Journal of Economic and Social Measurement 42 (2017) 123–149
DOI 10.3233/JEM-170445
IOS Press

Family matters: Development of new family interrelationship variables for US IPUMS data projects

Marina Mileo Gorsuch[a,*] and Kari Charlotte Wigness Williams[b]
[a]*Economics and Political Science, St. Catherine University, St. Paul, MN, USA*
[b]*Minnesota Population Center, University of Minnesota, Minnesota, MN, USA*

In demographic datasets, researchers frequently want to identify how members of a household are related. In this paper, we develop a new method of estimating parental and spousal relationships using data on fertility patterns and family interrelationships. The improved method includes cohabiting and same-sex couples and is comparable across all modern US IPUMS data projects. A detailed variable indicates how the relationship was inferred and the level of ambiguity around that inference. The new IPUMS family interrelationship variables are very accurate, matching self-reported spouse/partner for 99.99% and parent for over 99.00% of respondents. Among those identified as same-sex couples, we match self-reported spouse/partner for 100% of respondents, 87.57% of whom self-identify as lesbian, gay, or bisexual. We further demonstrate that the new family interrelationship variables closely track temporal variation in teenage fertility.

Keywords: Family interrelationships, demographic datasets, family composition

1. Introduction

Many demographic datasets enumerate people inside households. These datasets frequently include a variable showing how an individual household member is related to a reference person (e.g., how each person is related to the householder). This relationship variable is important in many analyses – for example, it is useful to know how many children a person has when examining her/his labor market activity. However, because datasets typically provide only one variable reporting how each member of a household is related to a single reference person, relationships among household members are frequently ambiguous. For example, in a household with a "householder", two female "children", and one "grandchild", it is not immediately clear who the parent of the "grandchild" is.

Using a variety of other data, including age, sex, and marital status, researchers can often infer relationships between household members. However, it cumbersome and inefficient for each researcher to create these family interrelationship variables for her/his own analyses, and the variables will inevitably differ between researchers.

*Corresponding author: Marina Mileo Gorsuch, Economics and Political Science, St. Catherine University, St. Paul, MN 55105, USA. Tel.: +1 857 247 4952; E-mail: mmgorsuch952@stkate.edu.

Different results between studies may simply be the product of different choices in the construction of these family interrelationship variables. These challenges are compounded by varying levels of detail provided in the relationship variables between datasets or across time within the same dataset. Researchers seeking to analyze multiple datasets or compare results across datasets add another layer of complexity because the construction of family interrelationship variables must be consistent across datasets as well as researchers.

Beginning in 1995, IPUMS released the original family interrelationship variables indicating the line numbers (or "location") of each person's co-resident parent(s) and spouse [1]. These groundbreaking variables allowed researchers to examine household and family structure in a completely new way. Characteristics of parents could be attached to children to examine how parental education or income was associated with child wellbeing. The number and age of children could be easily calculated, to see how family composition was associated with the labor supply of parents. In 2009, IPUMS International developed a set of family interrelationship variables that are flexible enough to accommodate the diversity of families in an international setting, including polygamous families and large variation in household size [2].

The IPUMS family interrelationship variables in US data projects have become outdated because of changing family structure and changes in how families are enumerated in datasets. For example, there has been a large increase in cohabiting couples over the past decades [3], and data collection agencies now identify cohabiting couples. Additionally, data collection agencies are increasingly reporting married and cohabiting same-sex couples rather than editing the sex or marital status of reported same-sex couples, as was previously common practice [4–6]. Because the original IPUMS family interrelationship variables are based on data that did not report same-sex or cohabiting couples, they do not include these types of families. The original variables were also developed for use in datasets where families were generally grouped together on a household roster. This meant that proximity and order could be used to infer relationships. However, many datasets no longer group household members in this way.

In this paper, we document a new set of family interrelationship variables for all modern US IPUMS datasets (IPUMS CPS, IPUMS NHIS, ATUS-X, and IPUMS USA for all ACS samples as well as Census 1970 and later) [7–10]. The new IPUMS family interrelationship variables include cohabiting and same-sex couples, and enable more comparable analyses across datasets and across researchers. We define a two-stage protocol for addressing ambiguity when assigning family interrelationship variables. First, we describe how we prioritize links based on the clarity of the relationship between the two people being linked. This step is specific to each dataset, because they provide different levels of detail about a respondent's relationship to the householder. Second, we create a set of logical assumptions for how to identify spouse/partner and parent locations when there are multiple potential links. In a detailed appendix, we present the evidence underlying our rationale for these logical steps. Each location variable has a corresponding rule variable, which denotes the

clarity of the relationship between the two people being linked and the logic used to assign the link, allowing researchers to include only those links made on assumptions acceptable for their analyses.

We test the new family interrelationship variables with two methods. First, we compare the new IPUMS family interrelationship variables to the self-reported family interrelationship variables available in recent years of the Current Population Survey (CPS) and National Health Interview Survey (NHIS). We show that the new IPUMS parent and spouse/partner location variables match self-reported locations extremely well. In particular, we examine same-sex couples and find that these couples perfectly match self-reported spouse/partner location and largely self-report being gay, lesbian, or bisexual. We also examine changes in teenage fertility, demonstrating that racial and temporal variation in teenage fertility is captured in our family interrelationship variables. Because teenage parents are more likely to live in a household they do not head, this is a challenging test for the new family interrelationship variables.

Finally, researchers who have used the original family interrelationship variables may be curious how the new variables compare to the original. We show that the new family interrelationship variables are identical to the original variables for the vast majority of respondents, with differences almost entirely due to additional links based on cohabiting and same-sex couples.

