access

to the PUMS much as BLS does to gain access to the geocoded data of the NLS or the NIA does for detailed person specific information in the Health and Retirement Survey. Such barriers will be small and inconsequential to the research community which has no interest in finding specific people. But, the barriers may be effective in barring users interested in finding specific users.

**Mark Fossett, Texas A&M University, Sociology (Faculty)**

I believe the promise of confidentiality is crucial to the success of the census. Realistically, the PUMS are not likely to be a major source of confidentiality problems -- better and more timely information could easily be obtained by other means (e.g., internet sources). But I understand that confidentiality issues must be emphasized over and above the degree necessitated by objective risk.

**Halliman Winsborough, ICPSR and University of Wisconsin, Madison, Demography (Faculty)**

The first PUMS file came out about 35 years ago. Has anyone been identified? I think not.

**Confidential Response**

What specific evidence is there to suggest that the current PUMS data structure has contributed to violations of privacy for those included in the sample? It seems to me that the existing geographic restrictions effectively preclude the possibility of linking the PUMS data to names and addresses. In fact, I would be in favor of increasing the size of the sample in order to enhance geographic precision. I suspect that we could drill down a bit deeper without posing a threat to anyone's confidentiality, particularly in urban areas.

**Michael Conaway, Institute for Social Science Research, University of Alabama, Demography (Researcher)**

I strongly support absolute confidentiality of respondent identity and that includes the revelation of information that could lead to respondent identification; but I don't see how that is likely to happen from the PUMS data. I am willing to be persuaded otherwise.

**Harold H. Kassarian, UCLA and Cal State Univ, Northridge, Other Academic (Faculty)**

Of all the breeches of privacy, this is among the least onerous. In fact I really do not see that much of a privacy problem here. Are you sure it is privacy and not politics that are the major issue here?

**James T. Ault, III, Creighton University, Sociology (Faculty)**

Breaches of confidentiality are clearly possible, although the historical record (as I understand it) suggests that it has not been a problem. Although improvements in Information Technology do make it easier to analyze --and to commit unethical and/or illegal breaches of confidentiality-- I believe there are other mechanisms available to deter such acts...without eliminating legitimate academic research opportunities.

**Leann Tigges, University of Wisconsin-Madison, Sociology (Faculty)**

Far greater assaults on privacy can come through the kinds of information now being gleaned from use of the internet.

**Confidential Response**

It may be possible to offer a number of PUMS files with users trading geographic level size in exchange for less detailed coding of variables. An example would be a PUMS with geographic units of 10000 pop but with highly collapsed variable coding. From there you could have 50000 pop units, then 100000 pop units, 250000 pop units, etc. each with more detailed variable coding. Another alternative would be to increase the PUMS sample sizes. With today's high speed hardware and software, maybe it is time to offer the entire long form sample to interested users. Other comments;)

**Robert Fairlie, University of California, Santa Cruz, Economics (Faculty)**

It would be nearly impossible to identify someone based on the variables in the Census. It would require someone who has the expertise and training to work with these data and sophisticated statistical programs. Why would someone with that kind of training be interested in identifying someone in the Census? Also, it would only work for individuals who had a cross-tab of very unusual characteristics (e.g. age=102, race=black, ancestry=Irish, occupation=professional athlete, etc...).

**Confidential Response**

There is always statistical disclosure risk. I think it is very important for the Bureau to demonstrate the risk with current (1990) PUMS data, before we are asked to rely on less information (data).

**Charles B. Nam, Florida State University, Demography (Researcher)**

I am a great supporter of privacy and confidentiality in census returns and public-use data. With limited geographic detail, I see no way that confidentiality can be compromised.

**Judith Rowe, Princeton University (retired), Economics (Support)**

If one wanted to identify an individual there are better and cheaper ways to do it.

**Confidential Response**

I'm not really familiar enough with the data right now to comment on that. But I think sociologists doing academic research would abide by the ASA ethical guidelines on confidentiality. These guidelines are also present at the university level through the IRB board, so I'm confident that the privacy of individuals can be protected.

**JoAnn Dionne, University of Michigan Library, Other Academic (Faculty)**

While there is always the possibility that users will try to identify individuals I think the possibility is very remote. I am much more concerned about what marketers know about me than what the government knows.

**David Brownstone, University of California, Irvine, Economics (Faculty)**

It might be wise to require all users of PUMS to sign a statement acknowledging (under threat of some penalty) that they will not attempt to identify individuals using PUMS data.

**Patricia Becker, APB Associates/Southeast Michigan Census Council, Demography (Nonacademic)**

To Q11 above: absolutely not. I think a problem has been created where there is none. The data are three years old, limited to geography of 100,000 or more, and the files are primarily used by expert census data users who have no interest in this. As I told a high census official earlier today on the phone, if you knew someone had a long form and exactly what their responses were--such as for yourself--you might be able to find the record in PUMS if it made it there. If you don't know a household/person is in the sample to begin with, forget it!

**Richard O'Reilly, Los Angeles Times, Media (Nonacademic)**

In California, the 100,000-person population minimum for a PUMA serves quite well to prevent identification of an individual. I cannot imagine any scenario where a individual could be identified. Nor can I imagine any scenario where anyone would bother to try. Personal and economic data about individuals is much more readily available on the Internet without any ambiguity.

But what isn't available anywhere else is the kind of information that one can derive from PUMS data, such as the dominant industry of self-employed Asians in Los Angeles County being restaurants.

**Confidential Response**

I understand that confidentiality is a big concern and it seems to be a valid one. However, many other datasets have successfully explored ways to minimize the risk of confidentiality breaches without sacrificing data quality and I would urge that this be done in this case. Limiting data availability, asking for signed statements to uphold confidentiality, etc., are all good ways of addressing this important issue. But rendering data less useful to researchers seems to be a convoluted way to deal with it.

**Tracy Dietz, University of Central Florida, Sociology (Faculty)**

Specific categories are useful for researchers who wish to better understand and predict social behavior. Yet, because of the size of the US population, the data set is generally too large to make it feasible for anyone to identify individuals.

Moreover, given the technological advances of the last two decades, the PUMS simply does not contain any information about individuals that is not readily accessible to population anyway. In fact, it probably more easily obtained via other methods than through narrowing down using the PUMS.

**Susan Thistle, Northwestern University, Sociology (Faculty)**

I think it will be fairly easy to come up with ways to ensure the confidentiality of PUMS data and that effort should be made in this direction rather than in limiting collection of data.

**Donald J. Bogue, University of Chicago, Demography (Faculty)**

I have never heard of a case of a breach of confidentiality using PUMS data. Is there a documented case on record? If so, what privacy was breached and how serious was it? Abuse of privacy of credit, bank, and other records is rampant, with little efforts to control or regulate it. Persons seeking info about particular others would find PUMS about the least productive entry-point for gaining personal information.

**John Blodgett, University of Missouri (OSEDA), Sociology (Support)**

It seems possible to me that if someone had nothing better to do, and if they or someone they knew happened to have some unusual combination of characteristics, and if that person happened to be included in the PUMS sample and if they had the computer access and expertise, then yes -- it could happen. Just as it could have happened in each of the last 4 decades. Increasing the size of the PUMA's to render them next to useless for many applications will not change the situation.

**Confidential Response**

It is possible that someone would try to identify the characteristics of specific individuals, but it seems doubtful. Perhaps all users should sign a confidentiality agreement.

**Aaron Pitluck, UW-Madison, Sociology (Graduate Student)**

Given that data is already censored to preserve confidentiality for small demographic groups in some datasets, I fail to see why privacy or confidentiality is a problem. If it were a realistic problem, naturally I would support further aggregation.

HOWEVER, relative to true information available on the web (credit card information, history of addresses, telephone numbers, and so on) traceable through a social security number, a name, and a date of birth, the IPUMS information is not only unspecific, it is well nigh impossible to track to a certain household with any degree of certainty.

This appears to me to be a budget issue rather than a confidentiality issue. As a researcher, needless to say I find the IPUMS data unique and a further aggregation of its data would be highly unfortunate.

**Confidential Response**

It's implausible that someone would be able to identify respondents, and hard to imagine why any researcher would do so.

**Confidential Response**

I think anything is plausible. Yet, breaches of confidentiality are extremely rare; I know of no such case. One response would be to license access to the data; some federal government entities have done this and such programs have been successful. The licenses, however, tend to make the data available only to academic researchers. PUMS data is used by a wider variety of communities, and thus perhaps this approach may not be feasible.

**Stephen Snyder, University of Maryland, College Park, Economics (Graduate Student)**

PUMS data will not be used to identify individuals by academic and policy researchers. My guess is that very recent PUMS data could possibly be of interest to marketing types. I suspect they have less expensive ways to get rich data sets on target populations, but they might possibly find it useful to add PUMS to the mix. This would clearly be undesirable, as it would affect confidence in census confidentiality. This would be true even if 95% of the data came from commercial rather than Census sources. I suspect it would be possible to prohibit the commercial use of PUMS data while allowing academic research to continue. I think this should be our goal.

**Hayward Derrick Horton, Dept. of Sociology, SUNY-Albany, Sociology (Faculty)**

I think that the Congress should rely on the Census Bureau to ensure the confidentiality of the PUMS data. The bureau has done a fine job and there is no reason to suspect that it won't continue to protect the confidentiality of all citizens.

**Abbott Ferriss, Emory University, Sociology (Faculty)**

In my research, no reference to individuals are required. Only aggregated statistics are employed.

**Abdelhani Guend, UW-Madison, Demography (Graduate Student)**

I used PUMS to generate data needed for Federal Development grant for very small community but never encountered any such issue as individual identification.

**Michael Dennis, UC BERKELEY, Economics (Graduate Student)**

A) Given the ability to investigate people's personal lives over the internet, the sale of DMV information about specific individuals to marketing firms, the ease obtaining credit reports and that sort of thing, it seems to me that the IPUMS is one of the smaller privacy/confidentiality issues that confronts us as a society.

B) It is probable that most individuals who are afraid of people knowing the content of their Census forms will lie or omit answering the questions they object to. Most of the information on the Census form would be easily gathered from other sources--that's exactly how we think people might be identified.

C) In light of B, perhaps those individuals to be included in the IPUMS might be informed of the process and consulted as to whether or not they would like to change any of their answers to non-response. Probably economically infeasible I suppose.

**Marco van Leeuwen, NEHA/IISG, Sociology (Faculty)**

The Historical Sample of the Netherlands obtains individual level data from government agencies on persons born in the 19th and early 20th C.. These data contain confidential information, esp since some of the persons are still alive while in other cases children are alive. To protect privacy, the HSN has rules and regulations concerning privacy to which every one using the public use sample must obey. In addition, requests by researchers that need more detailed information are considered by the board of the HSN and customized datasets are provided where possible, at a cost, and provided the data are kept confidential and only the results are published in a way securing privacy. That way both the interests of the academic world that needs detailed information and those of the general public that is concerned about questions of privacy are secured.

**Jan Mutchler, University of Massachusetts Boston, Demography (Faculty)**

I have worked with Census Public Use Microdata for almost 20 years and am convinced that confidentiality is a non-issue.

**Amy Guptill, Cornell University, Sociology (Graduate Student)**

There's simply no way of knowing which specific people or families fit specific profiles, and that is not of interest to people in my field. It is the overall trends and relationships among variables that count.

**Martha Hill, ISR, University of Michigan, Economics (Researcher)**

Why collect the data if it is not available for research? Care should be taken with confidentiality matters, but fear should not be the dominant force in decisions about how to balance guarantees of confidentiality with the creativity in expanding our understanding of the world that comes with wide availability of data to the research community. We must consider the value of foregone research if restrictions are placed on what is publicly available.

**Confidential Response**

Names are not necessary, except for tracking purposes.

**William Dow, UNC-Chapel Hill, Economics (Faculty)**

At the least, the usual detailed data should be made available at the regional Census Research Data Centers.

**Kathryn Hornberger, HealthFirst Management Group, Industry (Nonacademic)**

From an historical perspective, I would be interested in knowing how helpful PUMS data has ever been to private commercial marketers (insurance, credit, direct mail/sales, etc) in targeting \*specific individuals\*. There are already commercial programs which identify demographic characteristics down to block numbers (and specific homes within the block) based on commercial surveys. Marketers use this data, not PUMS data. The argument that privacy will be lost is a moot point. The question should be: who uses PUMS data and for what reasons? Governmental decision makers need to consider that with or without concern for PUMS privacy issues, our citizens are already pinpointed on the Net by such enterprises as Mapquest and through other demographic data gathering commercial enterprises. It is actually a non-issue. Just check out a VALS survey. "They" already know who we are. If I don't mind paying for the information, I can get anything I want TODAY - including a satellite picture of my neighbor's property.

