THE (MIS)MEASUREMENT OF SUBFAMILIES IN U.S. CENSUS DATA*

Matthew B. Schroeder

Minnesota Population Center

Abstract: Subfamilies--family units residing in someone else's household--are an important subject of research, but they have proved difficult to measure. This research documents trends in and dynamics of the Census Bureau's identification of subfamilies by comparing them to highly refined and temporally consistent subfamily measures newly available in the Integrated Public Use Microdata Series (IPUMS). I show that the Census Bureau's measurement of subfamilies leads to highly unlikely interpretations of family interrelationships and that these apparent errors have grown worse over time, affecting even the most recent American Community Survey data. Furthermore, errors are particularly high among young adults, nonwhites, and persons without a high school diploma--precisely those populations that subfamily researchers are most interested in. Researchers may wish to consider avoiding the U.S. Census Bureau's subfamily measures in favor of the IPUMS subfamily measures.

THE (MIS)MEASUREMENT OF SUBFAMILIES IN U.S. CENSUS DATA

Subfamilies, as defined by the U.S. Census Bureau, comprise a married couple (with or without children) or a parent and child living in the household of another person. While they do not retain a separate residence, subfamilies are important parts of household structure and often constitute a relevant economic unit. Single mothers living with their own children, for example, are a central concern of stratification and poverty research (e.g., Cohen 2002; Snyder and McLaughlin 2004) and demographic analyses (e.g., London 2000). Simply put, researchers who focus on families rather than households require accurate subfamily identifiers.

Several years ago, researchers identified apparent problems in the U.S. Census Bureau's measurement of subfamilies in the 1990 decennial census data (Ruggles and Brower 2003).[1] This research examines whether these problems in the Census Bureau's measurement of subfamilies continue for the most recent public use microdata. I compare original subfamily measures in 1960-2008 census microdata samples to newly available subfamily measurement variables from the Integrated Public Use Microdata Series (IPUMS) (Ruggles et al. 2009). The IPUMS contains microdata samples of every surviving decennial census since 1850; all variables have been harmonized for maximum comparability over time.

Subfamily Measures in Census Data and in the IPUMS

Before 1990, census enumerators and coders performed most of the subfamily identification. Since 1990, the coding of subfamilies has taken place in a computer program. The historical details of this program have to my knowledge never been documented, but the key variable appears to be one's relationship to the householder, and the program apparently performed poorly in 1990 (Ruggles and Brower 2003:84-85). Currently, though, the Census

---

[1] This followed identification of other problems in the Current Population Survey (London 1998) and the Panel Study of Income Dynamics (Rendall 1997).

Bureau relies on relationship codes, age differences between household members, marital status, and sex to construct subfamily codes. Furthermore, it uses information on names not available in public use data. Where family interrelationships are ambiguous--for example, between two unspecified "other relatives" of the householder--a parent and child must have the same surname.

All this would seem to make Census Bureau subfamily variables quite accurate--at least in the most recent data. However, the details of the program appear to create problems, as I show below. The treatment of age differences between household members, for example, leaves room for some illogical links. And the reliance on surname similarity ignores subfamily members who do not have the same last name, excluding many people who are likely in a subfamily--for example, a woman who did not change her name upon marriage.

The IPUMS subfamily variables differ from the Census Bureau's subfamily measures in three main ways. First, IPUMS subfamily variables contain more information and are built from rules that are more flexible than the Census Bureau's program. Importantly, the IPUMS contains family interrelationship variables that identify individuals' likely mothers, fathers, and spouses, and the IPUMS subfamily measures are built from them. Like the Census Bureau's subfamily variables, these "pointer" variables are based on one's relationship to the householder, but they also carry information about individuals' marital status, age, sex, fertility, and sequence in the household.[2] The extra information contained in the pointer variables makes the subfamily measurement more precise, as I attempt to demonstrate below.

Second, IPUMS subfamily variables are coded consistently over time. The method of constructing the pointer variables on which the subfamily variables are based is the same for all years of data (with minor adjustments for differing availability of variables). Census Bureau

---

[2] See http://usa.ipums.org/usa-action/variableDescription.do?mnemonic=POPRULE, http://usa.ipums.org/usa-action/variableDescription.do?mnemonic=MOMRULE, and http://usa.ipums.org/usa-action/variableDescription.do?mnemonic=SPRULE for a full description of these procedures.

coding, in contrast, changes dramatically over time, and small tweaks occur even between the

years of the American Community Survey (ACS), an ongoing survey that has replaced the

decennial census long form.

Finally, for the American Community Survey, IPUMS subfamily variables are based on a

more refined classification of in-laws. In-laws are crucial components of subfamilies. But the

2000-2007 ACS data provides only a general "in-law" category that erases distinctions among

children-in-law, siblings-in-law, and parents-in-law; and the 2008 ACS data lumped siblings-in-

law with the catch-all "other relative" category. Classifying an in-law as the wrong generation

(e.g., calling a child-in-law a sibling-in-law) can change one's subfamily type or even change

whether one belongs to a subfamily. The IPUMS includes detailed rules for classifying in-laws

that have been empirically validated.[3] As far as I am aware, the Census Bureau has no such

procedures in place.