2. New IPUMS interrelationship variables

IPUMS uses five variables to link a person to her/his probable parent(s) and spouse/partner. The variables MOMLOC, POPLOC, MOMLOC2, and POPLOC2 identify the location of a person's probable parent(s). MOMLOC2 or POPLOC2 is used when a child has two parents of the same sex.[1] The variable SPLOC identifies the location of a person's probable spouse or partner; SPLOC is used for both same-sex and different-sex couples. By design, these variables will always be consistent with each other: spousal locations are always reciprocal; when a child has two parents, they are also identified as each other's spouse or partner.

The new IPUMS family interrelationship variables link family members using a two-stage protocol. In the first stage, we prioritize links based on the clarity of the relationship between the two people being linked. For example, in NHIS a "child" and "child-in-law" are a clear spousal link, but occasionally a "child-in-law" will be listed as an "other relative". The "child" to "child-in-law" link is less ambiguous and therefore takes priority over a "child" to "other relative" link. The second stage of the

[1]For example, MOMLOC and MOMLOC2 would be used to show the location of a child's parents when both parents are women. MOMLOC and POPLOC would be used when the child's parents are different sex.

protocol systematically selects the best candidate among multiple potential spouses or parents. For example, we describe how we form links in a household with multiple married "children" and multiple married "children-in-law".

A large, complex household is likely to have both less clear relationships between the two people being linked (e.g., more "other relatives" may exist in a complex household) and more candidates for linking among potential spouse/partners and potential parents (e.g., more potential parents for a "grandchild"). While these two sources of ambiguity are interdependent, our protocol parses the overall question by first ranking the clarity of relationship and then using logical steps to select the best candidate among multiple potential links.

Each family interrelationship variable has an accompanying two-digit rule variable that indicates how the link was formed. The first digit indicates how direct the relationship is between the two people. The second digit indicates how the protocol selected between multiple potential candidates for a person's spouse/partner or parent. These variables are SPRULE, MOMRULE, MOM2RULE, POPRULE, and POP2RULE. The detailed rule variables are a valuable tool for researchers to understand how the spouse/partner and parent location variables were inferred and the level of ambiguity around that inference.

In the following sections, we will refer to "linked parent" and "linked spouse" as the person assigned as an individual's parent or spouse/partner. A "potential parent" or "potential spouse" is a person who meets the requirements to be the linked parent or spouse/partner of an individual. A "linked child" is the child half of the parental link.

2.1. Prioritizing links based on relationship

2.1.1. Prioritizing spousal links

We first prioritize links within a household based on the clarity of the relationship between the two people being linked. Assignments are first made within the highest priority level, followed by each subsequent level. For example, if a "child" reports being married and the household contains both a married "child-in-law" and a married "other relative", the "child" to "child-in-law" will be prioritized because this is a clearer and more direct link. Spouse/partner links are formed only for people whose marital status indicates a spouse/partner is present; because not all datasets include "Living with partner" as a marital status, pairings between a "householder" and an "unmarried partner" are made regardless of marital status.

Because each data project has a different level of detail in the "relationship to householder" variable, specific acceptable pairings are created for each data project. Table 1 shows two examples of the prioritization of spousal links for datasets with different available relationship categories: IPUMS CPS and IPUMS NHIS. The prioritizations of pairings for ATUS-X (part of IPUMS Time Use) and IPUMS USA are found in Appendix 2. Regardless of different relationship categories across datasets,

Table 1
Priority of spousal links based on how clear the relationship is between the two people being linked

	IPUMS CPS		IPUMS NHIS	
1st level links	Householder	Spouse	Householder	Spouse
SPRULE value:	Parent	Parent	Child-in-law	Child
1-	Housemate	Housemate	Sibling-in-law	Sibling
	Roomer	Roomer	Aunt/uncle	Aunt/uncle
	Non-relative	Non-relative	Grandparent	Grandparent
			Parent	Parent
			Parent-in-law	Parent-in-law
			Housemate	Housemate
			Roomer	Roomer
			Non-relative	Non-relative
2nd level links	Householder	Partner	Householder	Partner
SPRULE value:				
2-				
3rd level links	Other relative	Other relative, grandchild, child, stepchild, sibling	Other relative	Child, sibling, sibling-in-law, other relative, child of partner, niece/nephew, grandchild, child of ineligible householder
SPRULE value:				
3-				
	Non-relative	Roomer, housemate	Non-relative	Roomer, ward, housemate
			Unknown	Unknown
4th level links	Householder	Other relative, non-relative	Other relative	Grandparent, aunt/uncle, parent, householder
SPRULE value:				
4-	Child	Child	Niece/nephew	Niece/nephew
	Grandchild	Grandchild	Child	Child
	Unknown	Roomer, housemate	Grandchild	Grandchild
	Sibling	Sibling	Unknown	Roomer, ward, housemate
			Sibling-in-law	Sibling-in-law
			Sibling	Sibling

the first level of spousal links are unique pairings, where the relationship to the reference person clearly identifies the expected relationship of that person's logical spouse/partner. For example, the "householder" clearly will link to the "spouse". Similarly, the reference person's "parent" will link to the other "parent", while a "sibling" would link to a "sibling-in-law". Spousal links assigned via the first priority level of relationship clarity have a value of "1" in the first digit of SPRULE.

The second level of links is exclusively for the reference person and a "partner" rather than "spouse". This link does not require that the person report being married. Couples linked at the second priority level will have a "2" as the first digit of the SPRULE variable.

The third level of links introduces pairings that are not unique based upon reported relationship: an "other relative" may be married to persons reporting many different relationships to the reference person. Spousal links at the third level are based on less clear relationships; the first digit of the SPRULE variable for these links is 3. The fourth level of links includes those links that seem appropriate given reported marital status, but are somewhat questionable: for example, we expect an " aunt/uncle" to

Table 2
Example household of spouse/partner pairing level 3

Person number	Relationship to householder	Sex	Marital status	Age
1	Householder	Male	Married, spouse present	67
2	Spouse	Female	Married, spouse present	67
3	Child	Female	Divorced	46
4	Grandchild	Male	Married, spouse present	23
5	Other relative	Female	Married, spouse present	22

IPUMS NHIS 2015 serial 25424.