**Charles Rambeck, St. John's University, Economics (Faculty)**

While I appreciate the need for privacy and confidentiality, I believe current procedures are adequate. This proposal seems to be an excessive reaction to current political criticism of census questions. I could live with an effort to protect truly unusual individuals, such as persons over 100 in geographic areas below 250,000, but hope they will not reduce the information content of the PUMS in response to unthoughtful political attacks.

**Wilbert Van der Klaauw, UNC-Chapel Hill, Economics (Faculty)**

If there is a concern about privacy issues, why not release an initial version of PUMS with more aggregated data during the period in which these privacy issues are further investigated. I doubt that any variable is finely enough categorized that in combination with other variables any specific individual could be uniquely identified. If this is the case at all, than this must be because some values of a variable are so rare that it may be justifiable to consider grouping values only for that particular variable. I see no reason to aggregate age, for example, as there are large numbers at each specific value. Similarly, except for very large income values (where one may want to consider topcoding), there will always be many individuals with similar income levels.

As far as geographic variables are concerned, the most relevant issue here is that the PUMS data can be linked to other aggregated geographic data from other surveys. As long as the geographic data collected permits this, I don't believe it necessary to release finer, more precise, data.

**Carl Backman, Buffalo State College, Sociology (Faculty)**

While I am very interested in seeing maximal use made of census data, I am even more committed to the importance of maintaining confidentiality. I wonder if it makes sense to "fuzzify" some of the micro data, as is done with small area summary data. It might also be possible, though a true pain in a bureaucrat's backside, to use different levels of outlier topcoding with different geographic areas (eg, an area with lots of nonogenarians might topcode age higher than an area with relatively few of the old old). To be sure, this creates some problems in merging files from different geographies, but these problems are not insurmountable.

**John McHenry, Demographic Data For Decision-Making, Inc., Demography (Nonacademic)**

It is my understanding that over the past two decades, one or more newspaper reporters has been able to use PUMS data to identify one or more individuals. If true, this is regrettable but a sense of proportion must be maintained. The "good" afforded by easy public access to PUMS data has far outweighed the "bad" that results from such a sensationalized story. Moreover, while I believe that a skilled researcher with lots of time on his/her hands might be able to start with a PUMS database and eventually identify one or more individuals from it -- the reverse situation is not true -- i.e. no matter how highly skilled, a researcher cannot start with a list of specific individuals and then with any certainty, identify their PUMS responses. Nor can I think of any reasons (apart from a story, apart from PUMS security testing, or apart from an exercise in academic curiosity) that a skilled survey researcher would waste his/her time on such an exercise.

**Joseph Persky, University of Illinois at Chicago, Economics (Faculty)**

The PUMS and Census data in general are gathered under a promise of confidentiality. Respondents should have every confidence that this promise will be kept. Although a PUMS user for more than three decades, I do not actually know what penalties, if any, exist for breaking such confidentiality and using these data to gain information about individuals. If no such penalties currently exist they should be legislated and if they do exist perhaps they should be given greater publicity.

**Jeanne L. Anderson, Jeanne Anderson Research, Industry (Nonacademic)**

It seems to me that the Census Bureau could mask identities by issuing several versions of the PUMS file, each with \*less\* detail on some variables but not all. Creating a \*long\* ID number, and attaching only a portion of the ID number to a record (or all records for persons in the household) on a specific file might enable users to obtain the detailed data they need but, even if one user purchased a series of files, would be unable to create a merged file with all detail data for a given respondent.

A given individual's "ID" would be different on different files. Here is a hypothetical but simplified example:

Of course, the ID # would have to be more than 3 digits, but with the use of letters or duplicating some digits in ID's, it would be possible to create an ID number for an individual on a given file that is unique for that file but also identifiable by the Census Bureau in case full data record for one individual is necessary.

**Mary Grace Kovar, NORC, Demography (Researcher)**

The PUMS data have to remain confidential. That's the promise. However, the means of doing so are improving. This is no place to discuss them, but I hope your group will consider means other than collapsing categories into useless groups.

**Confidential Response**

Regarding question 11, it is plausible (almost anything is plausible) that someone might use public use data to try to identify a specific individual, but the likelihood of this happening is pretty remote in my opinion. Most policy research looks for patterns and indicators of behavior, not investigate the behavior of a particular person.

While I think privacy and confidentiality are important factors to assure that the public participate in data collection efforts, I am not aware that, to date, there have been concerns about these factors with the PUMS files.

**Confidential Response**

Maintain confidentiality is important but should not be taken to the extreme.

**Peter Philips, University of Utah, Economics**

The notion of that there are privacy problems with the Census PUMS is false. I have used these data for 20 years in a host of public settings and never once was there ever a question of any specific individual being identified from these data.

**Charles Rynerson, San Diego Association of Governments, Government (Nonacademic)**

You would have to already know a lot about an individual in order to identify them in PUMS, particularly since there is only a 1 in 20 chance of their record being included. The additional information that you might learn from PUMS is no more revealing than what you could learn from a credit check, DMV, county tax assessor, etc. These sources are more reliable since they include all debtors/drivers/homeowners rather than a sample. If a researcher is weird enough to investigate

someone with PUMS, and then use the information in a way that would harm the person, they probably would use other data sources instead.

**Virginia Appl, Houston-Galveston Area Council, Government (Nonacademic)**

The Census Bureau is the only people left who are protecting privacy and confidentiality of respondents. When they do not provide data for fear of identification of individuals, researchers are forced to spend large amounts of money to purchase databases on individuals from private marketing companies. I personally would prefer to use Census products as I believe their data is far more reliable and representative of the general population than the data that is sold by marketing companies.

**James Peterson, Sociometrics Corporation, Sociology (Nonacademic)**

The Bureau might consider a two-tiered model of data release similar to that used by the Carolina Population Center for the AddHealth data. They have a less detailed public release version (distributed by Sociometrics) accessible to anyone (though requiring a signed pledge of confidentiality) and a more detailed restricted release version (distributed by the Center). The latter contains more detail on contextual variables that in theory might be used by an unscrupulous user to attempt to identify individual respondents. The restricted version is released only by application to the Center with a description of the intended use and a significant payment. This allows the Center to be highly confident that confidentiality will be well protected. Protecting confidentiality was of the utmost importance to them, and my understanding is that they feel confident that the procedures they have implemented with this two tiered system are satisfactory for this concern. You might contact J. Richard Udry at the CPC at UNC for more details.

**Bill O'Hare, Casey Foundation, Demography (Nonacademic)**

Since the raw data used to construct the PUMS files will have already gone through a data swapping process to protect confidentiality, I don't understand what the problem is?

**Ken Hodges, Claritas Inc., Demography (Nonacademic)**

It might be possible to identify individuals on PUMS, but I see the risk as extremely remote. Businesses and others whose applications involve the identification of individuals have far better resources available than PUMS. PUMS is great for its intended purpose (do-it-yourself tabulations), but for identifying the characteristics of specific individuals, PUMS is one of the worst resources you could use.

**Jeanne Gobalet, Lapkoff & Gobalet Demographic Research, Inc., Demography (Nonacademic)**

Those using PUMS data need to be extremely sophisticated users of computer software and must be well-trained in data analysis. I cannot begin to figure out how to go about identifying individuals (which I would never try to do because of my scientific training and ethical views). Why would anyone even want to try to identify individuals using PUMS, when countless commercial databases exist that already contain far more than the information that PUMS might supply? Those who have a few dollars can buy detailed information about individuals anytime they want from data vendors and don't need PUMS.

**Yoshinori Kamo, Louisiana State University, Sociology (Faculty)**

I don't see how it's possible with PUMA level identification.

**Mark Mather, Population Reference Bureau, Demography (Nonacademic)**

It seems possible that certain individuals could be identified on the basis of detailed occupation, race, gender, etc. (for example, black female college professors in North Dakota). Since I am less interested in the data on detailed occupations, I would not mind if certain occupation data were combined to preserve confidentiality. With the current public thinking on this issue, it is extremely important for the credibility of the Bureau that they maintain respondents' confidentiality.

**Julie Hudson, The Urban Institute, Economics**

Currently the BLS uses a system with the NLSY data that seems to work well. Access to "non-geocode" data is unregulated, while access to "geocode" data requires a special three-year contract through one's employer. This seems to work well.

**James Holland Jones, Harvard University, Demography (Post-doc)**

How about making it illegal to use PUMS data for commercial enterprises? If this were the case, who would have an incentive to violate confidentiality?

**John Straub, University of Wisconsin-Madison, Economics (Graduate Student)**

I'm not sure about question #11. I suppose the Census could ask some people on its staff to try and see if they can figure out a way to identify characteristics of specific individuals with PUMS data. I've never tried, but I suspect it would only be possible in some extreme situations, if at all.

**Jim Hinterlong, Washington University, Other Academic (Graduate Student)**

I am confident that, even if it were readily possible to identify the characteristics of specific individuals, scholars and researchers in the social sciences would follow the same ethical and professional guidelines that apply to the use of human subjects and/or personal information.

**Francine Atwell, Hawaii SBDC Network Business Research Library (under the Research Corporation of the University of Hawaii), Economics (Researcher)**

About PUMS data being used to identify specific individuals: it could be possible if the Census Bureau isn't careful, though I doubt many researchers actually would use it to that end. Private industry collects much more detailed data about individuals, and it is easily accessible if you have the money--e.g., I could get a mailing list of local residents who own homes and spend more than \$1,000/month on their credit cards in 2 days or less.

**Tara Gibbs, University of Minnesota, Other Academic (Graduate Student)**

I don't understand question 11. My goal is to profile communities and figure out how many people in a specific language group are isolated by language or have specific language needs in order to be prepared for various jobs, or to communicate with schools about their children, or the courts, or hospitals, or other public services. To accomplish this, I need to have all of the variables, language, age, number of children, amount of contact with various institutions, location, etc. attached to a specific individual x, so that I can run chi-squared tests on how certain variables affect other variables.

It would be nice to know how many times this specific person interacted with specific institutions not asked about on the census, but I doubt anyone would have the time, money, or manpower to actually figure out who the person is in order to survey them more specifically about which institutions they interact with. We just make estimates based on the statistical characteristics provided by the census and statistically numbers provided by the institutions.

**Andrew Pearlman, Yale University, Economics (Graduate Student)**

With all the other ways now available for obtaining information about individuals, it's very hard to imagine anyone using PUMS for this purpose.

**Alexander Bucur, USC/UCLA, Sociology (Researcher)**

While I believe that one might be able to obtain characteristics of some specific individual in some very low populated area, I cannot see what the benefits would be for such a convoluted and very tedious work. Those that possess the know-how and the hardware necessary for such analysis would not have any reason/purpose to identify such person(s).

**Deborah Alexander, University of Minnesota, Other Academic (Graduate Student)**

It is my opinion that the concept of individual privacy is under much more pressure from market research/internet snooping than census/GIS data. Scholarly researchers must be aware of, and protect, the individual privacy of subjects.

**Kenneth Sylvester, University of Victoria, History (Post-doc)**

To me, the geographic detail is the most sensitive subject area that would allow for identifications to be made. Some adjustment of this information might be attempted to preserve confidentiality.

**Karen Norberg, National Bureau of Economic Research, Other Academic (Faculty)**

I think it is extremely unlikely that the PUMS would be used in a way that would violate an individual's confidentiality. With the growth of the information industry & the proliferation of ways in which information can be collected about private individuals, I can understand that the public has become very sensitive about such issues. However, I think that the concerns about the use of publicly collected national data, such as the PUMS, are a displacement - public officials may be targeting this data, which I believe has very high public value - because it is much harder to address the genuine invasions of privacy from other sources.

**William Sundstrom, Santa Clara University, Economics (Faculty)**

It seems highly implausible that someone would go the PUMS to try to identify actual individuals, and I believe the effort would generally be futile even under less stringent confidentiality measures.