I proceed in four stages. First, I outline trends in subfamily membership over time,

comparing IPUMS and Census Bureau estimates. I then examine specific households in the

2000 decennial microdata and recent ACS data and offer examples of how the extra information

contained in the IPUMS codes provides more plausible household structures than do the Census

Bureau subfamily codes. Third, I look at overall trends in the misidentification of subfamilies

between 1960 and 2008, providing unparalleled detail on the Census Bureau's difficulties in

classifying subfamilies. Finally, I examine what kinds of people are most likely to be

misclassified by the Census Bureau.

---

[3] Applying the in-law classification rules to the 2000 5% PUMS, which contained full detail on in-laws, less than 4 percent of in-laws were misclassified.

How Has Subfamily Membership Changed Over Time?

Figure 1 gives overall trends in subfamilies as measured by the IPUMS (the grey line with diamond markers for each sample) and the Census Bureau (the black line with circular markers for each sample).[4]

(Figure 1 about here: Trends in Subfamily Membership, 1880-2008)

Panel A shows the percentage of U.S. residents living in married-couple subfamilies (a husband and wife living together, with or without their own children) from 1880 through 2008. The presence of married-couple subfamilies rose steadily from 1880 until 1950, plummeted between 1950 and 1970, and rebounded slightly after 1970. The Census Bureau's data shows lower rates of married-couple subfamily membership in every year, but the difference is small (about 0.15 percentage points); the trends are quite similar to the IPUMS.

Panel B shows the percentage of U.S. residents living in father-child subfamilies (a father living with his own child under age 18). Father-child subfamily membership rose sharply between 1980 and 2000 and remained at roughly 2000 levels thereafter. Important here is that the Census Bureau measures have yielded higher estimates of membership in father-child subfamilies since 1990 than the IPUMS measures have, and this tendency is particularly strong in 2006-2008 ACS data.

Finally, Panel C shows the percentage of U.S. residents living in mother-child subfamilies (a mother living with her own child under age 18). After underestimating mother-child subfamily membership in 1980 and 1990 relative to IPUMS measures, the Census Bureau overestimated it in 2000; but estimated rates are quite similar in IPUMS and Census Bureau ACS data from 2001 to 2008.

---

[4] The Census Bureau began measuring subfamilies only in 1960. I examine only related subfamilies—that is, subfamilies in which members are related to the householder by birth, marriage, or adoption. The Census Bureau does not measure unrelated subfamilies, although these codes are available in the IPUMS.

In sum, overall trends in subfamily membership appear to be similar in the two data sources, but Census Bureau estimates appear to be more erratic: consistently applied procedures for obtaining estimates would produce trends that parallel each other and not cross. The reason becomes clear when examining specific households in 2000, 2007, and 2008: the Census Bureau's procedures for classifying subfamilies have been applied inconsistently across time.

Why Are IPUMS Subfamily Codes Preferable?

There are three kinds of discrepancies between Census Bureau and IPUMS subfamily measures. First is what I call *underidentification*, in which the Census Bureau does not see subfamilies where the IPUMS does.[5] Table 1 documents two households with this problem.

(Table 1 about here: Underidentification of Subfamilies in Census Bureau

Public Use Microdata Samples)

In example 1, the most plausible household structure is that the householder's child (person 2) and child-in-law (person 4) are married to each other and are the parents of the householder's grandchild (person 3). The IPUMS indeed classifies this as a married-couple subfamily, but the Census Bureau does not. Example 2, from the 2007 ACS data, is more clearly erroneous: the parents of the head (persons 2 and 3) are explicitly identified as a married-couple subfamily by the relationship code, yet the Census Bureau does not classify them as such, probably because their last names do not match.[6] (The Census Bureau does not forbid links between parents of the householder; many parents *are* linked as a married-couple subfamily in

---

[5] The term "underidentification" suggests that IPUMS subfamily identifiers are the benchmark against which the quality of Census Bureau identifiers can be measured. I make no claims that IPUMS identifiers are flawless, but I do use them as a convenient standard *for terminological distinctions* because they are based on temporally consistent and publicly documented procedures. And, as I attempt to show in this section, IPUMS subfamily classification procedures yield more plausible family interrelationships than do Census Bureau identifiers.

[6] Conversations with Census Bureau staff revealed that this is not an intended part of the procedures for linking the householder's parents: rather, an error in the subfamily classification program applied the condition that parents' surnames match. This will be fixed in the 2009 ACS, but it is doubtful that revised data for previous years will be released.

other households.)

The surname requirement bedevils links between other relationship categories. In example 3, taken from the 2008 ACS, the 32-year-old grandchild of the head is divorced from her husband and has moved in with her grandmother. The 10-year-old and 8-year-old grandchildren are probably her children and great-grandchildren of the householder, thus constituting a mother-child subfamily. Though the surnames do not match, it is easy to see why: she has probably shed her married name, while the children kept their father's name.