Table 3
Example household of spouse/partner pairing level 4

Person number	Relationship to householder	Sex	Marital status	Age
1	Householder	Male	Married, spouse present	73
2	Spouse	Female	Married, spouse present	71
3	Grandchild	Female	Married, spouse present	19
4	Grandchild	Male	Married, spouse present	27

IPUMS NHIS 2015 serial 24743.

be married to an "aunt/uncle", but occasionally the spouse of a "aunt/uncle" is listed as a "other relative". To show that these spousal links are based on questionable relationship combinations assigned in the fourth level of relationship clarity, the first digit of the SPRULE variable is set to 4.

Because the spouse/partner links in level 3 and 4 are based on less direct relationships and may be less intuitive to data users, we provide two example households from IPUMS NHIS that show two ways a "grandchild" can be linked to a spouse. There is no specific relationship value meaning the spouse of a grandchild, although this person would be considered a relative of the householder – consequently the spouse of a "grandchild" would most logically be listed as an "other relative". Table 3 is an example of this type of household, where person 4 is a "grandchild" who is likely married to person 5, an "other relative". A "grandchild" to "other relative" spousal link is assigned in level three, so the value of SPRULE for this link will begin with a 3.

More rarely, the spouse of a "grandchild" will simply be listed as another "grandchild". As shown in Table 4, person 3 is a "grandchild" who reports their marital status as "Married, spouse present". There is no "other relative" in the household, but there is another "grandchild" who also reports being "Married, spouse present". Person 3 and person 4 will be linked via SPLOC under the fourth relationship clarity level. The value of SPRULE for this link will begin with a 4, to show that this is a fourth level link.

2.1.2. Prioritizing parental links

After spousal links are formed, we create parental links. Parental links are more complicated to infer than spousal links. For example, we know a respondent should be linked to a spouse if they report being "Married, spouse present", but there is no similar variable that indicates if a person's parent lives in the same household.

Table 4
Priority of parental links based on how clear the relationship is between the two people being linked

	IPUMS CPS		IPUMS NHIS	
Direct links	Child	Householder	Child	Householder
Parental rule value:	Householder	Parent	Householder	Parent
1-	Sibling	Parent	Sibling	Parent
			Child of partner	Partner
			Spouse	Parent-in-law
2nd level links	Grandchild	Child	Grandchild	Child
Parental rule value:	Child	Spouse, partner	Parent	Grandparent
2-			Child	Spouse, partner
			Niece/nephew	Sibling
3rd level links	Housemate	Housemate	Housemate	Housemate
Parental rule value:	Roomer	Roomer	Roomer	Roomer
3-	Other relative	Other relative, sibling (for years 1988 ASEC to current ASEC; 1994 to current basic monthlies)	Other relative	Other relative, aunt/ uncle, niece/nephew (for years 1969–present)
	Non-relative	Partner, non-relative	Non-relative	Non-relative
	Unknown	Unknown	Unknown	Unknown
			Grandchild	Child-in-law
4th level links	Other relative	Child, other relative (for years ASEC 1962 –1987; 1989–1993 basic monthly samples)	Other relative	Other relative, child (for years 1963–1968)
Parental rule value:				
4-				
5th level links	Other relative	Other relative, householder, spouse (for years basic monthly samples 1976 to 1988)		
Parental rule value:				
5-				

Additionally, a parental link is not one-to-one: a respondent could be living with zero, one, or two parents, and a parent could have multiple children. Because these links are harder to infer, we create a prioritization for parental links that imposes additional restrictions as the clarity decreases. The parental links occur after spousal links and draw on information in SPLOC: if one member of a couple linked through SPLOC is identified as a linked parent, that person's spouse/partner is also identified as a linked parent.

As with spousal links, we first prioritize parental links based on clarity of the relationship between the potential parent and potential child. Assignments are first made within the highest priority level of relationship clarity, followed by each subsequent level. Table 4 shows parental pairings between the linked child and linked parent. First level parental links (such as "householder" and "child") face no restrictions, because the relationship connecting the linked parent and linked child are explicit. The parental rule variables (MOMRULE, MOM2RULE, POPRULE, and POP2RULE) assigned at the first priority level will have a first digit of 1, indicating that these

Table 5
Example household of parental pairing level 3

Person number	Relationship to householder	Sex	Marital status	Age
1	Householder	Female	Divorced	61
2	Niece/nephew	Male	Married, spouse present	41
3	Niece/nephew	Female	Married, spouse present	39
4	Other relative	Male	NIU	8
5	Other relative	Male	NIU	6
6	Other relative	Male	NIU	3
7	Other relative	Male	NIU	3

IPUMS NHIS 2015 serial 810.

parental links are based on explicit relationships.

Second-level parental links are not explicit, but the relationship between the linked parent and the linked child is clear. For example, we expect the parent of a "grandchild" to be reported as a "child" of the householder, but it is not known which "child" is the true parent to the "grandchild", or if the true parent even lives in the household. The relationship is clear, but it is not explicit. To avoid unlikely links, we impose a restriction on the allowed age difference between the potential parent and the linked child for the second level of parental links. Any potential mother must be 15–44 years older than the child; potential fathers must be 15–60 years older than the child (these cutoffs are discussed in detail in Appendix 1). To show that these parental links are based on clear but not explicit relationships from the second level of relationship clarity, the parental rule variables (MOMRULE, MOM2RULE, POPRULE, and POP2RULE) will be assigned a value of 2 in the first digit.

The third, fourth, and fifth levels of links include household members whose relationships to the householder result in more ambiguity about the relationship between the linked parent and linked child (e.g., "non-relatives" or "other relatives"). In addition to the age difference restrictions, links assigned in the third, fourth, and fifth level also require that the linked child is 22 years old[2] or younger and is single/never married. Table 5 shows an example household with a clarity level 3 parental link, where there are four people listed as an "other relative" and a married couple who both have relationship values of "niece/nephew". An "other relative" may be a child of many different relationship values, so this pairing contains more ambiguity than first or second level links. The parental rule value for this link will have a 3 in the first digit, indicating the pairing occurred in the third level of relationship clarity for parental links.