**Confidential Response**

The PUMA already are quite large (i.e., 100,000 persons) so the chances of identifying individuals is remote. Age could be collapsed to five year age categories.

**Confidential Response**

If someone uses data fusion techniques, it is possible for a very small percentage of Americans to be identified to a level of detail that will allow targeted advertising. At present, PUMS data is too generalized to be useful for many geographic analyses. If PUMS data is generalized even more than it is now, the loss of potential geographic research is immense. Every week, many Americans voluntarily give specific data more detailed than PUMS in product warranty registration cards. To reduce the quality of PUMS data for privacy issues is very short sighted. If anything, we should be increasing detail.

**David Swanson, Helsinki School of Economics & Business Administration, Demography (Faculty)**

With census bureau data there is virtually no chance that individual confidentiality can be breached. I have complete confidence in the Census Bureau to safeguard privacy and maintain confidentiality. I wish I could say the same about private sector vendors who maintain large and generally inaccurate databases of their own on Americans.

**Robert O'Connor, Penn State University, Government (Faculty)**

Privacy and confidentiality in the PUMS is a non-issue. Nobody is trying to identify individuals.

**Maggi Rose, Open University, UK, Economics (Undergraduate)**

I live and work in the UK. Your data is the best available. I have no chance, opportunity or desire to identify individuals.

**Francine Stephenson, SDC, NC Office of State Planning, Government (Nonacademic)**

With areas of 100,000 or more, it would be possible, but very difficult to discover information about an individual. If the Census Bureau uses the disclosure avoidance method of data exchange, no one could ever be certain that they had discovered information on a particular individual.

**Confidential Response**

I think the Census Bureau is over-reacting somewhat. This is a somewhat complicated files. Most users of it are professionals.

**Christine Holt, Richmond Regional Planning District Commission, Government (Nonacademic)**

Sure anything is possible, but the chances of that are slim.

**Leonard Gaines, Empire State Development (NYS SDC LEAD), Government (Nonacademic)**

The dataset is fairly well protected for confidentiality as is. Some small amount of local "data switching" might be an acceptable way to add some noise to the data to protect confidentiality while keeping a useful dataset.

**Duane Marble, The Ohio State University, Geography (Faculty)**

I would strongly favor the type of approach that is used by the UN's Latin American Demographic Center in Santiago in their REDATAM software. Anonymous individual records are held in groups identical to small census reporting areas (e.g., our block groups or tracts) and the software permits reaggregation of these geographic units as the investigator wishes. This eliminates a LOT of statistical problems! And privacy is maintained.

**Richard Rathge, North Dakota State University, Demography (Faculty)**

The large geographic levels keep confidentiality for our state.

**Ronald Cossman, University of Central Arkansas, Geography (Graduate Student)**

It is unlikely that specific individuals could be identified using this data set, especially given the commercial data sets that are readily available (and much superior) for that purpose. However, if confidentiality remains a concern of the Bureau, a confidentiality agreement could be made part of the use agreement. Researchers often work with such restrictions on release and publication.

**Lisa Dillon, Institute of Canadian Studies, University of Ottawa, History (Faculty)**

First, it is crucial that the PUMS of the 2000 Census be constructed in such a way that it is compatible with previous PUMS in the IPUMS series. The power of an individual microdata set is increased exponentially when used in conjunction with previous microdata to produce a time series. The best research-based policies are informed by knowledge of long-term historical trends, and we simply cannot research such long-term trends if the 2000 Census PUMS is considerably different from previous PUMS. DO NOT underestimate the importance of long-term historical research for understanding the dynamics of poverty, race relations, sex discrimination, migration and aging, all issues which continue to affect the American population. Second, it is crucial that a fine level of detail is preserved in all variables, ESPECIALLY occupations, in order to maximize the abilities of IPUMSi and IMAG to integrate these microdata with parallel international data such as the 2001 Canadian census. Third, since, in 72 years, confidentiality rules will no longer apply to the 2000 Census, the U.S. Census

Bureau should preserve for that time a raw version of the 2000 PUMS to which \*no\* changes have been made. Finally, the U.S. Census Bureau should directly involve members of the IPUMS staff in planning and executing the PUMS of the 2000 Census.

**Diane Hite, Mississippi State University, Economics (Faculty)**

If PUMS is used for academic research there should be no confidentiality problem. If it is a concern there should be some means to punish breach of confidentiality, i.e. require a specific contract.

**Craig Olson, Univ of Wisconsin-Madison, Economics (Faculty)**

Could use a procedure such as that which is used with the NLS datasets when research request access to gain and use detailed geographic data.

**Confidential Response**

People do not realize that the Census information is actually more confidential than information gathered on the average web page, but they have no qualms about that...

**William O'Dell, University of Florida, Other Academic (Faculty)**

I have heard apocryphal stories about the identification of individuals through complex PUMS crosstabulations. If true, it seems more reasonable to provide resources within the Bureau to discover those unique situations and solve the problem for individual PUMA's rather than severely blunt such a valuable tool - there have got to be more creative solutions than those under consideration.

**Roy Williams, Mass. Inst. for Soc. & Econ. Research/Umass/Amherst (State Data Center), Government (Researcher)**

I think it would be rare but possible that someone would use PUMS to identify the characteristics of specific individuals. I appreciate that the Bureau has to take this into account.

**Stephen Greaves, Louisiana State University, Other Academic (Graduate Student)**

I don't understand why there is this belief that increasing the detail in the PUMS data will suddenly lead to big brother knowing our business. Ask any credit card company, telemarketer etc and they probably know more about our habits than we do - this to me is the real invasion of privacy. The PUMS data is the greatest source of disaggregate data for critical research in my field (transportation planning and policy investment). By bowing to some short-sighted politicians this country is shooting itself in the foot and holding back research and growth.

**Louis Kyriakouides, University of Southern Mississippi, History (Faculty)**

Privacy and confidentiality are important, but anyone familiar with PUMS data knows that identifying a specific individual is impossible.

**Confidential Response**

PUMS data is strikingly LESS useful in identifying the characteristics of specific individuals than many other sources of information. If I wanted information about individuals, I wouldn't start with PUMS data, I would start with internet "spy" software, or state, county, and city resources, even the local phone book.

**Don Burrell, OH-KY-IN Regional Council of Governments, Demography (Nonacademic)**

Insuring the privacy and confidentiality of Census respondents is essential. As the value of the PUMS data is in the subject detail, the geographic detail must be broad enough to accomplish this.

**Ted Mouw, University of North Carolina, Chapel Hill, Demography (Faculty)**

If there are concerns over confidentiality, then prepare two PUMS samples. One with small PUMAs and less demographic detail, and the other with large PUMAs and the same level of demographic detail as the 1990 PUMS 5% Sample.

**Confidential Response**

Census tract data could have been used to identify characteristics of specific individuals before this. Was that considered? If someone really wants to target specific people/individuals, the data is already there to do so.

**Sara McLafferty, Hunter College, Geography (Faculty)**

I really think that this is a non-issue. I cannot imagine being able to trace individual respondents from a PUMA population of 100,000 people, especially given the categorized nature of much of the PUMS information. I don't see why anyone would even attempt to do this, and if they did, it would be virtually impossible to accomplish.

There are commercial data sets that provide much more fine-grained, personal information, at much more detailed geographic scales -- anyone who wants personal data about someone can easily get it from commercial sources.

**Edward Schafer, San Diego Council of Governments, Government (Nonacademic)**

Why not announce a "hot deck" swap of a limited number of records, where the probability of swapping is positively correlated with the uniqueness of the record. That would reduce the chances of identifying unique responses.

**Jack Solock, Center for Demography and Ecology and Center for Demography of Health and Aging, University of Wisconsin-Madison, Demography (Support)**

I think that this, like many other Census based issues, is a politically driven issue, especially in an election year. Detailed PUMS records have been available for several Censuses and this has never been a problem before. All of a sudden this becomes an issue. PUMS records are set up already so defining individuals is almost impossible. What is going to happen if this goes through is that an important part of American social history is going to be lost, and future generations will see that this was a silly, stupid politically motivated decision.

**Stephen Wilson, SRF Consulting Group, Industry (Nonacademic)**

Most of us have better things to do with our time than chase down a particular individual, regardless of whether it violates any professional ethic.

**David Kruse, Alamo Area Council of Governments, Government**

In my nearly twenty years of experience using census data in general and PUMS data in particular, I have never seen or even heard of anyone attempting to identify specific individuals.

**Joyce Benson, Maine State Planning Office, Government (Nonacademic)**

If the geographic breakdown is the key to ensuring confidentiality, why not provide full detail for every state at the state and metro/non-metro level?

Then data for the smaller geographies could be put out in a more condensed version that groups variables into fewer categories or intervals.

**Jerome Deichert, University of Nebraska At Omaha, Other Academic (Faculty)**

The problem that I see is outliers in the 100,000 size puma. I think if values included in the PUMS file are bounded at a reasonable level, there should be no problem.

**Tammy Colson, History (Researcher)**

I believe that it is extremely difficult to identify a particular individual by the information provided in this database. This information is precious to many researchers in many varying fields and the decrease of such information would be detrimental on a broad-based spectrum.

**Keith Lawton, Metro, (MPO in Portland, Oregon), Government (Nonacademic)**

There are far better sources of information available in the private sector of interest to commercial applications (credit card data etc.) It would be extremely time consuming and inefficient to use PUMS for this purpose. The lack of real geographic specificity guards against most invasions of privacy.

**Pamela Klein, Penn State University, Sociology (Graduate Student)**

From what little work I have done using PUMS, I cannot possibly see how a specific individual could be identified through this data. Anyway, my field is not concerned with specific individuals.

**Patrick Costinett, Parsons Brinckerhoff, Industry (Nonacademic)**

Other data providers take the approach of masking some fields if there is any chance that individual persons, households or firms could be identified. This approach would be preferable to suppressing detail across the board.

**Chester Fenton, unaffiliated, Other Nonacademic (Nonacademic)**

Why anyone would use this data to try to identify specific individuals is unclear to me.

**Fred Wilcox, Claritas Inc., Industry (Nonacademic)**

As with any other Census data, confidentiality is of the utmost concern. One misstep now could have effects on data quality for many decades to come.

Having said this, I am not aware of any concerns about the PUMS data other than what has been expressed by the Census Bureau. Surely, if there was a potential problem with the confidentiality of the microdata available from the 1990 PUMS files, someone would have noticed.

**Michael Gillett, Oregon DOT, Government**

The data is very aggregate with all confidential data removed. Based on that I don't see a privacy issue.

**Nicolas Sloss, UC Berkeley, Other Academic (Graduate Student)**

The PUMS would never be used to identify individuals! Who are we kidding?

**John Pitkin, Analysis and Forecasting, Inc., Demography (Nonacademic)**

Detailed occupation seems the most likely characteristic to enable a breach of confidentiality because there are public databases of unusual occupations (yellow pages, state licenses). Since accurate public data on age or ancestry are not available, it would be impossible to prove a connection between a PUMS record and a specific individual based on either of these characteristics. (Suppose you found a PUMS record for a dentist of Armenian ancestry and found only one dentist with an Armenian surname in the yellow pages for the area. You still would not be sure that there is not another dentist with an Armenian mother and non-Armenian surname.

**Irene Browne, Emory University, Sociology (Faculty)**

I do not see how individuals could be identified with the PUMS data, given that the geographic detail is so broad. Isn't this an empirical question? Has someone at the Census Bureau successfully identified a PUMS respondent? If so, perhaps there could be "missing geographic data" on respondents whose characteristics are SO unusual that they could be identified.

**Frank W. Sweet, American Military University, History (Graduate Student)**

It is a 1/100 sample. The chances of finding the same individual person in two data sets would be one in 10 to the 4 -- less than the chance of being struck by lightning. The privacy concern is overblown. Perhaps they could remove people's names and house numbers (in address) to make the data appear less individualized.

**Jonah Gelbach, University of Maryland at College Park, Economics (Faculty)**

I understand why members of the public are concerned, but it is because they do not know the details of the confidentiality requirements placed on researchers (including the large amount of attention given to confidentiality by Census before it releases public use data).

I absolutely could not care less who is in the data, as long as I know the important demographic characteristics described above.