(Table 2 about here: Misclassification of Subfamilies in Census Bureau

Public Use Microdata Samples)

The second kind of discrepancy is what I label *misclassification*: the Census Bureau classified some people as belonging to a different type of subfamily. Table 2 gives examples. In Example 1, it is quite unlikely that the householder's 73-year-old sibling (person 2) and 83-year-old sibling-in-law (person 3) have a 10-year-old child (person 5).[7] The more plausible solution, as identified by the IPUMS, is that two subfamilies exist here. The 10-year-old is probably the child of person 4 (who has been married at least once), thus constituting a mother-child subfamily that is separate from the married-couple subfamily identified by the Census Bureau. Nor had this problem been fixed by the 2007 ACS. In Example 2, it is unlikely that the never-married 15-year-old (person 3) is the father of the infant (person 5). Rather, the householder's sibling is probably the mother of all three "other relatives," as suggested by the fact that she is married (but currently separated from her spouse). Thus what is probably a mother-child subfamily of 4 becomes an unlikely father-child subfamily of 2 in the Census Bureau data. The

---

[7] Another possibility is that the subfamily as it exists in confidential Census Bureau data is plausible, but that the Census Bureau's age perturbation procedures for reducing risks that individuals in PUMS data will be identifiable, carried out after the subfamily classification, made the individuals less likely candidates for linking.

IPUMS' more refined attention to other characteristics--marital status and age chief among them--gives its subfamily measurement more power.

Example 3 is slightly subtler, but the most parsimonious interpretation of this household structure is that the householder's sibling (person 2) and in-law (person 4) are married to each other, as suggested by the small age difference between the sibling and the in-law and their married status. (Indeed, the IPUMS rules classify person 4 as a sibling-in-law, although this information is not available in the Census Bureau data.) The Census Bureau does not make this link (again, it is possible that the surnames do not match); as a result, the married-couple subfamily is instead classified as a father-child subfamily.

The 2008 ACS questionnaire included more refined relationship categories, including for the first time options for "child-in-law" and "parent-in-law". While this enabled better spousal links for householders' children and spouses' parents, it made it more difficult to identify spouses of siblings because siblings-in-law were lumped into the "other relative" category.[8] Example 4 shows one consequence of this from the 2008 ACS: it is likely that the 29-year-old other relative (person 4) is married to the householder's sibling (person 3), thus making him a sibling-in-law. The IPUMS identifiers reflect this by classifying him as the husband in a married-couple subfamily, but the Census Bureau procedures--which do not link him as the sibling's spouse--make him instead the father in a father-child subfamily. Again, this is likely because the last names of persons 3 and 4 do not match. But the presence of a married sibling so close in age suggests that a spousal link is plausible even in the absence of a surname match, and the requirement that the surnames of "other relatives" match without exception for a subfamily link

---

[8] While there is no way to tell where respondent s actually placed siblings-in-law, the questionnaire instructions made it clear that anyone who did not fit one of the other categories was to be deemed an "other relative".

to be created--though rooted in the intentional avoidance of "false positive" subfamily links--is a

heavy-handed approach that is too restrictive when other evidence argues for a link.

(Table 3 about here: Overidentification of Subfamilies in Census Bureau

Public Use Microdata Samples)

The third and final kind of disagreement, highlighted in Table 3, is *overidentification*, in

which the Census Bureau identified subfamilies that probably do not actually exist. In Example

1, it is possible but unlikely that the 19-year-old, never-married son of the householder (person

3) had a daughter (person 4) when he was 13 years old. Such a link might be acceptable if the

19-year-old's marital status suggested a relationship, but this is not the case. In other cases, the

problem seems to be that the Census Bureau does not impose an upper limit on the age

difference that can exist between subfamily members.[9] In Example 2, it is fairly implausible that

the 33-year-old son of the householder (person 3) is married to the 85-year-old (person 6), who is

likely the householder's mother-in-law. And in Example 3, it strains the bonds of credulity that a

93-year-old bachelor (person 3) would have fathered a son (person 4) only 16 years ago. The

heterogeneity of the "other relative" category permits any number of alternate explanations. The

same is true for Example 4: if the Census Bureau's identification is to be believed, the never-

married 18-year-old child of the householder bore the 6-year-old grandchild when she was only

12 years old. While all these arrangements are not impossible, there are no compelling reasons

to accept the Census Bureau's classifications *given only the information in the data*.

Again, I do not argue that the IPUMS codes constitute the gold standard of subfamily

measurement against which Census Bureau data can be tested. As with any measure constructed

---

[9] As with the failure to link the householders' parents, this appears to be an error in implementation of the Census
Bureau's rules, not a shortcoming of the rules themselves.

after the fact from inexact data, they are not perfect.[10] However, a close examination of

thousands of households in multiple years of data revealed that the examples given here are by

no means exceptional. The number of cases in which the Census Bureau subfamily codes might

yield more plausible family interrelationships than the IPUMS is vanishingly small; and given

the ever-changing Census Bureau subfamily linking procedures, they do not inspire confidence.