The allowed relationship links between a linked child and a linked parent vary over time because of the level of detail collected regarding the relationship to the reference person. For example, in the 1976 to 1988 CPS basic monthly samples, the only options for relationship to the householder were "householder", "spouse", "other relative", and "non-relative". This means that a child of the householder would be listed

[2]See Appendix 3 for analysis illustrating that including "other relative" household members who are 19 to 22 in the child linking pool improves accuracy.

Table 6
Ambiguous parents of a grandchild in the ACS

Person number	Relationship to householder	Sex	Marital status	Age
1	Householder	Female	Married/in union	44
2	Spouse	Male	Married/in union	47
3	Biological child	Female	Single/never married	21
4	Biological child	Female	Single/never married	19
5	Grandchild	Male	Single/never married	< 1

(Serial 202000).

as an "other relative" in these samples. To appropriately identify family interrelationships in these early samples, both an "other relative" and the "householder" can be identified as the parent of an "other relative", but these pairings are not allowed in more recent samples containing more detail about the relationship to the reference person. We form links based on clear relationships with higher priority before exploring subsequent levels.

2.2. Selecting between multiple potential links

While the first step to finding a spouse/partner or parental link is based on the relationship to the reference person, this may not be enough to uniquely determine the link. Sometimes a person is in a household with multiple potential parents or potential spouse/partners. For example, in the household from the 2010 American Community Survey shown in Table 6, there is a "householder", "spouse", two "children", and one "grandchild". Even though there is a clear relationship between "child" and "grandchild", both "children" are potential parents to the "grandchild".[3]

To select among multiple potential parents, we apply a set of logical steps *within* each relationship priority level. The steps for parental links were developed based on analysis of two datasets with more detail about family interrelationships; our data-driven approach is described in Appendix 1. Importantly, these logical steps are the same across relationship priority levels and across all IPUMS US data projects, because the logical steps do not depend on the specific relationships being paired. For example, the same set of logical steps will apply when selecting among potential parents for a "niece/nephew" in IPUMS NHIS as for an "other relative" in IPUMS CPS, where "niece/nephew" is not an available relationship category. In the section below, we first describe the logical steps used to select among multiple potential spouse/partners and then among multiple potential parents.

2.2.1. Logical steps to assign spouse/partner links

Table 7 shows the order of logical steps in which spousal links are formed *within* each relationship clarity level from Table 1. Couples with only one potential spouse

[3]It is important to note that the order of members of a household in the ACS depends only on age and the relationship to the householder and does not reflect families within a household.

Table 7
Order in which spousal links are formed within each priority level from Table 1 when selecting among potential spouse/partners

Logical step 1 SPRULE value: -1	There is only one potential spouse/partner (including both same-sex and different-sex spouse/partners)
Logical step 2 SPRULE value: -2	There are multiple potential spouses/partners but only one potential different -sex spouse/partner
Logical step 3 SPRULE value: -3	There are multiple possible different-sex potential spouses/partners, relative age is used to link couples
Logical step 4 SPRULE value: -4	There are multiple different-sex potential spouses/partners but there are two people of the same age or age otherwise does not allow a clear link, location in the household is used as a "tie breaker"
Logical step 5 SPRULE value: -5	There are multiple same-sex potential spouses/partners in a household, again use relative age to determine who will pair together
Logical step 6 SPRULE value: -6	If there are multiple same-sex couples in a household and there are two people of the same age or age otherwise does not allow a clear link, location in the household is used as a "tie breaker"

or partner are linked as a first logical step; this is done for both same-sex and different-sex couples. These links are the clearest, because there is only one possible spouse/partner to link to: there is no inference. The SPRULE will have a second digit of 1 to show that the household composition made this particular pairing very clear.

If a person has multiple potential spouse/partners they will not be linked under logical step one. For those who have multiple potential spouse/partners, but only one different-sex potential spouse/partner, the different-sex potential spouse/partner will be assigned. The second digit of SPRULE will be a 2, indicating that this link is also quite clear and assigned using the second logical step. When there are multiple different-sex potential spouse/partners, relative age is used to link couples.[4] In the event that there are two different-sex potential spouse/partners of the same age, proximity to each other in the household roster is used as a "tie breaker".[5] These links will be denoted with a 3 and 4, respectively, on the second digit of SPRULE to indicate more ambiguity around the link. People who were not paired in the first four steps and have multiple same-sex potential spouse/partners are paired in the last logical steps, using relative age and location in the household to determine the best potential spouse/partner. The second digit of SPRULE for these links will be a 5 and 6, respectively.

2.2.2. Logical steps to assign parent links

Table 8 shows the order in which parent links are formed *within* each priority level based on relationship clarity from Table 4. Parent links that are direct (priority level

[4]The potential spouses are ranked by age within sex and then paired. For example, the eldest female "child" will be paired with the eldest male "child-in-law" and the second eldest female "child" with the second eldest male "child-in-law".

[5]Starting at the top of the household, each person to be paired will link to the closest potential spouse who has not yet been linked.