**Blaine Billings, Boise State University, Sociology (Undergraduate)**

Name, Social Security number and address information can very simply be tagged by computer engineers as operational files that cannot be accessed, except for duplication of information search in the data cleanup process. Once that is achieved assigning a persons data to a different data code would allow for geo/socioeconomic tracing

**Andrew Harvey, Saint Mary's University, Economics (Faculty)**

It is basically a red herring.

**Benjamin Ben-Baruch, StarWorks, Inc., Industry (Nonacademic)**

The ability of researchers to merge databases and to match confidential data with individuals is very real.

**Bill Godair, Old Dominion University, Other Academic (Graduate Student)**

Unfortunately, I can foresee aggressive Marketer attempting to identify characteristics of specific individual to "tailor" their marketing. However, I also believe that it is possible to protect the confidentiality of specific individual characteristics.

**L. W. Shell, Nicholls State University, Other Academic (Faculty)**

I disapprove of most work- and income-related questions on the census form, except those that create the broad occupational categories. They are too intrusive.

**Confidential Response**

It would be exceptionally difficult to identify any person(s) in the data on the PUMS system, particularly because the number of variables that you can request at any one time is limited by space and technical considerations. This necessarily reduces

any chance of obtaining an 'identifier' variable that could distinguish one person from many given that the amount of linked information to any one case would likely produce the same profile among a reasonably large segment of the population.

**Kenneth Sunderland, University of California - Riverside, Economics (Graduate Student)**

It seems unlikely that individuals could ever be identified by means of the PUMS. Even supposing such an unlikely possibility, it would be extremely difficult and expensive in time and effort. The product of my own research precludes this possibility, since it does not report individual statistics.

**Wendy Gordon, Plattsburgh SUNY, History (Faculty)**

I frankly do not understand why the 70-year confidentiality rule for censuses before 1940 was not applied to later censuses, and would have no problem with not releasing individuals' names until they could be expected to be deceased.

**Jill Marlette, , Geography (Nonacademic)**

I can't imagine anyone in my field of market research would use this data to invade the privacy of any individual. We're interested only in groups of people, their geographic location, migration patterns, population expansion, industry, etc.

**Charles Starnes, Oregon State University, Sociology (Faculty)**

Concerns about privacy or confidentiality in PUMS data are reasonable but risks to privacy and confidentiality via PUMS data pales in significance to risks arising from other sources. In this era of electronic information flow, PUMS data rank nearly dead last in any informed list of sources where one can find information that might conceivably invade privacy or confidentiality. Anyone who thinks he/she needs PUMS data for such invasive purposes is grossly ill informed. The benefits of more, rather than less, refined PUMS data far outweigh any risks to privacy or confidentiality.

**Paul Voss, University of Wisconsin-Madison, Demography (Faculty)**

Since the PUMS is built from sample records, suppose you find a record that looks like someone you know. How do you know it's not someone else? And who cares? I do not know of any documented case where someone misused the PUMS data to attempt to identify individuals.

Also, I've forgotten whether or not the 1990 PUMS file was created from the file that contained the Confidentiality Edits (swapped records). If so, this is one further bit of protection.

**Jason Fields, U.S. Census Bureau, Demography (Nonacademic)**

Title 13 confidentiality is the most important aspect of collecting data, certainly for a federal agency. The power of computing resources has brought the ability to spin data into the home, and has increased the importance of protecting confidentiality in all of the data we collect. There is likely not much more risk of breaches in confidentiality in the 2000 data than there was in the 1990 data. One option (unofficially) would be some type of registration and limited access to the data to allow academic research to continue with a more complete data file, while a more protected file could be released for broader use, although this has severe implications for the IPUMS.

**Leslie McCall, Rutgers University, Sociology (Faculty)**

As someone who has been working with the PUMS data for seven years or so, I find the concern for privacy and confidentiality ludicrous. It is impossible for a researcher to determine individual identity and, perhaps more important, there is *\*no reason\** for a researcher to be interested in this. We want as many "cases" as possible in order to establish "patterns". Singling out individuals goes against the very nature of research using census data, which is to *\*generalize\** from the data.

**Dunbar Brooks, Baltimore Metropolitan Council, Government (Faculty)**

PUMS data are statistically significant, but at the 100,000 person level in our metropolitan area, there is no way to identify an individual. Further, all of the researchers that I have encountered during the past 25 years (in our area and across the country: the number is significant) are not the least bit interested in identifying an individual. Rather, they are interested in the volume of data when the cells of a crosstabulation matrix based on PUMS. PUMS data, even if one could identify an individual is meaningless. No one makes policy based on one individual, there is no statistical significance.

**Confidential Response**

As a beginning researcher, I am not able to track any individuals on the PUMS sites. Further, I am not interested in individual persons, but the aggregate of my chosen variables. I believe my fellow graduate students who work with the site have no interest in individual names, thus helping maintain privacy and confidentiality.

**Peter Gttschalk, Boston College, Economics (Faculty)**

There is little balance in the current preoccupation with confidentiality in data sets. The government goes to huge expense to collect data and then severely restricts its use in the face of a miniscule probability that a specific individual might be identified. If the argument is that people won't respond to the Census if they think that they will be identified then we end up in the odd situation in which we assure response but then don't use the gathered data.

**David Olive, Southern Illinois University, Other Academic (Faculty)**

If the census is worried about confidentiality, would they be willing to make part of the data available immediately and part of the data available in 5 years (say)?

**Dr. Gene George, Butler County Community College, Demography (Support)**

Researchers such as myself have little time to wonder which particular individuals may have provided information to the Census, let alone do anything to compromise a respondent's privacy. Moreover, we're bound by professional ethics not to invade an individual's privacy. We are focused on our job, which is to provide the best possible analysis of aggregated data to provide the best possible understanding of what's going on in society at large. There are already too many "blank spots" in the population history of the United States. Reducing the information available in PUMS will only hurt the capacity of government and industry to manage resources effectively.

**Michael Stoil, Conwal Incorporated, Industry (Nonacademic)**

Most of the key data available through PUMS is based on the long form collection of data from 1 in 6 households. It therefore is not feasible to assume that the data description for a single household is defining of one household when it is statistically possible that 6 households in the track meet identical criteria.

**Confidential Response**

I wrote a review of a recent book on such issues. It is ridiculous to believe that academic researchers would try to do identify individuals -- what would be the point (all, or almost all academic research using the PUMS is concerned with moments of the distribution, such as the mean). Commercial researchers have far better ways to identify individuals than the PUMS.

**Mary Kritz, Cornell University, Demography (Researcher)**

I do understand the importance of maintaining privacy and recognize that the subgroups on which I do research would be most vulnerable. I am not aware, however, that the 1990 guidelines resulted in any breach of privacy and hope that the 1990 guidelines can be maintained.

Making it easier and less costly for researchers to access the confidential files at the Census Bureau would be the only real option.

**Douglas Gurak, Cornell University, Demography (Faculty)**

The current system might already be considered "overkill". Suppressing specific items for specific geographic areas when the odds of individual identification are still microscopic already guarantees confidentiality. I very much value confidentiality, but reducing information further is not needed and would be highly damaging to efforts to monitor trends in U.S. society.

**Ken Blum, Center for Labor & Community Research, Other Nonacademic (Nonacademic)**

I'm sure it is impossible to find data on any individual in the data sets I used.

**Confidential Response**

If it is necessary to collapse categories, the Bureau Director should reappraise the 72-year rule for release of the original data. Even if we cannot access the complete records for 72 years, the uncollapsed categories could presumably be released in a much shorter interval.

**Bruce Western, Princeton University, Sociology (Faculty)**

In years of research, attending conferences, discussing with colleagues a wide variety of scientific issues, and working in a demography center, I have simply never heard of any researcher trying to identify individuals with the PUMS or any other major social survey data source. I would imagine that if someone was trying to obtain census-style information about a specific individual there would be many more efficient ways of doing this than through the census. Thus, I tend to think that concerns about privacy exceed the real threats to privacy posed by the PUMS.

If the PUMS does carry threats to privacy, these seem much more likely to originate in the private sector than the scientific community. Perhaps concerns about privacy could be addressed by delaying release of the data to commercial users, while following the normal timetable in releasing data for scientific purposes.

In the spirit of current trends in criminal justice policy, perhaps privacy concerns can be allayed simply by imposing harsh penalties on those who would use the census to identify individuals. This focuses the cost of malfeasance on the perpetrators, rather than penalizing the entire scientific community. If the new penalties were well-publicized, the large deterrent effect would doubtless eliminate violations of privacy.

**Confidential Response**

Confidentiality must be respected, but data access cannot be sacrificed as a result. There are ways to achieve both goals.

**Beverly Daly, Kentucky State Data Center, University of Louisville, Demography (Support)**

Make it available on cd-rom only as it was in 1990. Don't release it on the Internet and don't make it "too easy" to use. If there haven't been problems in the past, why borrow trouble?

**Confidential Response**

We do not care WHO is at a particular geographic location, but the characteristics are essential.

**Brian Cushing, West Virginia University, Economics (Faculty)**

It is hard to imagine how I would ever be able to extract personal information from the PUMS data. Perhaps there are a few cases where this can happen, requiring some marginal changes. If such changes are necessary, they should make as few changes as necessary, and try to minimize the negative impact. They should consider what information PUMS provides that is not available elsewhere versus information that someone could easily extract from some other source. They should also try to gauge which details are not really used for basic or policy research, e.g., for example maybe there is room for some reduction on the level of detail on race or ancestry. Most importantly, they should be sure there really is a problem so that they do not try to fix something that wasn't broken in the first place.

**Eberts Randall, W.E. Upjohn Institute for Employment Research, Economics (Nonacademic)**

I have used CPS, SIPPS, and LRD, and I do not believe that confidentiality can be violated with a detailed PUMS data set.

**Confidential Response**

In regard to question #11, through the use of selection criteria and cross-tabulation, individuals may be identified if the Census Bureau does not keep vigilant in regard to the privacy issue.

**Roger Avery, Brown University, Population Studies & Training Center, Demography (Faculty)**

Require users to sign statements that would require them not to violate privacy and impose severe penalties on those that might seek to do so for any reason.

**Confidential Response**

Is the amount of detail that will be available for public use increasing from previous data releases; if not, what prompted making this an issue now. If it is increasing the level of detail, we should only be concerned with those variables that could identify individuals. We need to keep what we have been using.

**Heidi Wong, Asian & Pacific Islander American Health Forum, Other Nonacademic (Nonacademic)**

Anything is plausible now a days, with technology being the way it is...but I do not know enough about the safe guards that were used in 1990 and their effectiveness. Is the question whether or not to put the PUMS online? I think it is desirable, however, if it means that the level of detail is reduced, I would rather just have the PUMS issued on CD in as much detail as available in 1990, or even more! 1990 PUMS was very useful, but my only suggestion would be to format it in such a way so that the end user doesn't need to spend a lot of front end time processing the data as we had to for 1990 PUMS.

**Confidential Response**

In the wrong hands, good data could always be misused. But the answer is not to get rid of the data, but to monitor the usage. There has to be a better way to protect confidentiality than losing precious detail. I won't claim to have an answer and I do realize that the issue is important. It may not be as cost effective to take great pains to protect individual confidentiality, but I would plead that that is a better option than decreasing detail in data.

**Cynthia Rogers, University of Oklahoma, Economics (Faculty)**

I suppose that someone marketing type of companies might think it useful to track down specific individual characteristics. However, it doesn't seem that the time and effort involved would make it profitable to do so. Researchers certainly don't have the time to do such a thing.

**David Pearson, Texas Transportation Institute, Other Academic (Researcher)**

Privacy is a moot issue in 99.9 percent of the cases. I would challenge the so-called experts to produce a case where they could identify any individual or illustrate how the information provided could be used to harm them.

**Confidential Response**

Our work would not begin to get into this issue. We never want to reveal confidentiality. We need groupings on race data to use for planning and comparisons purposes; also some income groupings. Again, confidentiality could be protected without compromising such groupings.

**Confidential Response**

I don't see why this is an issue. The numbers of people within even the most detailed occupation are sufficiently large to keep people from knowing who people are. The way to handle confidentiality issues are to keep the geographic areas sufficiently large, as in previous years. Aggregating occupations would essentially destroy the ability of people in my subfield being able to look at sex differences in occupational attainment. Even the current classification scheme isn't detailed enough, so going in the other direction to aggregate further would be a disaster.