The refined and flexible rules of the IPUMS subfamily variables, along with their temporally

consistent nature, make them a clear choice for researchers analyzing change over time.

What Are The Temporal Dynamics of Subfamily Misclassification?

The above tables suggest that the Census Bureau's measurement of subfamilies did not

improve between 2000 and 2008. Examining the full complement of data confirms this. For

each type of subfamily (married-couple, father-child, mother-child), Figure 2 shows the three

kinds of discrepancies identified in Tables 1, 2, and 3.

(Figure 2 about here: Errors in Census Bureau Subfamily Identification, 1960-2008)

Panel A shows trends in the underidentification of subfamilies--that is, the percentage of

people belonging to IPUMS-defined subfamilies whom the Census Bureau does not classify as

belonging to any subfamily. Underidentification is quite high before the advent of automated

subfamily classification in 1990, particularly for father-child subfamilies (the medium-grey line

with diamond markers). Despite the improvements in 1990, ACS classifications (after 2000)

appear to have gotten worse for father-child and mother-child subfamilies, with

underidentification rates of around 20 percent.

---

[10] For example, the IPUMS allows children to be linked to mothers who are as little as 11 years older than they are; this makes less sense for 2008 data than it does for 1880 data. And the reliance of the family interrelationship pointer variables on record order has its disadvantages in the more modern surveys, when the order in which people are listed is chosen by the respondent and not trained enumerators. This lack of standardization and the ability to sort electronic records eliminates an important clue to household structures.

Panel B shows trends in the misclassification of Census Bureau subfamilies--that is, the percentage of people belonging to IPUMS-defined subfamilies whom the Census Bureau classifies as belonging to a different type of subfamily. These rates are fairly low, and the Census Bureau's manual coding worked quite well before 1990: although enumerators and coders often missed subfamilies that were present or perceived subfamilies that were not, they do not often disagree about the type of subfamilies that both consider to be present. In the ACS, however, father-child subfamilies remain a problem, with misclassification rates hovering around 15 percent.

Panel C shows trends in the overidentification of subfamilies--that is, the percentage of people belonging to Census Bureau-defined subfamilies who probably do not belong to any subfamily at all. Overidentification rates have become dramatically worse over time, especially among father-child subfamilies, and for father-child and mother-child subfamilies they are worse in the ACS than in any previous year.

In sum, the issues that create differences between the Census Bureau's and the IPUMS's subfamily measurements appear to have grown more prominent over time. In certain respects, the system of manual classification in place before 1990 actually yielded results closer to the IPUMS's techniques than did automated classification, which has become particularly questionable in its identification of father-child and, to a lesser extent, mother-child subfamilies.

What Are the Consequences of Subfamily Misclassification?

One might object that the consequences of poor subfamily measurement are relatively minor, given the relatively small net discrepancies in estimates of subfamily membership outlined in Figure 1. However, the size of the gross error rates in Figure 2 suggests otherwise: a large proportion of people in subfamilies are not picked up or are classified as the wrong type by

the Census Bureau; another large proportion of people do not belong to subfamilies at all. If these errors in subfamily identification are not random with respect to important sociodemographic characteristics, then researchers using the Census Bureau's subfamily variables run the risk of erroneous conclusions.

I focus now on the probability of subfamily errors with respect to four key characteristics: age, race, educational attainment, and immigrant status. (Sex and marital status are part of the very definition of subfamilies and are highly collinear within subfamily types, so I have not included them here.) Estimates of underidentification and misclassification are based on all individuals in an IPUMS subfamily and come from multinomial logit models measuring the log-odds of underidentification (not being in a Census Bureau subfamily) or misclassification (being in a different kind of Census Bureau subfamily) as opposed to agreement between IPUMS and Census Bureau identifiers. Estimates of overidentification are based on all individuals in a Census Bureau subfamily and come from similar multinomial logit models. Standard errors have been corrected for clustering within households, and full details are available upon request. To conserve space, I describe results for only mother-child subfamilies in the 2001-2008 American Community Survey data, the most numerous subfamily type and the one that researchers are most concerned about. Gradients in subfamily errors are similar for other subfamily types, although the error levels differ.

(Figure 3 about here: Variation in Census Bureau's Subfamily Errors, 2001-2008 American Community Survey Data)

Figure 3 gives results. Panel A shows the likelihood of *underidentification*, where people in IPUMS married-couple subfamilies were not placed in any subfamily by the Census Bureau. Across all individuals in IPUMS married-couple subfamilies, the average probability of

underidentification was about 20 percent. Probabilities of underidentification are displayed as deviations from this overall average. What is immediately apparent is the greater likelihood of underidentification among young adults (age 18-34), nonwhites, persons without a high school diploma, and the foreign-born.

Panel B gives equivalent results for the probability of *misclassification*. These rates were quite low; only about 1 percent of persons in IPUMS married-couple subfamilies were classified as being in a father-child or mother-child subfamily. Yet misclassification rates leap to 5 percent for Asians and are above 2 percent for children (under age 18) and Hispanics. The high incidence of misclassification among children suggests that the Census Bureau's requirement that most spouses have the same last name leads to coding many children as being in a father-child or mother-child subfamily--rather than the married-couple subfamily they should belong to (as in Example 3 in Table 2). Except for educational gradients and black-white differences, all effects shown here are statistically significant.