Table 8
Order in which parent links are formed within each priority level from Table 4 when selecting among potential parents

Logic level 1 Parental rule value: -1	This step identifies linked parents through links that are uniquely identified by relationship (priority level 1 from Table 4)
Logic level 2 Parental rule value: -2	When a child is in a household where there is one married couple among his/her potential parents. This couple is assigned as the linked parents
Logic level 3 Parental rule value: -3	When a child is in a household where there are multiple married couples among his/her potential parents. Relative age is used to assign the linked parents
Logical step 4 Parental rule value: -4	When a child is in a household with no married couple among his/her potential parents, but there is only one unmarried potential parent in the household, this person is identified as their linked parent
Logical step 5 Parental rule value: -5	When a child is in a household with no married couple among his/her potential parents and there are multiple potential parents, the first option selected is a previously married woman. Relative age is used to assign the linked parent in households with multiple previously married women
Logical step 6 Parental rule value: -6	When a child is in a household with no married couple and no previously married woman among his/her potential parents, and there are multiple potential parents, this step will identify a previously married man as the linked parent. Relative age is used to assign the linked parent in households with multiple previously married men
Logical step 7 Parental rule value: -7	When a child is in a household with no married couple or previously married people among his/her potential parents, the eldest single woman is selected as the linked parent
Logical step 8 Parental rule value: -8	When a child is in a household with no married couple, previously married people, or single women among his/her potential parents, the eldest single man is selected as the parent

1 in Table 4) are made first, because these relationships are explicit. For example, a person identified as the "child" of the householder will link directly to the householder. Potential children who have no explicit relationship to a parent are linked in logical steps two through eight.

The second stage of the protocol identifies potential parents based on the possible relationship pairings outlined in Table 4. Logical step two links potential children and parents when there is only one married couple who are potential parents. The second digit of the parental rule variables will be a 2, indicating that this link was made under logical step 2. Logical step three links potential children and parents when there are multiple married couples who are potential parents. In this case, children are split among couples by age, with the eldest potential children linking with the eldest potential parents.[6] The logical step is *always* reflected in the second digit of the parental rule variables.

[6]When possible, children are split equally among the married couples who are potential parents with the eldest child(ren) being assigned to the couple with the eldest mother. If children cannot be split evenly (e.g., three children and two couples) the children will be split by age into sibling groups; there will be as many sibling groups as there are married couples. The eldest sibling group will have, at most, one more

Logical steps four through six attempt to link children who have no married couples among their potential parents. If a child has only one potential parent in the household, logical step four will link this person as the child's parent regardless of the linked parent's marital status. Among children with multiple potential parents who are not married, previously married women are prioritized as potential parents in logical step five. If there are no previously married women, logical step six links potential children to previously married men. As with married couples, children are split among multiple potential previously married parents by relative age in logical steps five and six.

The pool of potential parents for children who were not linked in logical steps one through six consists of never married people. Logical step seven links children to the eldest single woman who is a potential parent. Logical step eight links potential children to the eldest single man who is a potential parent.

3. Testing the new family interrelationship variables

We have implemented a series of tests to determine how well the new IPUMS family interrelationship variables perform. In recent samples, the NHIS and CPS both collect self-reported family interrelationship information, where the respondent indicates who in the household is their spouse/partner and parent(s). Our spousal and parental links were assigned independently of this self-reported information, so we are able to use it to test the new IPUMS family interrelationship variables' accuracy. We first examine how accurately IPUMS family interrelationship variables perform among those we identify as same-sex couples. We match self-reported spouse/partner location among same-sex couples identified by IPUMS spouse/partner links and examine the proportion who self-identify as lesbian, gay, or bisexual. We then compare how well IPUMS spouse/partner and parent location variables match self-reported family interrelationship variables among all people. We further examine if the IPUMS family interrelationship variables are consistent with observed changes in teenage fertility over time. All analyses are performed with STATA 13 using a variety of computers; the statistical code was written by the authors [11].

Finally, researchers who have used the original family interrelationship variables may be curious how the new variables compare to the original. We demonstrate that the new IPUMS family interrelationship variables are identical to the original variables for the vast majority of respondents, with differences almost entirely due to additional links based on cohabiting and same-sex couples.

---

child than the youngest sibling group. Sibling groups are assigned to married couples by the age of the mother.

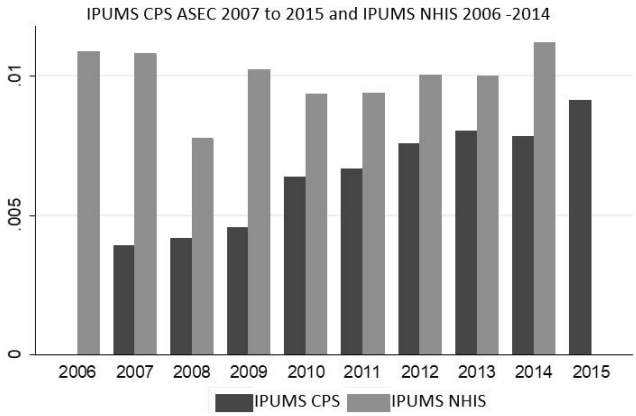


Fig. 1. Proportion of those with a spouse or partner who are in same-sex couples. Figures use appropriate sampling weights.

3.1. Testing the new family interrelationship variables: Same-sex couples

Same-sex couples can be difficult to identify in demographic datasets for numerous reasons. First, same-sex couples have historically not been able to legally marry, limiting consistently available information to identify these couples (i.e., marital status). Additionally, different-sex couples will occasionally report the sex of one person incorrectly and appear to be same-sex couples [6,12,13]. Because same-sex couples comprise a small portion of the population, different-sex couples with errors can constitute a large proportion of same-sex couples. Finally, different datasets have different rules about how same-sex couples can report their marital status, and these rules have changed over time [4,6,14]. For example, prior to 2010 the CPS would change the sex of one of the persons in any reported same-sex married couple; after 2010 they would change the marital status of the members of the couple.

The new IPUMS family interrelationship variables are very accurate when identifying same-sex couples. Of those identified by IPUMS as same-sex couples in both IPUMS CPS and IPUMS NHIS, 100% match the self-reported spouse/partner location. As shown in Fig. 1, the IPUMS family interrelationship variables consistently identify between 0.4% and 1% of couples in IPUMS CPS and 1% of couples in IPUMS NHIS as being same-sex. These percentages are consistent with other estimates of reported same-sex couples [14]. The increases in same-sex couples in 2010 and 2015 recorded by IPUMS CPS correspond to CPS changes in editing procedure and relationship-to-head question [4,5].