**Donald Hirasuna, House Research Department, Economics (Nonacademic)**

The data does not uniquely identify individuals. In looking at income tax data, I think the minimum number within a cell is 14 persons. You could make simply write na for cells with fewer than 14 observations. Or, you could re-weight the sample within each PUMA to meet such small cell sizes.

**James Kinard, Government (Post-doc)**

I can not understand this issue. The purpose is to identify population characteristics, not individuals. This is an issue of proper methodology design that is supposed to ignore any personal identification. Confidentiality is extremely important and must be strictly adhered to.

**Confidential Response**

Given the low level of geographic detail typically available in the PUMS, the whole idea that a researcher would or could seek to identify information for a known respondent is simply absurd. Individuals' privacy is already adequately protected in the PUMS.

**Richard Ogburn, South Florida Regional Planning Council, Government (Nonacademic)**

The assurance of confidentiality is critical to the long-term mission of the Census Bureau, as the primary provider of survey data on the socio-economic characteristics of the population. Maintenance of the public perception of the ability of the Census Bureau to guarantee this confidentiality always should prevail over other objectives.

**Marianne Hill, Mississippi Center for Policy Research, Economics (Nonacademic)**

Our obligation to help our children is strong, as is our obligation to fight poverty. The occupations of mothers by race is needed to talk about solutions. There may be some trade-offs involved, but I have never heard of any individual suffering as a result of a loss of privacy through Census data. But I will add that training of professionals using data such as Census data should include discussion of privacy issues.

**Alicia Sasser, Harvard University, Economics (Graduate Student)**

Given that any identifying information is held confidential for nearly the respondent's lifetime, there is little possibility of following an individual longitudinally. Given the wealth of personal information available on the web, the PUMS seems like it would pose far fewer risks to individual privacy.

**Deborah Garvey, Santa Clara University, Economics (Faculty)**

Given the aggregation of population into 100,000 member PUMAs since 1980, it is impossible to identify individuals on the Census; to suggest otherwise is preposterous. If privacy and confidentiality are truly a concern, require researchers to sign a confidentiality agreement prior to use of PUMS data. Scholarly researchers have no interest in revealing respondents' personal information that might jeopardize people's faith in the confidentiality of the Census and further reduce response rates.

**Philip Wuest, Clark County, Washington, Government (Graduate Student)**

Privacy and confidentiality are very important issues, and I can understand why there is concern about making detailed information widely accessible. On the other hand meaningful research requires that detail, so there is a balance to reach. Often in my work is it essential to have detailed individual information for model development, but results used for policy analysis rarely present specific details.

**Judy Olson, Michigan State University, Geography (Faculty)**

I am extremely respectful of confidentiality. I believe it can be protected. I am not sufficiently facile with the data yet to know how it might currently be used to identify characteristics of specific individuals (though it would probably be easy to use it to increase the probability of finding someone with a known specific characteristic).

**William Mead, Gloucester County New Jersey Planning Division, Demography (Nonacademic)**

Severe financial penalties for any individual or entity found in violation AND lifetime prohibition from utilizing PUMS data.

**Kate Kelly, University of Central Arkansas, Other Academic**

The census bureau is a government agency and must follow the Freedom of Information act. They have a choice of sharing the information with PUMS or handling the questions on an individual basis. The information will be needed in research, if they share, they cut their workload; if they don't share, they increase their workload, considerably. We do not need personal information as to who someone is, but we do need the demographics. What else is the census really good for (representation in Congress???)

## Responses to Question 13: Other Comments.

### Confidential Response

I've only begun to explore the PUMS. It provides a unique opportunity to observe a representative sample of the U.S. population at different points in time. Right now, the 1960--1990 data are mostly comparable. It would truly be a shame if that were not true of the 2000 data.

### Alair MacLean, University of Wisconsin - Madison, Sociology (Graduate Student)

Keep up the good work!

### J. Wesley Null, The University of Texas at Austin, Other Academic (Graduate Student)

IPUMS is a valuable source for colleagues throughout academia, and the level of specificity should not be changed.

### Robert Miller, University of Michigan, Other Academic (Graduate Student)

You have an html problem with question 6 & 7. You cannot answer each item separately. I have no opinion for item set 5.

### Matthew Sobek, University of California, Riverside, History (Researcher)

I know of no instance in which the privacy of any person has been violated using the PUMS. The confidentiality criteria applied to the 1980 PUMS seem entirely adequate. The impetus for stricter criteria seems like a solution to a problem that does not exist.

### Confidential Response

On occupational categories: There is absolutely no adequate compromise, at least as far as my research is concerned. Many of the categories are already far too broad to be meaningful as social units. Also, see comments to question 10.

Generally: If the government is truly concerned with confidentiality, and not just with saving money, it should look first at cookies! I would bet that people disclose far more identifying information about themselves in the course of surfing the net than they do in filling out the long form. And, unlike academic researchers, internet companies and web site developers are not constrained by professional codes of ethics or institutional rules against profit-mongering.

### Peter Meyer, Northwestern University, Economics (Graduate Student)

I value IPUMS very highly. I am impressed by its continuing progress.

### Sarah Hayford, University of Pennsylvania, Population Studies Center, Demography (Graduate Student)

I have only used the PUMS data once, for historical work (using 1910 Census data). I'm therefore not very familiar with the issues involved with more current census data.

### Confidential Response

This survey should allow respondents to answer more than one category.

### Lisel Blash, Public Research Institute/San Francisco State University, Other Academic (Researcher)

It seems that the real interest in limiting the data has little to do with confidentiality and much to do with saving money and promoting a political agenda. The whole anti-census issue seems a political ploy to rally public opinion against "big government" by attacking a relatively defenseless target, meanwhile limiting the ability of pesky social science researchers to ask important and difficult questions.

### Joseph Doherty, UCLA School of Law, Other Academic (Researcher)

PUMS should not compress the top age brackets to 65+. The population is aging and more people are entering that bracket, not fewer. Surveys of California and Arizona voters indicate as many 65-75 as 75+. There are significant differences between the two groups.

### Confidential Response

I understand that The Bureau is obliged to protect confidentiality under a de jure definition, not a de facto one, but this seems really dumb. The detailed data allow future researchers to do analyses that we have not thought of yet; why ruin it for them?

### Confidential Response

PUMS is the way for the public/academia to do research freed from the shackles of the government deciding what crosstabs are of interest. This data is useful immediately and for decades to come, so don't make a shortsighted decision that will impair us in areas that we might not even be able to imagine yet.

**Rebecca Sandefur, University of Chicago, Sociology (Graduate Student)**

It saddens me that many citizens seem not to realize the DAILY benefit they receive from policies formulated using information from Census data.

It also amazes me that people who will happily allow unknown corporations to track their internet reading, electronic shopping behavior and spending patterns are worried that egg-heads like myself will try to find out where they live.

**J. David Hacker, California Institute of Technology, History (Faculty)**

The PUMS data have been available for years. The history of PUMS research suggests that individual privacy is a non-issue. It is simply not plausible that low density sample data with missing names and geographic detail will ever be used to identify the characteristics of specific individuals.

**Confidential Response**

Your questions do not allow for much in the way of explanations. For example, I've only used PUMS 2-5 times in my own research but I'm a new Assistant Professor. If I've needed it that many times so far, how much more work will I do with it over my career?

**David Clark, Marquette University, Economics (Researcher)**

The quality of research that can be conducted from PUMS data continues to grow as other spatial data sets become more readily available. Thus, using GIS tools, in conjunction with PUMA's allows for much stronger matching of spatial characteristics (e.g., amenities, employment growth, etc..) to PUMS data. If you diminish the spatial or occupational/industrial detail of the data, you reduce the value of this database.

**Patricia Kelly Hall, University of Minnesota, History (Graduate Student)**

If there is any threat at all to the American people from detail in the PUMS, it is puny. But the research gains from the anonymized individual level data in the PUMS are tremendous. Access to these data has produced research that successfully challenges the findings of work done only on aggregate data. It is essential for sound policy and planning research for the future -- as well as time series analysis of the past -- that no more detail be lost in the PUMS. It would be a serious and foolish mistake of the Census Bureau to put so much invaluable research at risk for so little gain.

**Joy Williams, Environmental Health Coalition, Other Nonacademic (Nonacademic)**

Thank you for your conscientious attention to the quality of census data.

**Confidential Response**

PUMS is indispensable for the microeconomic research. The more PUMS provides data, the more productive our economic research becomes.

**Lawrence Nitz, University of Hawaii at Manoa, Political Science, Other Academic (Faculty)**

When I think of my own privacy, I am more bothered by the fear of unauthorized access to the Safeway POS files-- somebody might find out how much of which brand of beer I buy. More seriously, I venture that we are all harassed more by telephone marketers whose interview schedules are derived from matching block group income to phone book listings.

Finally, I have never heard of a Census PUMS user identifying or seeking to identify a person in the file. Learning the PUMS is far too much work to go through for meaningless curiosity, commercial exploitation, or private investigation.

There is the prospect that some data in the PUMS could embarrass political interest groups. I am sorry--inequitable distributions and miserable governmental performances at the level of the PUMA, the blockgroup and the block deserve to be reported and corrected.

**Joelle Saad-Lessler, Columbia University, Economics (Graduate Student)**

The census data is the lifeblood of so much research, it is a shame to play politics with it. Already, the quality of the 2000 census is being compromised by not allowing statistical sampling. There is no need to tinker more with it, just to satisfy society's temporary panics, which are usually overblown and unjustified. Please help maintain the integrity of the census. We really need it!

**Reeve Vanneman, University of Maryland, Sociology (Faculty)**

Thank you for undertaking this survey!!!!

**Confidential Response**

It makes a lot more sense to manage who has access to the data than it does to refrain from collection. Attempts to further limit the information collected will almost certainly have costs over a number of decades, as we find that we are ill-equipped to analyze new policy issues that arise.

**Gabriela Galescu, Cornell University, Sociology (Graduate Student)**

I know it sounds paranoid, but to me it looks as if someone has an interest for people not to write papers on stratification and inequality. No data, no research, no pain. Who wants people to know that we don't live in El Dorado? Well, I do. And I'm not the only one.

**Huey-Chi Chang, University of Wisconsin-Madison, Sociology (Graduate Student)**

After all the resources and energy have been spent in the collection of census information, it would be a waste if more detailed information can't be made available to the researchers. New perspectives and new ways of analyzing the data are being developed every day. These new perspectives or methodologies may call for different ways of organizing the data. Preserving detailed information would make application of these new perspectives and methods possible. Future researchers shouldn't be punished by the lack of insights of contemporary researchers.

US government has also made such an effort to convince people how census results are going to affect their welfare. In response, people make a commitment by filling out the census form. Implicitly, people are willing to take certain risk, in the hope that it will help themselves or their groups in some way. If the data turn out to be not as useful as they have promised to be, it would be an ultimate betrayal to the people.

Finally, concerns on the breach of confidentiality should be backed up by actual evidence. Statistics on how many incidents and under what circumstances these incidents have occurred in the past should be used as guideline to justify how detailed the census information can be released to the public. It is also important to ask if the issue of confidentiality affects the majority of the people or if it is only the concern of certain interest groups. I can't help but to construe the issue as a political one if protecting the confidentiality of a few industries is in conflict with obtaining knowledge that would benefit the million.

**Charles Wetherell, University of California, Riverside, History (Faculty)**

The IPUMS provide historians and other social scientists with the most important data source for historical social analysis in the United States. Reducing the compatibility of the 2000 PUMS with previous releases would have profound short and long term effects for an expanding user population of social scientifically oriented research.

**Robert Adelman, University at Albany-SUNY, Sociology (Graduate Student)**

Without detailed PUMS data, researchers will be unable to complete thorough investigations of a variety of social issues and problems. And, without data compatibility across time, the hard work of the IPUMS group and the promising avenues the IPUMS has opened for many researchers will be lost.

**Albert Monroe, Harvard University, Economics (Graduate Student)**

The PUMS is good the way it is currently constructed, and as long as we use decennial censuses, we should not reduce the detail currently found in the pums.

**David Katz, Loyola University Chicago, Sociology (Graduate Student)**

Census data are an invaluable resource to social scientists, but maintaining historical comparability of the data is vital. Changing the way data is grouped or categorized severely limits researchers' abilities to understand changes in the American population.