Finally, Panel C shows results for *overidentification*--the likelihood that respondents identified as being in a married-couple subfamily by the Census Bureau are not in an IPUMS subfamily. This too is highest among young adults and nonwhites. All differences shown here are statistically significant except for the differences between middle-aged and older Americans; among blacks, Hispanics, and members of other races; and between foreign-born and native-born Americans.

In these three sets of results, Wald tests revealed that each group of coefficients--age, race, education (except in the analysis of misclassification), and immigrant status (except in the analysis of overidentification)--is jointly statistically significant, although differences between racial groups and age groups are not always significant. However, the main contrasts I wish to

draw--children and young adults versus other age groups, nonwhite groups versus whites, and those without a high school diploma versus those with more formal education--were consistently significant.

Furthermore, descriptively speaking, these multivariate probabilities actually *understate* the disparities in subfamily errors. Because nonwhites are disproportionately represented among the young and those without much formal education, the unconditional gradients between these variables and subfamily error are even stronger. Thus researchers who care about simply estimating the preponderance of subfamily membership within specific groups, rather than controlling for other characteristics, are likely to be led even farther astray than these figures suggest.

Conclusion

This research has advanced four main points that users of census microdata should heed:

- The Census Bureau has not classified subfamilies consistently over time.

- The Census Bureau's subfamily identifiers appear not to fully consider important characteristics of household members such as age and marital status.

- The Census Bureau's subfamily identifiers appear to be diverging farther from the temporally consistent IPUMS measures over time.

- The Census Bureau's errors are particularly high among children, young adults, persons of color, and persons without high school diplomas--*precisely the populations that subfamily researchers are most interested in*.

Conclusion

It is important to note that both the IPUMS and the Census Bureau identifiers are based on a limited set of information. Each set aims to make the most informed guesses possible given the rest of the data, but they remain only guesses. The ideal, of course, would be the explicit identification of family interrelationships on the survey form. The Current Population Survey implemented such an approach starting in 2007, and future work will test IPUMS subfamily identifiers against the CPS pointer variables. For the most recent data, CPS subfamily identifiers may well prove more precise than either IPUMS or Census Bureau measures. Yet for pre-2007 samples, and particularly for pre-1990 data, the IPUMS subfamily identifiers are essentially the only available reliable option.

And the tentative conclusion of Ruggles and Brower (2003) holds true for post-1990 census data: subfamily codes in the original Census Bureau data should not be trusted without close scrutiny. IPUMS subfamily identifiers, which are informed by the characteristics of other householders and which are built according to a publicly documented set of rules that is applied consistently across time, provide a superior alternative for analyzing historical, and possibly contemporary, household structures.

REFERENCES

Cohen, Philip N. 2002. "Extended Households At Work: Living Arrangements and Inequality in Single Mothers' Employment." *Sociological Forum* 17:445-63.

London, Rebecca A. 1998. "Trends in Single Mothers' Living Arrangements from 1970 to 1995: Correcting the Current Population Survey." *Demography* 35:125-131.

------. 2000. "The Dynamics of Single Mothers' Living Arrangements." *Population Research and Policy Review* 19:73-96.

Rendall, Michael S. 1997. "Identifying and Misidentifying Single Mothers in the Panel Study of Income Dynamics." *Journal of Human Resources* 32:596-610.

Ruggles, Steven and Susan Brower. 2003. "Measurement of Household and Family Composition in the United States, 1850-2000." *Population and Development Review* 29:73-101.

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. 2009. *Integrated Public Use Microdata Series: Version 4.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor].

Snyder, Anastasia and Diane K. McLaughlin. 2004. "Female-Headed Families and Poverty in Rural America." *Rural Sociology* 69:127-49.

Table 1. Underidentification of Subfamilies in Census Bureau Public Use Microdata Samples

Example and Person Number	Relationship to Householder	Age	Sex	Marital Status	Census Subfamily	IPUMS Subfamily
Example 1 (2000 5% PUMS)						
1	Householder	66	F	Widowed		
2	Child	39	F	Married		Wife
3	Grandchild	20	F	Never married		Child
4	Child-in-law	41	M	Married		Husband
5	Other nonrelative	70	F	Divorced		
Example 2 (2007 ACS PUMS)						
1	Householder	23	M	Never married		
2	Parent	61	M	Married		Husband
3	Parent	56	F	Married		Wife
Example 3 (2008 ACS PUMS)						
1	Householder	62	F	Widowed		
2	Grandchild	32	F	Divorced		Mother
3	Grandchild	10	F	Never married		Child
4	Grandchild	8	M	Never married		Child

Table 2. Misclassification of Subfamilies in Census Bureau Public Use Microdata Samples