The IPUMS family interrelationship variables will carry through any errors in the data; that is, they may identify a couple as same-sex because of an error in the reported sex of one member. To examine how many heterosexual people are identified in same-sex couples, we use recent years of IPUMS NHIS, which include self-reported sexual orientation for the sample adult. In 2013 and 2014, 87.57% of

Table 9
The proportion of IPUMS parental locations that match self-reported parental location (left columns) and the proportion of people in each "Relationship to head" (right columns)

Relationship to householder	Match self-reported mother	Match self-reported father	Proportion of total	Proportion of those under 18
Head/householder	100.0%	100.0%	39.2%	0.3%
Spouse	99.3%	99.7%	19.3%	0.01%
Child	99.3%	98.4%	29.9%	88.8%
Parent	99.3%	99.8%	1.5%	0
Sibling	99.6%	99.1%	1.3%	0.9%
Grandchild	87.6%	91.1%	1.9%	6.3%
Other relatives	93.2%	96.3%	1.9%	1.9%
Unmarried partner	100.0%	100.0%	2.1%	0.02%
Housemate/roommate	99.0%	99.1%	1.6%	0.3%
Roomer/boarder/lodger	98.9%	98.5%	0.4%	0.1%
Foster children	100.0%	100.0%	0.1%	0.3%
Other nonrelatives	89.3%	78.6%	1%	1.2%

IPUMS-CPS ASEC 2007–2015.

people in same-sex couples identified by IPUMS family interrelationship variables self-reported being gay, lesbian, or bisexual. Only 7.95% report being heterosexual and 4.5% answered with "something else", "unknown", or "don't know". This suggests that for IPUMS NHIS, relatively few same-sex couples are the result of an error in the sex variable.

3.2. Testing the new family interrelationship variables: Matching self-reported interrelationships

The IPUMS family interrelationship variables are very accurate. The IPUMS spouse/partner variable identifies the same person as the self-reported variable over 99.99% of the time (IPUMS CPS ASEC 2007–2015, IPUMS NHIS 2006–2014). The mother and father links are likewise very accurate, consistently identifying the self-reported mother 99% of the time and father 98%–99% of the time (IPUMS CPS ASEC 2007–2015, IPUMS NHIS 1998–2014).[7]

Table 9 shows that the match between self-reported mother and father location is high across all relationship-to-head categories, although it is noticeably lower for "non-relatives" and "grandchildren". IPUMS links "non-relative" children to both the "householder" and the "unmarried partner", while CPS and NHIS will only link such children to the "unmarried partner".

[7]The self-reported family interrelationship variables in IPUMS CPS refer to LINENO, the line number in the original CPS enumeration. In IPUMS NHIS, the self-reported family interrelationship variables refer to PX, the original NHIS person number. PX is not necessarily unique without household, but it is unique within family (identified with FMX). The IPUMS family interrelationship variables refer to PERNUM. To compare the self-reported family interrelationship variables to IPUMS family interrelationship variables, users will need to compare the PERNUM of the person identified with the self-reported family interrelationship variable, rather than the value of the self-reported family interrelationship variable itself.

The lower match rate for "grandchildren" is driven by IPUMS variables linking a "grandchild" to a "child" who is the correct age to be a potential parent, but is in fact the aunt or uncle of the "grandchild". For example, in IPUMS CPS we link 87.6% of "grandchildren" correctly to their mothers.[8] 6.5% of "grandchildren" do not have a co-resident mother in the household, but do live with another person who is the correct age and is a "child". IPUMS will identify the "child" (erroneously) as the mother of the "grandchild", when in fact she is likely the aunt of the "grandchild". 5.8% of "grandchildren" have a co-resident mother in the household, but IPUMS links to a different woman who is an eligible age and relationship to be their mother.

The rule variables signal how confident a researcher can be in the link contained in the parent and spouse/partner location variables. For example, those who are linked to a mother through a direct relationship (relationship clarity level 1) with logical step 1 have a MOMRULE value of 11 and are correct 99% of the time in IPUMS CPS. Those linked with a less direct relationship (relationship clarity level 2), but in households where the potential parents include only one married couple (logical step 2) have a MOMRULE value of 22 and are correct 91% of the time. Links made among higher relationship priority levels (relationship pairs with more ambiguity) and assigned under later logical steps (steps addressing harder decisions among multiple potential links) are less often correct than links assigned in the face of less amiguity. The nuanced rule variables thus enable researchers to understand how the parent location variables were calculated and the degree of uncertainty in their assignment.

3.3. Testing the new family interrelationship variables: Teenage fertility

As an empirical test, we used the new IPUMS family interrelationship variables in IPUMS CPS to track the well-established fall in teenage fertility that has occurred since the early 1990s. Teenage parents are hard to identify because they are less likely to head their household and be married than other parents. Moreover, the CPS does not have a variable directly indicating the number of births a woman has had, and provides limited detail in the relationship-to-householder variable. Identifying parents who may not be householders and may not be in couples is a challenging test for family interrelationship variables.

Beginning in the early 1990s, there has been a sustained drop in teenage fertility in the U.S. [15]. The observed decline was particularly strong for Latina and African American women [16]. If the IPUMS family interrelationship variables correctly link children and parents, we would expect that patterns of young parenthood as measured by the new IPUMS family interrelationship variables would reflect the patterns observed in teenage fertility. Note that because parenthood is a long-term

[8] This includes when IPUMS correctly links to the self-reported mother of the "grandchild" and when IPUMS correctly does not link those without a co-resident mother.