**Carolyn Liebler, University of Wisconsin at Madison (sociology & demography), Sociology (Graduate Student)**

For the occupation codes, please do not combine them into something like a Duncan SEI. It conflates occupational education and occupational earnings. I suggest something like what Hauser and Warren have done: keep the two characteristics of occupations separate and report two things: 1) the proportion of people in that occupation who have completed schooling above some threshold, like one year of college 2) the proportion of people in that occupation who have an income above some threshold like \$30,000

These suggestions are ONLY in case you really really HAVE to reduce the detail reported for occupations. If you do reduce the detail, my work and the work of others will be stopped short. It will be a big problem.

I care even more about the race and ancestry questions.

**Confidential Response**

I think researchers would be in better defending grounds if they can make direct links between census research and policy questions and how answers to those policy-relevant questions can be affected by what is being proposed.

**Confidential Response**

I sincerely appreciate the efforts of everyone involved in maintaining the IPUMS website. It is something that every researcher who works with data gets around to using sometime during his/her career. With the expansion in the capabilities of computers, more analysis can be done. Limiting the data at the time would severely incapacitate the ability of researchers to provide sound, useful studies.

**Steve Holden, Johns Hopkins University, Economics (Graduate Student)**

Could not have done my dissertation without this data. The data should be in the public domain, but guards should be in place to protect privacy. This can be accomplished through legal contracts.

**Ernest Schreiber, Lancaster New Era, Media (Nonacademic)**

It seems to me that those with political agendas may be directing the Census Bureau in directions that its own statistical staff would not support -- if freed of political loyalty. No statistician would argue for aggregated data when specific data is available. It's insane.

**Cristian deRitis, Johns Hopkins University, Economics (Graduate Student)**

As a researcher of historical data, I am often faced with the problem that detailed statistics from earlier periods are simply lacking or are much cruder than modern data. My colleagues and I work around these issues as best we can but are often left with a degree of dissatisfaction. If the Bureau does choose to limit the data released, I wonder if researchers fifty years hence might not be faced with the same issues. I wonder what questions one might be able to ask with a long series of highly detailed data.

**Eugene Hammel, University of California Berkeley, Demography (Faculty)**

PUMS data are also very important in TEACHING across a broad range of fields, sociology, demography, economics, history.

**Anupam Tyagi, University of Utah, Economics (Graduate Student)**

In question 9. Please try to release less details sooner, but don't forget to release more details after a more informed decision.

Is it possible to include in the samples the information about response rates from various small areas if this doesn't conflict with privacy concerns?

Your team has done excellent work. The datasets are superb, the site is very user friendly and a treat for academic and policy researchers. If I can help in any way, please get in touch.

**Confidential Response**

Good luck at the meeting!

**Barbara Reskin, Harvard, Sociology (Faculty)**

Re question 7, there is no adequate proxy for the present coding system.

**Louis Chauvel, Institut d'Etudes Politiques de Paris, Sociology (Faculty)**

Please stop with confidentiality paranoia: ethics and deontology of information is not simply a question of individual confidentiality, but a balance between individual confidentiality and socio-statistical knowledge. Macrosociologists are definitively not interested by specific individuals but by social knowledge: let us know!

**William Collins, Vanderbilt University, Economics (Faculty)**

The IPUMS are a remarkable and unique dataset, forming the basis for a great deal of high-quality economic research. I hope that the Bureau values it accordingly.

**Richard Mueller, University of Maine, Economics (Faculty)**

Related to question 9 above: would it be possible to have the less detailed data released early, followed by the more detailed data at a later date? It would seem to me that this doesn't have to be an either/or decision.

**Caroline Hoxby, Harvard University, Economics (Faculty)**

Census data is collected at considerable expense in order that the United States make informed policy decisions. If the data are released in format that makes them unusable for policy analysis, then the expenditure on the Census is entirely wasted.

**Allen Glicksman, Polisher Research Institute of the Philadelphia Geriatric Center, Sociology (Researcher)**

The IPUMS is critical to so many research and policy efforts that making it more difficult to use would have serious and unforeseen consequences.

**Confidential Response**

With the advent of cookies in the internet space corporations are already tracking consumer behavior and patterns at a highly individual level. Most of the time the ordinary web surfer does not even know that they are being tracked in terms of their surfing patterns, interests, purchasing patterns and a host of other variables. It is a remote possibility that after having access to such personal level data, anyone would be interested in cracking aggregated and sampled data to reach personal profiles. However the quality of research would substantially be affected in the case that the level of detail were reduced in the PUMS. The high computing power available on the PC today is another factor that makes patrons demand better quality data even if it voluminous so that multiple models and or scenarios may be tested. Reducing the data detail would reduce the accuracy of predictions using this source and reduce comparability with previous datasets.

**Confidential Response**

Thank you for the good work.

**Olivia Mitchell, University of Pennsylvania, Economics (Faculty)**

There appears to be much current concern about provision of data to researchers. Nonetheless there is NO evidence that any researcher has ever created problems in this area. Let's not kill the golden goose of scientific information; rather if there are privacy concerns, what should be done is to devote more attention to developing mechanisms that bind the data user to responsible data use when necessary.

**Cheri Minton, Harvard University Sociology Department, Sociology (Support)**

I am not an "academic", but rather a programmer/analyst working in support of students, faculty and research staff in the Sociology Department at Harvard. I see projects depending on significant detail in all the areas you outline. I also see that the first question involves current data and the second invariably looks to comparison across decades. Comparability is extremely important. It should not however sacrifice keeping up with changes, such as developing occupational codes as new areas expand. I had been working with tapes of Pums data, and more recently the DES extraction feature making the 1990 5% sample accessible. Students introduced me to IPUMS, which provides an invaluable service. Regarding my "other" on occupations in question #7 -- Pums is THE place to get a sufficient sample at the detailed occupational code level -- eliminating this by providing a broader grained description would be a real shame and I simply could not choose a way that would be "best".

**Ellen Meara, Harvard Medical School, Economics (Faculty)**

The PUMS is probably the most widely used set of data across disciplines, and directly impacts what we know about overall trends in economic and general well-being of the US population. It would be devastating to reduce the detail.

**Susan Brudvig, Other Nonacademic (Nonacademic)**

We use the data for school districts' applications. Geographic detail is INDISPENSIBLE for local areas, as it impacts policy & financial decisions. We have also found socio-economic characteristics to be CRITICAL in accurately projecting school-age populations for small geographies (ie, our school districts). We need this data for local governments!

We also use this data to assess "brain drain" impacts. I can not rely to you how CRITICAL it is for the State of Indiana to be able to assess the situation in the year 2000, and compare it to our 1990 situation. We absolutely must have historically comparable data.

**Confidential Response**

We do not need 63 racial categories in the PUMS. Many of the privacy concerns could be addressed by reducing the number of racial categories, for example by combining the less common ones.

**Confidential Response**

As for the minimum occupational categories (#7), it is difficult to choose just one of the options. Different types of research would find one categorization scheme more useful than another. I, myself, have made use of typologies and SEI scores but have read studies that use occupation in all the ways listed.

Again, the great danger in minimizing the available detail will be that researchers will have to rely on other sources for their research--sources that may not be as representative of the population or as accurate.

**Darcy Hango, Ohio State University, Sociology (Graduate Student)**

I hope that the PUMS does not get more restrictive. As a graduate student preparing to embark on an academic career in sociology I cringe when I hear that a major data source like this may not be as fruitful for future research, especially on topics that involve the study of small group populations like Native Americans.

**Jamie Strickland, Western Kentucky University, Geography (Faculty)**

I am deeply concerned that misinformed politicians are at least partially responsible for this heightened concern over confidentiality. I draw this conclusion from some of the reports I heard prior to April 1, in particular, telling constituents not to fill out census forms over privacy concerns. In my experience, large-scale surveys such as the U.S. Census and the Panel Study of Income Dynamics currently go to great pains to maintain a high level of confidentiality without sacrificing much needed research detail. Furthermore, I feel that my ethics as a researcher are being questioned when these types of issues surface. Thank you for providing me with a forum on this extremely important topic.

**Eric Millis, Federal Reserve Bank of Dallas, Economics (Support)**

This seems to have begun as partisan political haggling, and I never imagined that such a trivial issue would be considered seriously. All social science work requires actual data, and the legislators who will ask for this data later must be made aware that their future inability to answer basic questions, if this is pushed through Census, will be due entirely to their own efforts. Policy cannot target problems it cannot see.

**Gregory Reeves, The Kansas City Star, Media (Nonacademic)**

The U.S. Census is the envy of the world. Its "special beauty" was to provide a fairly consistent portrait of every part of America, down to the census tract level. The Census is already completely confidential. Privacy concerns about it are misguided.

**Joshua Angrist, MIT, Economics (Faculty)**

In my view, the PUMS is the most important research data set in Economics. One of the most important features of the PUMS is that it is the only representative data set that allows us to look back into the pre-CPS period on a roughly comparable basis. Historical comparability was regrettably compromised between the 1980 and 1990 PUMS, for very little research or policy benefit. Any further changes should be made with an eye to preserving comparability of PUMS files as much as possible.

**Susan Losh, Florida State University, Sociology (Faculty)**

I will note that I found about this survey through AAPOR-net, which is a collection of (generally) survey and public opinion researchers. Please consider posting to AAPOR-net again as developments occur.

**Confidential Response**

I think the British geographers have had a somewhat similar problem and have got some "concessions" from their Census people. I am not sure of the specifics but some pilot trials were started with some type of sampling scheme which actually enhanced small area socio-economic-geographical analysis.

**Confidential Response**

I think these proposed changes would be disastrous for research.

**Confidential Response**

I wonder why detailed data are being collected if they are not provided to the scientific community for analysis. It seems a gargantuan waste of money.

**Robert Wolff, Central Connecticut State University, History (Faculty)**

The census data form not only the backbone of demographic research but provide invaluable details about the contours of American history. More and more scholars rely on the PUMS to inform their research, even if those scholars do not use the census data themselves. For the Census Bureau to limit what information would be available entails the arrogant premise that we know today what will be important decades from now.

**Jeff Claassen, Fort Worth Star-Telegram, Media (Nonacademic)**

The Census Bureau has been a great help to journalists and genealogists. It would be a terrible waste of highly valuable data if worries about privacy -- which should be respected -- cut off access to information that could never unveil an individual.

**Confidential Response**

To my mind, a neglected issue is the legal tension between the FOIA and the Privacy Act (and other privacy provisions like Title 13). Yes, everyone is quick to tell you, privacy act material is specifically exempted from FOIA. But this means ONLY that individually identifiable records can not be required to be released under FOIA. In court cases involving administrative records, a much much looser standard is regularly applied in releasing records than any statistical agency has ever used in creating a public use file. Although it is very likely that a Federal court would give great latitude to an agency in protecting privacy, it is still possible that a suit seeking broader public use files could be brought some day. Agencies should not assume that all the legal risk is on disclosure.

**Confidential Response**

The more data that better. We can always collapse ourselves.

**Andrew Beveridge, Queens College and Grad Center CUNY, Sociology (Faculty)**

I think the threat to individual confidentiality through the PUMS data is quite limited, since the data are actually quite benign. I understand the genuine concerns for confidentiality, but I hope that some accommodation can be reached.

The 1960 PUMS for New York is almost of no value for work dealing with NYC, because in 1960 the only thing that was released was Central City Status by State. So NYC is lumped with Buffalo, Rochester, Syracuse and Albany.

For the Census to be a viable means of analyzing the US and changes in the US the researcher should not be hamstrung by what are well meaning but what may turn out to be severely limiting data release restrictions.

**William Bridges, University of Illinois at Chicago, Sociology (Faculty)**

Being on a campus where the IRB zealots have been give free rein, I applaud you for resisting. There is a rumor that the pendulum is now swinging the other way in Congress, however. That is, swinging back towards more reasonableness in the balance between research and protection.

**John Blandford, University of Chicago, Economics (Post-doc)**

Many thanks for bringing this to my attention and for taking the initiative in addressing this matter. Please let me know if there is a need to make personal appeals to agency heads or elected representatives to help sway the Census Bureau from adopting the proposed policy.

**Michael Haines, Colgate University, Economics (Faculty)**

This seems to be part of a more general effort by Congress to reduce data going to the public. They do not seem to understand the "public goods" nature of this.

