

The (Mis)Measurement of Subfamilies in U.S. Census Data

Matt Schroeder

Minnesota Population Center, University of Minnesota

RESEARCH QUESTION

How accurate are subfamily identifiers in U.S. Census Bureau public-use microdata samples and in the IPUMS?

THE NEED FOR SUBFAMILY IDENTIFIERS

Subfamilies are family units (a married couple with or without children or a single parent and child) residing in the household of another person. Accordingly, they are important parts of household structures; single mothers living with their children, for example, are a central concern of family researchers. To distinguish them from other household members, accurate identifiers are critical.

THE DIFFICULTY IN MEASURING SUBFAMILIES

Despite subfamilies' importance, no items in U.S. census data inquire about them, or about family interrelationships more generally. Instead, researchers must construct post hoc subfamily identifiers from other information—an often challenging process.

CENSUS BUREAU SUBFAMILY IDENTIFIERS

The Census Bureau's procedures for identifying subfamilies have changed dramatically over time (as I show below). In recent years, however, it uses five main pieces of information:

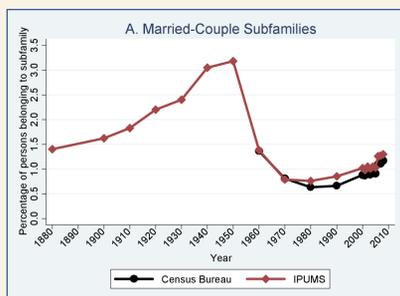
1. Relationship to householder
2. Age
3. Marital Status
4. Sex
5. Surname similarity (where relationship is ambiguous)

IPUMS SUBFAMILY IDENTIFIERS

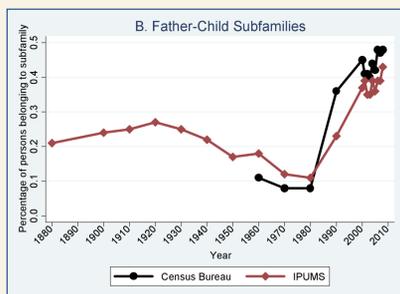
The Integrated Public Use Microdata Series (IPUMS) constructs its own subfamily identifiers. They differ in four ways from Census Bureau's subfamilies:

1. They use more information (despite the unavailability of surname similarity in public-use data), such as fertility.
2. They are coded consistently over time.
3. The implementation of the rules is more flexible.
4. In 2000-2007 American Community Survey data, in-laws are lumped into a global category; the IPUMS uses additional information to break them apart into the various kinds of in-laws—thus improving the accuracy of family interrelationships.

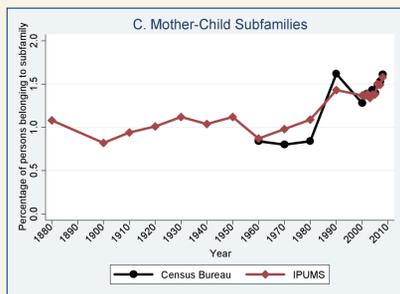
TRENDS IN SUBFAMILY MEMBERSHIP



Married-couple subfamilies comprise a husband and wife, whether with or without their own children. Census Bureau estimates are lower than the IPUMS' after 1970.



Father-child subfamilies comprise an unmarried man living with one or more of his own children. Note the rapid rise after 1980—a trend that is far more jagged as measured by the Census Bureau.



Mother-child subfamilies comprise an unmarried woman living with one or more of her own children. The Census Bureau's subfamily identifiers again yield more extreme results, which likely owe more to measurement changes than to real differences.

SOURCES OF CENSUS-IPUMS DIVERGENCE

What accounts for these differences? I focus on three dynamics in the identification of mother-child subfamilies, with examples drawn from the 2000 5% PUMS, the 2007 ACS, and the 2008 ACS. In each example, the IPUMS provides more plausible subfamily identifiers.

Relationship	Age	Sex	Marital Status	Census Subfamily	IPUMS Subfamily
Householder	62	F	Widowed		
Grandchild	32	F	Divorced		Mother
Grandchild	10	F	Never married		Child
Grandchild	8	M	Never married		Child

Here, the IPUMS identifies the householder's divorced, 32-year-old grandchild as the mother of the 10- and 8-year-old grandchildren (who are likely the householder's great-grandchildren). It is likely that she left her ex-husband and took her children. But, because the great-grandchildren's surnames do not match hers (perhaps they bear her ex-husband's last name), the Census Bureau does not call this a mother-child subfamily. This mother-child subfamily is **underidentified**.

Here, the IPUMS identifies the householder's separated, 37-year-old sibling as the mother of the three "other relatives". Again, though, her children likely share her estranged husband's surname that differs from her own, so the Census Bureau does not link the group as a subfamily—putting in its place a father-child subfamily, the unlikely pairing of a never-married 15-year-old boy and a baby. This mother-child subfamily is **misclassified**.

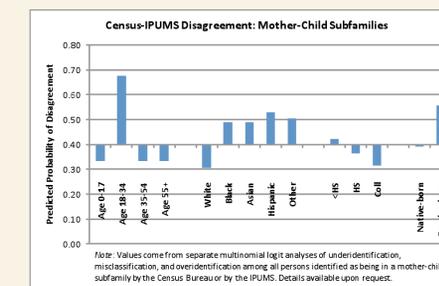
Relationship	Age	Sex	Marital Status	Census Subfamily	IPUMS Subfamily
Householder	41	F	Separated		
Sibling	37	F	Separated		Mother
Other relative	15	M	Never married	Father	Child
Other relative	7	F	Never married		Child
Other relative	0	M	Never married	Child	Child

Relationship	Age	Sex	Marital Status	Census Subfamily	IPUMS Subfamily
Householder	49	F	Divorced		
Child	18	F	Never married	Mother	
Grandchild	13	F	Never married		
Grandchild	6	M	Never married	Child	

Here, the IPUMS sees no subfamily: the age differences are too small. Yet the Census Bureau allows the householder's never-married, 18-year-old daughter to be the mother of the 6-year-old grandchild, even though she would have given birth at the unlikely age of 12. It is more likely that the householder has another child living elsewhere who is the parent of both grandchildren. This mother-child subfamily is **overidentified**: the Census Bureau classifies a subfamily that probably does not exist.

A scan of hundreds of households where Census Bureau and IPUMS subfamily identifiers disagree reveals that (1) the Census Bureau identifiers almost never yield household structures with more face validity than IPUMS identifiers; (2) the above differences are by no means exceptional and extend to married-couple and father-child subfamilies as well; and (3) the disagreement between the two sets of identifiers has in fact *increased* over time.

Furthermore, and most disconcertingly, the disagreements above are most prevalent among young adults, nonwhites, those without much formal education, and the foreign-born—precisely those individuals that subfamily researchers are most interested in.



CONCLUSIONS

Researchers studying subfamilies should strongly consider using the IPUMS subfamily identifiers instead of the Census Bureau's variables. Neither is ideal—explicit survey items about family interrelationships, such as those provided in the 2007-onward Current Population Surveys, are preferable—but IPUMS identifiers include more information, are temporally consistent, and provide more intuitively plausible household relationships. They are the clear choice for analyzing trends in subfamily dynamics.

ACKNOWLEDGEMENTS

I thank Martin T. O'Connell of the U.S. Census Bureau for his assistance in understanding the Census Bureau's measurement of subfamilies, and Charlot Meyer of the Minnesota Population Center for her design wizardry.