Table 10
The proportion of new IPUMS family interrelationship variables that match original IPUMS family interrelationship variables

	Mother location	Father location	Spouse/partner location
Old and new match perfectly	98.79%	98.71%	95.4%
Same-sex couple (matches one parent)	0.05%	0.01%	0.37%
Cohabiting couple (matches one parent)	0.66%	0.89%	4.21%
Doesn't match for other reasons	0.5%	0.39%	0.03%

IPUMS-CPS ASEC 2010–2015. *Figures use appropriate sampling weights.*

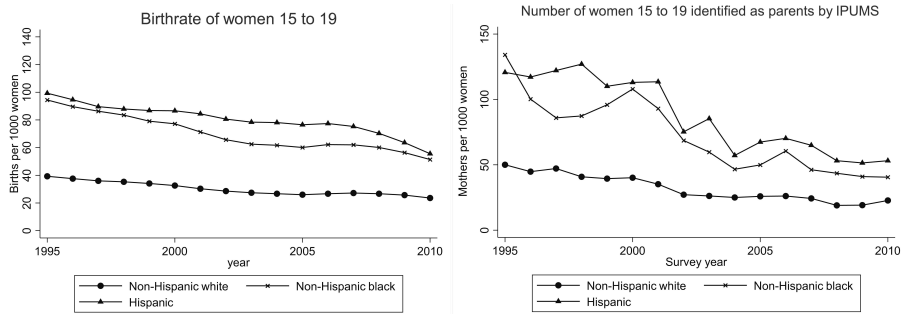


Fig. 2. Teenage birthrate from 1995 to 2010 (left) and the number of teenage women per 1000 who are identified as parents with the new IPUMS family interrelationship variables from IPUMS-CPS (right). *Left hand graph is based on data from [16] in Table 1, available at www.guttmacher.org/pubs/USTPtrends 10.pdf.*

status, the decrease in young women who are parents will lag behind the fall in teenage fertility.

The graph on the left in Fig. 2 shows the fall of the teenage birthrate beginning in 1995 (graph data in [16]). The graph on the right in Fig. 2 closely tracks this pattern, with a decrease in the proportion of teenage women aged 15 to 19 who are identified as parents by the new IPUMS family interrelationship variables. The noticeably larger decrease among Latina and African American women is also replicated by the IPUMS family interrelationship variables.

3.4. Comparison to original IPUMS family interrelationship variables

The new IPUMS family interrelationship variables improve on the original variables by including cohabiting and same-sex couples and increasing their comparability across IPUMS data projects, but researchers who have used the original variables may be concerned about differences between the two. The new family interrelationship variables agree with the original variables the vast majority of the time. As shown in Table 10, in the 2010–2015 ASEC, the parent location variables match the original variables for over 98.7% of respondents. The majority of respondents with non-matched parent locations do not assign a different parent, but rather an

additional parent. The original variables only assigned one parent to children of cohabiting or same-sex couples, while the new interrelationship variables identify that parent's spouse/partner as well.

In comparing the old and new IPUMS links, the spouse/partner location variable exhibits the largest change: 4.58% of respondents are in a cohabiting or same-sex relationship that was unidentified in the original variables. These newly included cohabiting and same-sex couples match self-reported spouse/partner location 100% of the time. These cases represent a sizable proportion of the sample and emphasize the importance of having inclusive family interrelationship variables.

4. Conclusions

In this study, we describe the new IPUMS family interrelationship variables that indicate a household member's probable spouse/partner and parent(s). These new family interrelationship variables build on the groundbreaking original IPUMS family interrelationship variables by extending both parent and spouse/partner location to include same-sex and cohabiting couples. The original IPUMS family interrelationship variables were becoming less accurate over time because of changes in family structure and in how families are enumerated in datasets. For example, nonmarital fertility and other complex families are increasingly common, making parents harder to identify than in the past. Additionally, families are becoming more diverse, particularly with data collection agencies increasingly reporting cohabiting and same-sex couples. It is important that researchers are able to accommodate many family types while also making clear and consistent assumptions.

The new IPUMS family interrelationship variables employ a two-stage protocol for addressing ambiguity when assigning family interrelationship variables. Both steps in the new protocol account for same-sex and cohabiting couples, allowing the links to more accurately reflect diverse family types. The protocol first prioritizes links based on the clarity of the relationship between the two people being linked. This first step is specific to each data project because of variation in the level of detail about a respondent's relationship to the householder. The second stage of the protocol uses a set of logical rules to select between multiple potential links. In a detailed appendix, we describe the development of these logical steps. Detailed rule variables indicate the level of ambiguity encountered at each stage of the protocol. These composite rule variables allow researchers to use only those family interrelationships that are based on the assumptions acceptable for their analyses and to communicate these details to other researchers.

To assess the performance of the new IPUMS family interrelationship variables, we implemented two tests. First, we compared the new IPUMS family interrelationship variables to self-reported parent and spouse/partner locations included in recent samples of the CPS and NHIS. IPUMS variables match the self-reported parent and spouse/partner locations very well, including among same-sex couples. Second, we

showed that the new IPUMS family interrelationship variables track racial and temporal variation in the teenage birthrate. They were able to replicate the fall in teenage fertility that began in the 1990s, including the more dramatic decrease for African American and Latina women. These finding suggest that the new IPUMS family interrelationship variables are able to capture trends in changes to family structure, even for a parental relationship that is not often explicitly identified in the dataset. Additionally, a direct comparison of the new and original IPUMS family interrelationship variables confirms that the new variables retain parent and spouse links from the original variables while adding to their power by including same-sex and cohabiting couples.