**Christopher Lyttle, Rush Primary Care Institute, Sociology (Researcher)**

It seems to me that question 9 presents a false dichotomy. If fact, it seems to me to be foolish to contemplate a single product. My true answer to this question is that I want whatever I can get ASAP, and I want as great detail as possible as soon as that is possible.

**Charles Brown, University of Michigan, Economics (Faculty)**

Re item 9, a sensible compromise might be to release a seriously gutted PUMS (acronym: LUMS--"limited use microdata sample") that would include only variables/level of detail that you're confident would not be regretted later.

Along the same lines, independent samples that gave detail on different variables would be better than one file which provided little detail on any file. So, e.g., one might give detailed geography on an "A" file with limited race/ethnicity detail and detailed race/ethnicity on a "B" file with only gross geographic coding.

**Confidential Response**

Please allow us to have these data--it is the "life's blood" of our work. You will stifle major academic research if this occurs. Similarly, policy analysis will be hindered as well.

**Carl Mason, UC Berkeley, Demography**

The IPUMS is a terrifically valuable dataset. As it is currently released it seems extremely unlikely that any data could ever be matched to an individual. Incremental (redundant) increases in data confidentiality could result in huge losses to science.

**Jon Lorence, University of Houston, Sociology (Faculty)**

If the Census Bureau is forced to reduce the availability of detailed information in the PUMS, I would hope the bureau would then resume producing all of the detailed special reports of the 1950s, 60s, and 70s which were discontinued with the 1980 Census. However, the purpose of the PUMS was to enable the Bureau not to have to publish those reports.

I realize the Bureau is under political pressure to protect the confidentiality of individual respondents. However, I found it ironic that those congressmen which vehemently oppose the level of detail in the decennial census are also the ones who are loathe to place restrictions on the use of data held by financial and medical institutions. Companies in the private sector already have considerable personal information on individuals which can be used to harm citizens.

**Catherine Weinberger, University of California Santa Barbara, Economics (Researcher)**

Ipums is an extremely valuable complement to the more detailed data that I usually work with. Having access to this resource often enables me to put the results of research based on a smaller subset of the population into a broader context. As a labor economist, the ability to match populations by income, age, race, and occupation is vital.

**Masao Suzuki, Public Policy Institute of California, Economics (Researcher)**

I didn't answer #11 because I really don't know. However, given the long lag time in getting the data out, what would be the point anyway?

**Mark Fossett, Texas A&M University, Sociology (Faculty)**

Thanks for your efforts. Please let me know if you are aware of avenues for getting further geographic detail in conjunction with PUMS data. Also, if loss of geographic detail does ultimately occur, perhaps provisions could be made for academic researchers to get relative easy access to special, restricted files that would provide greater geographic detail

**Simone A Wegge, Visiting Scholar, Hoover Institution, UC Davis Ag History Ctr., Economics (Faculty)**

Micro data are key to careful and thoughtful study of many issues in economics, demography, and economic history. Various authors have commented on this in journals like Historical Methods and elsewhere. The alternative is more aggregated data which can produce results that often do not illuminate the complexities of human behavior.

**James T. Ault, III, Creighton University, Sociology (Faculty)**

The products of the Census Bureau have been a national treasure for over 200 years. I would encourage the Congress to strengthen rather than weaken these resources.

**Confidential Response**

Or the Bureau could offer choices as to the sample size of the PUMS; allowing users to take a 50% or 30% or 5% sample. The smaller samples could have less detailed variable coding.

**Confidential Response**

These data are essential for public policy in a wide array of domains.

**Confidential Response**

I would like more information as to why there is a concern over confidentiality. Did something happen when someone was using for research?

**Patricia Becker, APB Associates/Southeast Michigan Census Council, Demography (Nonacademic)**

I will be at the May 22 meeting and I would appreciate having a report on this survey before I leave (i.e. by May 19). It will help me a lot in the discussion. If you have only raw data then and want to send me the data set, I'll be happy to throw it into SPSS and tabulate it myself.

**Richard O'Reilly, Los Angeles Times, Media (Nonacademic)**

Two releases of PUMS data would be acceptable, the first being an early release with data collapses and the second being a final release without collapses.

**Confidential Response**

I also think it would be disastrous to cut back on race variables -- while I have not done research specifically related to race, we are now at a time where there is increasing focus and interest in investigating experiences of specific races. In the health area, for example, it is now well known that there are differences in reactions of different races to drugs and dosages, so that it is vital to be able to study differences across races. We have finally reached a point where the importance of studying other

ances is widely acknowledged and it seems ridiculous that the chances of a confidentiality breach outweigh the enormous benefit from research that would result in being able to identify race.

**Tracy Dietz, University of Central Florida, Sociology (Faculty)**

For decades, social researchers have been calling for more detail, not less so that we can make better generalizations and predictions. Limiting the detail will serve only to limit our progress in the social sciences.

**Susan Thistle, Northwestern University, Sociology (Faculty)**

I have found the Ipums a wonderful and indispensable source of data for the big picture questions that I ask, and I frequently find myself thanking Steve Ruggles and the rest of the staff for creating this data set.

**Donald J. Bogue, University of Chicago, Demography (Faculty)**

The PUMS data have been comparatively under-utilized in past censuses because of the difficulty of access, and need for special computing facilities to handle cartridges. The 2000 census will greatly expand the potential for use, and for studies of change 1970-2000. It would be a major loss to have the 2000 data abbreviated so as to cripple what could be a major gain in research.

**Marie Mora, New Mexico State University, Economics (Faculty)**

I have been looking forward to the release of the 2000 PUMS because much of the research (and policy recommendations) based on the 1990 census (and there are volumes out there) is becoming outdated.

(Incidentally, my answers to the above questions regarding the reduction in the level of detail assume that SOME information on these items would still be provided. If these items are deleted altogether from the PUMS, it would be catastrophic for my own research as well as that of my colleagues.)

**Confidential Response**

The census is collected once every ten years; hence, analysts are familiar with the need to wait. Therefore, if waiting until 2004 or 2005 means it is possible a more extensive and cross-time comparable dataset will be available, I see no reason to rush release of the data. My preference also is for a historically comparable dataset; this is the nations archival process, and if we do not maintain a historically comparable record, it will not be maintained by any other entity. International comparability is nice, but virtually out of our control, for other nations can end the comparability at any time. Therefore, it seems most prudent to me to focus on historical comparability.

**Danielle Longerbone, Indiana University, Geography (Graduate Student)**

In response to question 8, I would like to choose more than one level of geographic detail. As a geographer, I use the data at many different spatial scales, so it is important to retain as much detail as possible.

**Hayward Derrick Horton, Dept. of Sociology, SUNY-Albany, Sociology (Faculty)**

A greater threat to confidentiality is the abuse of customer data by private companies and corporations. It is there that Congress should focus.

**Ross Stolzenberg, University of Chicago, Demography (Faculty)**

I serve on the Census Advisory Panel of Professional Organizations as a representative of the PAA. Please let me know more about this study.

**Michael Dennis, UC BERKELEY, Economics (Graduate Student)**

With regard to timing of release, would it be possible to release first a file of limited detail while consideration was given to whether or not, or which, variables would be included with more detail?

If that were impossible, I would prefer to wait for more detail.

**Larry Hotchkiss, University of Delaware, Sociology (Support)**

I don't use PUMS in current research. But, generally, I am very much opposed to reduced detail in these data. I hope you press the Bureau for evidence of abuse due to "excessive" detail in past releases of PUMS.

**Kathryn Hornberger, HealthFirst Management Group, Industry (Nonacademic)**

I have two thoughts about why this issue is even being discussed. (1) an attempt to reduce manpower hours, or (2) a political response to squeals from an ill-informed public who mistakenly believe that there is such a thing as privacy in this country. The sad thing is that rather than simply answer questions about privacy by stating the truth - that commercial demographers already know WHO WE ARE - government officials would rather eliminate data used by researchers to develop needed

programs and services - and all the while, the professional commercial demographers will continue gathering their explicit data - and eventually the rest of us will have no choice but to pay for access to it.

You are welcomed to use my comments, as I feel strongly about this issue. In fact, you have given me a great idea for an article on the subject of privacy and the availability of demographic data now in use in our country.

**R.S. Oropesa, Penn State University, Sociology (Faculty)**

It would be useful to have some background information on who and what is generating the attempts to modify the PUMS: Where is the pressure coming from? Is this pressure motivated by real or imagined fears that the PUMS data will be used to identify specific individuals? If real, how specifically has the PUMS been abused in the past to identify specific individuals?

**John McHenry, Demographic Data For Decision-Making, Inc., Demography (Nonacademic)**

My company and many others would certainly sign any reasonably phrased confidentiality waiver -- certifying that we would, under no circumstances, use PUMS data to identify individuals and/or their family/household members in any way, shape, or fashion.

**Joseph Persky, University of Illinois at Chicago, Economics (Faculty)**

Like jury duty, filling out the long-form of the census is time consuming. It would be reasonable to compensate households who are asked to undertake this highly useful task.

**Jeanne L. Anderson, Jeanne Anderson Research, Industry (Nonacademic)**

I have not had occasion to use PUMS for many years, but I did when I was at Met Life. I have answered the first few questions as per my present status. However, there are chances that I may need to use PUMS in future. That is why I am replying to this survey questionnaire. When I answered the questions on how I use the PUMS (now), I am answering for the project I conducted in the past.

**William Yancey, Temple University, Sociology (Faculty)**

The data sets you have generated have greatly facilitated our understanding of the historical changes in the social structure of the United States. What was at one time extremely difficult, if not impossible, have been made feasible and readily available.

**Jeff Burr, UMASS Boston, Demography (Faculty)**

I urge you to fight this nonsense!

**Ken Hodges, Claritas Inc., Demography (Nonacademic)**

I'll be at the May 22 meeting. See you there.

**Yoshinori Kamo, Louisiana State University, Sociology (Faculty)**

Please keep as detailed information as before. PUMS is critical for sociological/demographic research and our work should benefit the general population much more than a possible harm regarding identification, whose probability is very very small.

**Jim Hinterlong, Washington University, Other Academic (Graduate Student)**

Good luck on holding the line.

**Confidential Response**

I think some people in my profession might be interested in exploring other trade-offs, perhaps trading more aggregate socio-economic characteristics for more disaggregate geography ... perhaps mini-pumas of 25,000 to 50,000 persons?

**Confidential Response**

My focus is on the study of relatively small minority populations, for whom the PUMS is the \*only\* nationally representative source of data with sufficiently large sample sizes. Any reduction in data availability will make it more difficult to study minority populations.

**Alexander Bucur, USC/UCLA, Sociology (Researcher)**

We have learned from the past that badly modified/specified data have diminished the type of analysis were possible (e.g., analysis of race in the 1700s and 1800s is impossible due to no information in the interviews). Also, we have learned that we cannot compare periods since occupations or education were not properly recorded. By minimizing data records for 2000 not only we do not allow for researchers to properly analyze the 2000s, but we also throw away tedious and elaborate work carried out in 1980s and 1990s. Any analysis of those periods would have to be redone for simplified occupations or race.

As for comparing results from different publications, they may require many extra work from researchers to find the proper medium for comparison.

**Karen Norberg, National Bureau of Economic Research, Other Academic (Faculty)**

Best of luck in your endeavors.

**Sarah Reber, Harvard University, Economics (Graduate Student)**

The PUMS is very important for research in my field! Please don't take it away!

**Laura Leete, Willamette University, Economics (Researcher)**

The 1980 and 1990 PUMS have been invaluable tools in understanding a myriad of social welfare issues. I hope that we will be able to continue doing this kind of work.

**Confidential Response**

I only used this data set once, but it was absolutely integral for my research. I think it is very thorough and easy to use, and it would be very unfortunate were it not available for future research.

**Ed Christopher, Bureau of Transportation Statistics, Government (Nonacademic)**

I understand that this survey is geared to the academic researcher. However, pums has also been used in the field of travel demand modeling. Over the last two decennial censuses its use has been gaining momentum in this field, as has been its acceptance. The transportation planning community is an active user of census products including pums. To suggest changes without discussions in and amongst this community will only cause to alienate what has been a vocal political ally in the development of census products.

**Confidential Response**

Now that we have the technology to use the data extensively, it would be sad to learn that the Census Bureau has decided to make it less useful for researchers in various fields.