Example and Person Number	Relationship to Householder	Age	Sex	Marital Status	Census Subfamily	First IPUMS Subfamily	Second IPUMS Subfamily
Example 1 (2000 5% PUMS)							
1	Householder	72	M	Never married			
2	Sibling	73	F	Married	Wife	Wife	
3	Sibling-in-law	83	M	Married	Husband	Husband	
4	Other relative	47	F	Separated			Mother
5	Other relative	10	M	Never married	Child		Child
6	Other relative	33	M	Never married			
Example 2 (2007 ACS PUMS)							
1	Householder	41	F	Separated			
2	Sibling	37	F	Separated		Mother	
3	Other relative	15	M	Never married	Father	Child	
4	Other relative	7	F	Never married		Child	
5	Other relative	0	M	Never married	Child	Child	
Example 3 (2007 ACS PUMS)							
1	Householder	34	F	Never married			
2	Sibling	35	F	Married		Wife	
3	Parent	60	F	Divorced			
4	In-law*	37	M	Married	Father	Husband	
5	Other relative	4	M	Never married	Child	Child	
Example 4 (2008 ACS PUMS)							
1	Householder	38	M	Married			
2	Spouse	31	F	Married			
3	Sibling	30	F	Married		Wife	
4	Other relative*	29	M	Married	Father	Husband	
5	Other relative	4	M	Never married	Child	Child	

* - Classified by IPUMS as a sibling-in-law

Table 3. Overidentification of Subfamilies in Census Bureau Public Use Microdata Samples

Example and Person Number	Relationship to Householder	Age	Sex	Marital Status	Census Subfamily	IPUMS Subfamily
Example 1 (2000 5% PUMS)						
1	Householder	53	M	Married		
2	Spouse	45	F	Married		
3	Child	19	M	Never married	Father	
4	Grandchild	6	F	Never married	Child	
Example 2 (2007 ACS PUMS)						
1	Householder	56	M	Married		
2	Spouse	53	F	Married		
3	Child	33	M	Married**	Husband	
4	Child	25	F	Never married		
5	Child	23	M	Never married		
6	In-law*	85	F	Married**	Wife	
Example 3 (2007 ACS PUMS)						
1	Householder	56	F	Never married		
2	Grandchild	2	M	Never married		
3	Other relative	93	M	Never married	Father	
4	Other relative	16	M	Never married	Child	
Example 4 (2008 ACS PUMS)						
1	Householder	49	F	Divorced		
2	Child	18	F	Never married	Mother	
3	Grandchild	13	F	Never married		
4	Grandchild	6	M	Never married	Child	

* - Classified by IPUMS as a parent-in-law

** - Classified by IPUMS as married, spouse absent
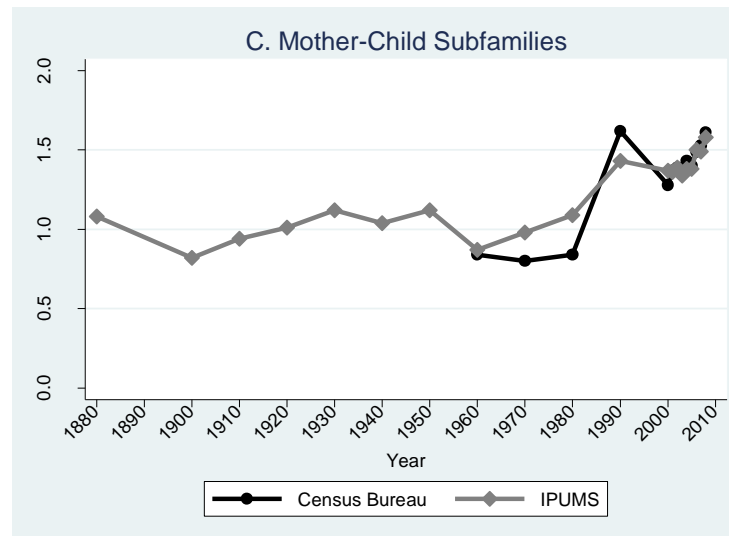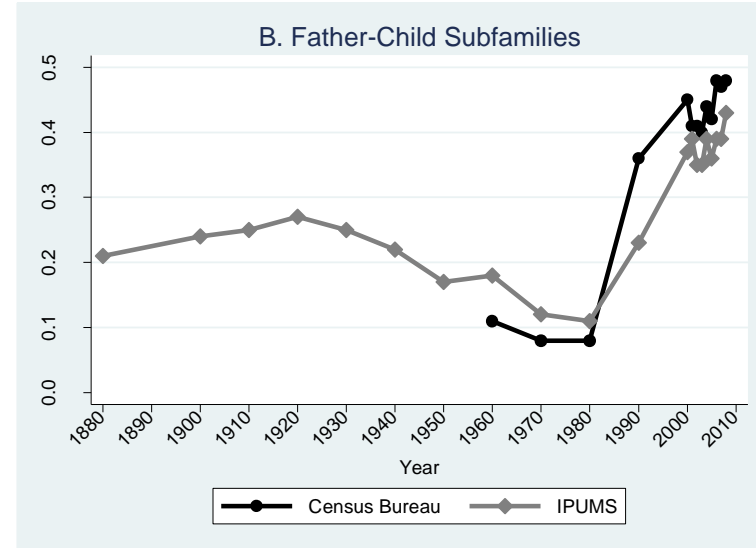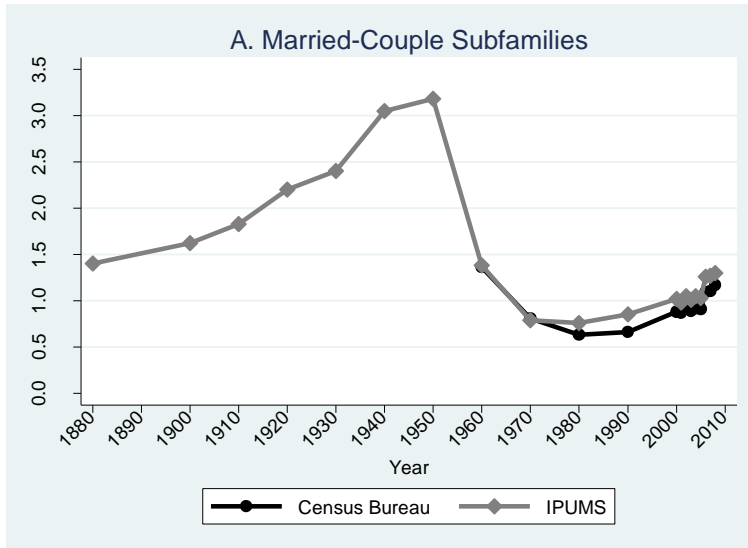
Figure 1. Trends in Subfamily Membership, 1880-2008