**Michael Spar, Weldon Cooper Center, University of Virginia, Demography (Researcher)**

I would like some statement (from the Bureau) as to why they feel changes in the PUMS architecture are necessary at this time. Was there some sort of breach of confidentiality based on either the 1980 or 1990 PUMS data? If such is not the case, then what is the rationale for tinkering with PUMS?

**Leonard Gaines, Empire State Development (NYS SDC LEAD), Government (Nonacademic)**

Generally a good survey, but too targeted to the academic setting.

**Richard Rathge, North Dakota State University, Demography (Faculty)**

The PUMS data files are vital for our research and our ability to assist policy makers in their efforts. We need to keep as much detail in the information as possible.

**Roy Williams, Mass. Inst. for Soc. & Econ. Research/Umass/Amherst (State Data Center), Government (Researcher)**

In our work as a State Data Center, geographic detail is more important than the kinds of content detail mentioned here. We almost always group age, income, etc. when providing data. Assuming the Bureau groups age with the standard break points used in their published tables & summary files, we would not be harmed by such grouping. I certainly hope the Bureau will group infrequently reported multiple-race combinations; that will facilitate rather than harm the kind of work we do here. Of course the devil is in the details of exactly how that grouping is done.

**Ren Farley, University of Michigan, Demography (Researcher)**

So far as I know, there has been no attempt to identify individuals using information from the 1940 through 1990 or through the CPS public use files. So long as some random changes are made in the data, a user could never be certain that he or she has identified a specific individual.

**Confidential Response**

People who think that census data in general or PUMS data specifically is useful for identifying individual characteristics just haven't used either.

**Willis Goudy, Iowa State University, Sociology (Faculty)**

While I have not worked much with the PUMS files, our research unit has received many inquiries about them. We've indicated where they are available and helped researchers access them. For our own work, we found that these files were the only ones available that would allow us to answer certain questions with state-level information.

**Don Burrell, OH-KY-IN Regional Council of Governments, Demography (Nonacademic)**

My last application of the PUMS data was to stratify household size by number of housing units in a structure. The tabulation verified my hypothesis that the fewer units per structure, the higher the household size; or apartment have lower household sizes than single family homes and big apartments have smaller household than >10 unit structures. The available PUMAS also allowed this comparison to be made for the central county and the outlying counties of the CMSA. I did this as part of estimating population based on residential building permit data which included # of units per structure.

**Jack Solock, Center for Demography and Ecology and Center for Demography of Health and Aging, University of Wisconsin-Madison, Demography (Support)**

I wish the Census Bureau would not feel compelled to be pressured by silly political partisanship in an election/decennial census year. The politicians could care less about the Census Bureau at any other time than this. Caving in to political partisan pressure will be seen in the future for what it was, a cowardly cave-in to political pressure.

**Don Larrick, Office of Strategic Research (the State Data Center in Ohio), Government (Nonacademic)**

Our concern is less with occupations than with industry classifications. We would like a scheme that uses new classifications and allows a regrouping for historical comparability.

**Richard Sherwood, Maine State Planning Office, Sociology (Nonacademic)**

I am willing to accept reduced geographic detail as a trade off for retaining greater subject matter detail.

**David Kruse, Alamo Area Council of Governments, Government ()**

After all this time in use, it is difficult not to feel that this is a backdoor attack on an invaluable data source by overzealous or anti-government data factions. Either that or someone is just fear mongering.

**Joyce Benson, Maine State Planning Office, Government (Nonacademic)**

The geographic groupings cover such a large number of people that I don't see how identifying an individual would be a problem, esp. since only about 6% of the population answer the long form in the first place.

**Jerome Deichert, University of Nebraska At Omaha, Other Academic (Faculty)**

If someone wants to find out information about specific individuals, I think there are far easier ways than trying to match records from PUMS.

**Claudia Goldin, Harvard University and the NBER, Economics (Faculty)**

Regarding occupation-industry classification -- why not a simple coding using 2-digit or 3-digit occupations and industries. The Duncan-SES should definitely NOT be used because it hides the underlying information and aggregates it in a manner that is particular to some period. It is NOT primary information; it is secondary and should probably be omitted altogether.

**Walter Treichel, Wilder Research Center, Demography (Nonacademic)**

In attempting to identify the risk to confidentiality, it is important to address both aspects of the risk, the possibility that it can technically be accomplished and the probability that someone would actually try.

**Confidential Response**

We are a planning commission, and on rare occasions, we get requests for data that can only be answered with PUMS. One of our more common requests for PUMS data is for persons by age, by number of bedrooms in dwelling units. This is sometimes used by committees planning school facilities.

**John Pitkin, Analysis and Forecasting, Inc., Demography (Nonacademic)**

Is this anxiety about disclosure a response to Senatorial and talk-show comments about the intrusiveness of the Census long-form questions? Whatever its source, it's much exaggerated. As a practical matter, private databases have so much of the same information and are so much more complete and specific, that no one looking for it would waste their time sifting through the PUMS for a 1-in-20 chance of finding a particular person's record. The most malicious thing someone might be able to do would be to sort throughout the PUMS for a small (100,000) population area, identify a handful of individuals with unusual combinations of characteristics and then, based on outside information, say, for example, that "it is very likely" that person so and so leaves for work at 8:45 or had over \$100,000 in income in 1999. You couldn't prove the inference. The

worst you could do is mail some catalogs for expensive merchandise to the person on spec, which is no more a violation of confidentiality than the publication of aggregate income information for block groups. It at most enables probabilities inferences about people's characteristics.

**Irene Browne, Emory University, Sociology (Faculty)**

I find this issue very troubling. My research agenda for the next 10 years would have to be scrapped if the 2000 PUMS makes the proposed changes to "protect confidentiality."

**Confidential Response**

I am an undergraduate student and used data to do a project on the impact of Hispanic travel time on youth employment rates.

**Frank W. Sweet, American Military University, History (Graduate Student)**

Without IPUMS, I would not have had the data I used to write my book "History of U.S. Inter-marriage"

**Jonah Gelbach, University of Maryland at College Park, Economics (Faculty)**

I am really fed up with the way a few politicians have created this hysteria, when these questions (and many much more personal ones) have routinely been asked by the Census, the CPS, the NLS, the SIPP, the PSID, a zillion smaller surveys -- and have you seen what's in the NSFG???

We are in danger of seriously handicapping researchers' ability to answer critical policy questions, and someone needs to stand up to the moronic behavior of these politicians. Maybe the Census ought to spend some money advertising ways in which the data are kept confidential.

**Blaine Billings, Boise State University, Sociology (Undergraduate)**

More than ever before we need more census data. Not less. We face greater challenges today and in the future as our population continues to grow, not to mention that only through the calculation of real number populations can any productive public policy be established.

**Confidential Response**

The PUMS system is an invaluable resource for social research and any limitations on the quality and/or amount of detail in the data would undermine the ability of researchers to identify socio-demographic trends and further limit the types of policy implications that could be gleaned from research.

I indicate that I find PUMS to be indispensable to my research and yet I have only used it once. To clarify, I have just recently been introduced to the system and will be using it extensively in the future.

Also, for question #7 I indicated that socio-economic status should be stressed in relation to occupations. This is extremely important, but given an opportunity, I would also have chosen the type of work as well.

**Martye Scobee, University of Louisville, Demography (Support)**

We have both 5% & 1% PUMS data for all of the states. This data has been used in many, many research projects. It has provided invaluable information about our local area and also at the MSA, state, and national level.

**Dr. Gene George, Butler County Community College, Demography (Support)**

People have complained about the Census since the first one in 1790. No one likes to provide information about themselves, but I've found that most people cooperate willingly in research when they are identified as subjects -- if they know the project's purpose and trust the research staff. The news media are overplaying the invasion of privacy angle of the Census. From what I've seen locally, people are responding to the long form. I think Kansas reported a 65 percent or higher response rate to both forms. We need to keep talking about the value of social research as a tool of managing limited public resources, and the vital need for detailed Census data to make social research as meaningful as possible.

**Mark Killingsworth, Rutgers University and NBER, Economics (Faculty)**

If you haven't already done so, you should (and I will myself!) contact government agencies like EEOC, DOJ and DOL. They should be very concerned about the loss of detail in the Census PUMS files, because they and their consultants use this info. All the time.

**Confidential Response**

I do hope that the Census Bureau will stand fast against what is obviously a right wing attempt to limit access to information that is crucial for the furthering of social justice in the United States.

**Helen Burke, Minneapolis Public Library, Other Nonacademic (Nonacademic)**

As a reference librarian in a federal depository library, I myself do not use PUMS data in doing research. Rather I assist others who may do their own research or who may use PUMS data to make policy decisions. Our library users are best served when we can provide access to data that is comparable across time and also as detailed as possible. These criteria are of critical importance to our users and the mission of MPL. As a federal depository library, MPL dedicates staff time and effort to making information we receive available to library users of today and preserving it for the future, too. Having data that is both comparable over time and as detailed as possible allows MPL to maintain its "tradition of providing excellent collections and preserving cultural heritage as an investment for future generations."

**Eileen McConnell, University of Notre Dame, Sociology (Graduate Student)**

PLEASE PLEASE keep the data as detailed as possible to allow for important studies of the population, and maximize historical comparability--critical for understanding how the locational context has changed and/or promoted the increase of such trends as Latinos in the Midwest, Mexican immigrants in non-traditional receiving areas, etc.

**Bruce Western, Princeton University, Sociology (Faculty)**

Before the committee moves ahead with restricting data detail in the PUMS it would be useful to know how many individual's privacy was actually violated by use of the 1980 or 1990 census data. Has anyone complained of such a violation?

Perhaps the committee could use a small part of its budget to commission the efforts of a social scientist to actually identify an individual using the 1990 PUMS.

**Brian Cushing, West Virginia University, Economics (Faculty)**

My responses above assume that they would not gut too much detail on characteristics such as age, occupation, race, etc. Too much reduction in any of these areas could be catastrophic. For me, income and geography are where I have little wiggle room.

**Confidential Response**

There is no way that summary data will be able to satisfy the data needs that arise from decennial census data. The broader the data element categories, the more it will be like summary data.

If geographic areas are defined as larger than the 1990 ones, people will have a difficult time in using it for planning and research. People felt that even the 1990 PUMS census boundaries were too big, although they understood the need for confidentiality.

**Roger Avery, Brown University, Population Studies & Training Center, Demography (Faculty)**

I hope you are successful. PUMS is very important.

**Robert Gibbs, Economic Research Service, USDA, Economics (Nonacademic)**

PUMS is the ONLY source of detailed demographic and labor-related information widely available to the research community that can be used to construct micromodels with local area data for a variety of topics, including migration, human capital and earnings determinants.

**Marie Bousfield, City of Chicago, Government (Nonacademic)**

Please, do not tamper with the PUMS file lightly. It is too important to local governments. This may not be well known because local government workers are not very outspoken.

**Donald Hirasuna, House Research Department, Economics (Nonacademic)**

The need for comparability is never greater. We are now capable of conveniently storing several years worth of data on our cd-roms. To eliminate or reduce our capability by making larger categories is unthinkable.

**Confidential Response**

Beginning with 2001 I plan to use IPUMS for a historical research on Immigrants from the late Austrian Empire to the USA. I'm especially interested in the regional distribution of the places of origin. Any reduction of PUMS detail will make a comparison over time very difficult.

**Marianne Hill, Mississippi Center for Policy Research, Economics (Nonacademic)**

Black secondary school teachers, private household versus center childcare workers, black female textile sewing machine operators, and teachers' aides are other occupations of concern to me.

**Philip Wuest, Clark County, Washington, Government (Graduate Student)**

The more detail with locational specificity the better...at least for meaningful public policy analysis and research.

**Judy Olson, Michigan State University, Geography (Faculty)**

I am a very new user of PUMS data and mainly in connection with some student research. But I am becoming very interested in it and we have discussed how new technology should allow more detail without compromising confidentiality, though we realize it would not be an easy system to administer. I will be disappointed if the data become less useful instead of more useful just when locational technology suggests we should be moving in the opposite direction.

**Jeremy Atack, Vanderbilt University, Economics (Faculty)**

My work is primarily historical and I think that the Census needs also to consider the permanent long-term impact of these decisions on research when issues of confidentiality have become irrelevant. This decision has long-term consequences and will diminish the value of existing (e.g. the IPUMS databases) as well as future work.