A. Married-Couple Subfamilies

B. Father-Child Subfamilies

C. Mother-Child Subfamilies

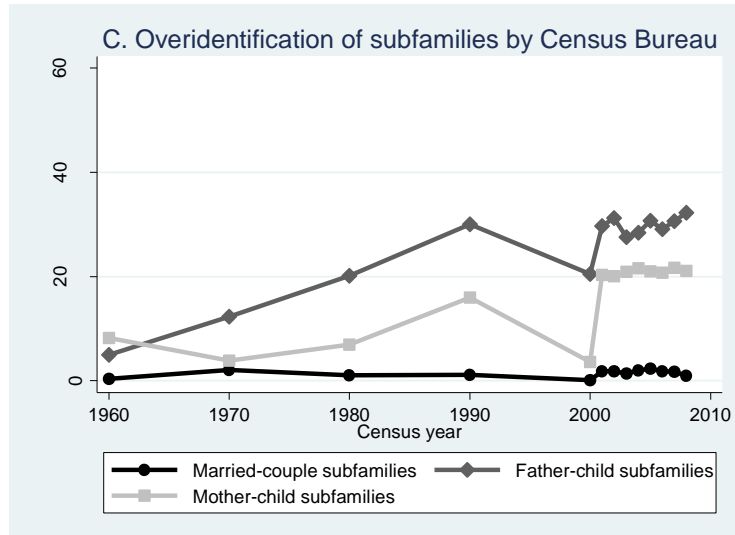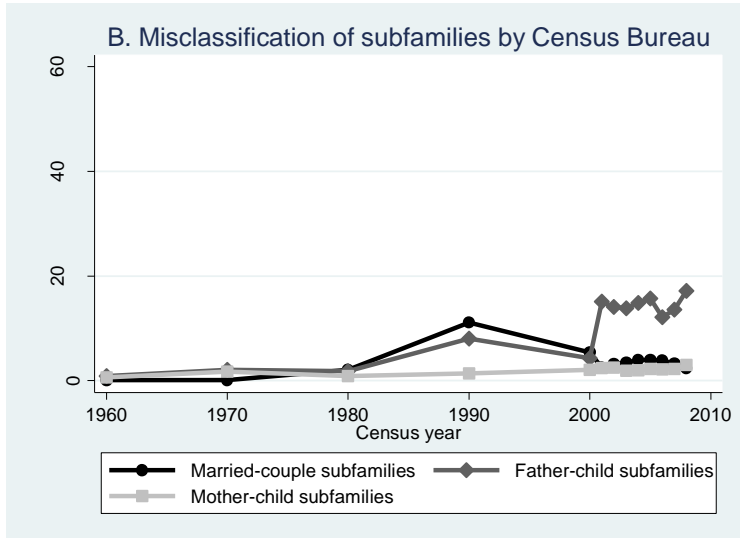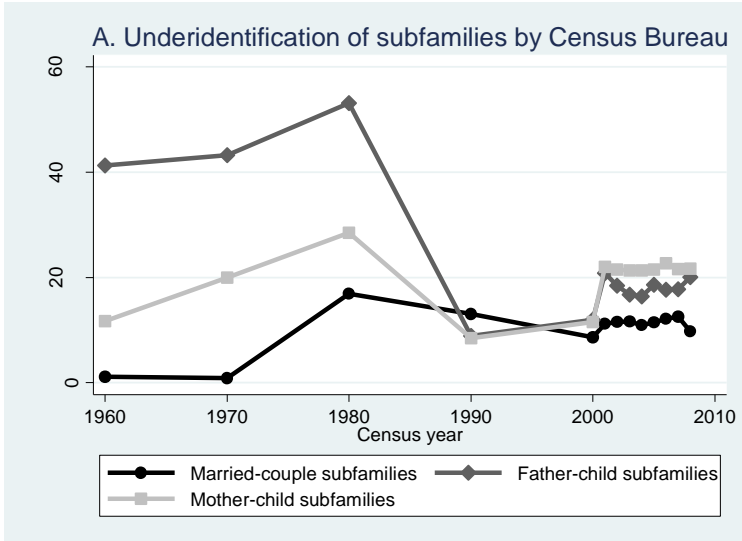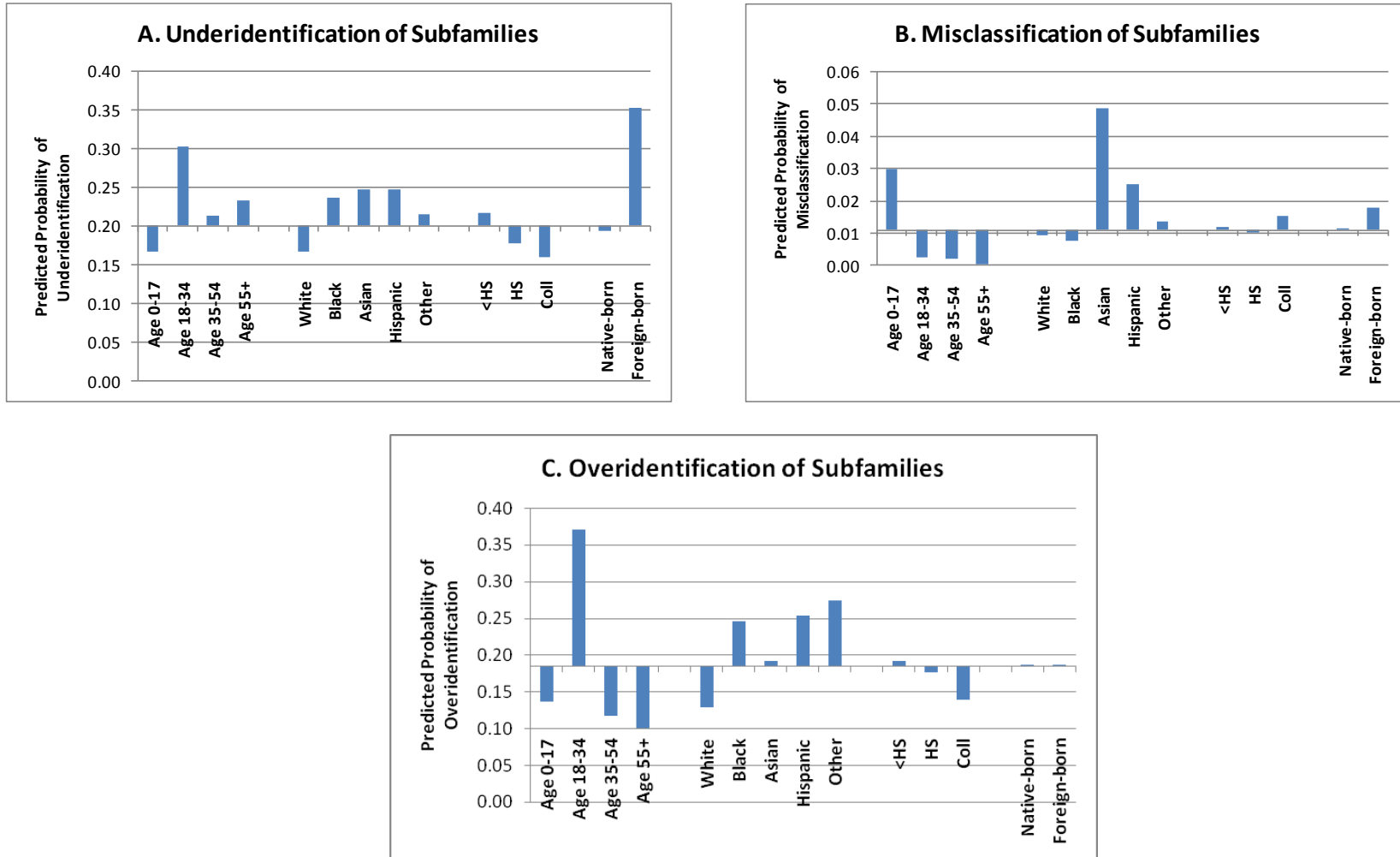**Figure 2. Errors in Census Bureau Subfamily Identification, 1960-2008**

Figure 3. Variation in Census Bureau's Subfamily Errors, 2001-2008 American Community Survey Data

Note: Results shown only for mother-child subfamilies. Probabilities are calculated from multinomial logit models (see text for details) and are displayed as deviations from the unconditional average.