References

[1] Ruggles S. Family interrelationships. Historical Methods 1995; 28(1): 52-58.
[2] Sobek M, Kennedy S. The development of family interrelationship variables for international census data. MPC Working Paper Series 2009. https://international.ipums.org/international/resources/misc_docs/pointer_working_paper_2009.pdf.
[3] Smock PJ. Cohabitation in the United States: An appraisal of research themes, findings, and implications. Annual Review of Sociology 2000; 26: 1-20.
[4] Lofquist D, Ellis R. Comparison of estimates of same-sex couple households from the ACS and CPS. U.S. Census Bureau 2011. https://www.census.gov/hhes/samesex/files/2011Final_PAA_Poster.pdf.
[5] Lofquist D, Lewis J. Improving measurement of same-sex couples. SEHSD Working Paper 2015-13. U.S. Census Bureau 2015. https://www.census.gov/hhes/samesex/files/Lofquist_Lewis_2015-13.pdf.
[6] National Center for Health Statistics. Changes to data editing procedures and the impact on identifying same-sex married couples: 2004-2007 National Health Interview Survey. Centers for Disease Control and Prevention, 2015. http://www.cdc.gov/nchs/data/nhis/samesexedits.pdf.
[7] Flood SM, King M, Ruggles S, Warren JR. Integrated public use microdata series, current population survey: Version 4.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2015. https://cps.ipums.org/.
[8] Minnesota Population Center and State Health Access Data Assistance Center, Integrated Health Interview Series: Version 6.11. Minneapolis: University of Minnesota, 2015. http://www.ihis.us.
[9] Hofferth SL, Flood SM, Sobek M. American time use survey data extract system: Version 2.4 [Machine-readable database]. Maryland Population Research Center, University of Maryland, College Park, Maryland, and Minnesota Population Center, University of Minnesota, Minneapolis, Minnesota, 2013. http://www.atusdata.org.
[10] Ruggles S, Genadek K, Goeken R, Grover J, Sobek M. Integrated public use microdata series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2015. https://usa.ipums.org/usa/.
[11] StataCorp. Stata statistical software: Release 13. College Station, TX: StataCorp LP 2013.
[12] Black D, Gates G, Sanders S, Taylor L. The measurement of same-sex unmarried partner couples in the 2000 U.S. Census. On-Line Working Paper Series 2007. http://escholarship.org/uc/item/72r1q94b.
[13] Kreider RM, Lofquist D. Matching survey data with administrative records to evaluate reports of same-sex married couple households. SEHSD Working Paper #: 2014-36, 2015. https://www.census.gov/hhes/samesex/files/Kreider-Lofquist-Working-Paper.pdf.
[14] U.S. Census Bureau. Frequently asked questions about same-sex couple households, 2013. http://www.census.gov/hhes/samesex/files/SScplfactsheet_final.pdf.

[15] Boonstra HD. What is behind the declines in teen pregnancy rates? Guttmacher Policy Review 2014; 17(3). http://www.guttmacher.org/pubs/gpr/17/3/gpr170315.html.
[16] Kost K, Henshaw S. U.S. teenage pregnancies, births and abortions, 2010: National and State trends by age, race and ethnicity. 2014. http://www.guttmacher.org/pubs/USTPtrends10.pdf.
[17] Anderson SE, Dallal GE, Must A. Relative weight and race influence average age at menarche: Results from two nationally representative surveys of US girls studied 25 years apart. Pediatrics 2003; 111(4): 844-850.

Appendix 1:

Developing a set of logical steps: Using the CPS fertility supplement

The logical steps described in the main text where we select between multiple potential parents are based on analysis of two additional sources of data on family interrelationships. We first use the CPS supplements that focus on fertility. Second, we use the American Time Use Survey which has been linked to CPS. In this section, we discuss the CPS fertility supplement survey, which provides insight into how to assign parent links when there are multiple potential parents.[9] These supplements do not explicitly identify family relationships, but they do contain information about the number of live births for each woman (aged 15 to 44) in a household. This information can be used to infer patterns in family relationships. In the absence of intentional ordering of persons within a household, supplemental information is central to making appropriate inferences. Because the CPS fertility supplement only has information on women, it is used to determine priority level among female potential parents. All analyses are performed with STATA 13 [11].

In assigning parents to children, we privilege married parents and then women who were previously married if there are no eligible married parents in the household. This step considers that married and ever married women are more likely to have given birth than never married women. Figure A1 shows the percent of female "children" who have had one or more live births in households with one "grandchild" and two female "children". The vast majority of those who are currently married or were previously married have given birth, whereas the majority of single women have not.

When there are no previously married potential parents, the next most likely potential parent is the eldest single woman. Figure A2 illustrates the logic for this step by showing the percent of eldest and youngest daughters who have given birth in households with two female "children" – the older daughter is far more likely to have given birth than the younger.

[9]The CPS and NHIS both include self-reported family interrelationship variables for recent years. Because we use the self-reported variables to examine the accuracy of the IPUMS family interrelationship variables, we do not use them to develop our algorithm.

Table A1
Ambiguous relationship between grandchildren and children of householder

Person number	Relationship to householder	Sex	Marital status	Age
1	Householder	Female	Married/in union	42
2	Spouse	Male	Married/in union	46
3	Biological child	Female	Single/never married	23
4	Biological child	Female	Single/never married	18
5	Grandchild	Male	Single/never married	6
6	Grandchild	Male	Single/never married	3

(Serial 1396426000).

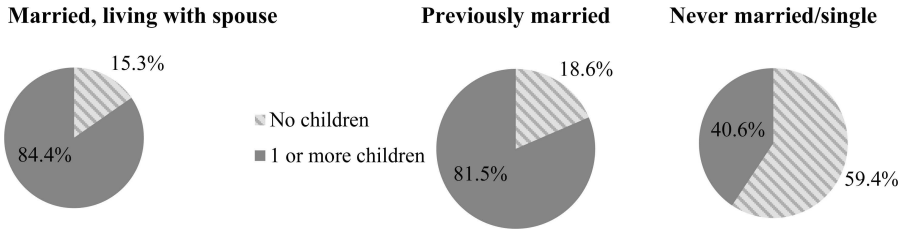


Fig. A1. Number of live births by marital status of female children in households with one grandchild and two female children. IPUMS CPS fertility supplements 1994, 1995, 1998, 2000, 2002, 2004, 2006, 2008, 2010, and 2012. $n = 1{,}293$. Calculated using supplement weights.

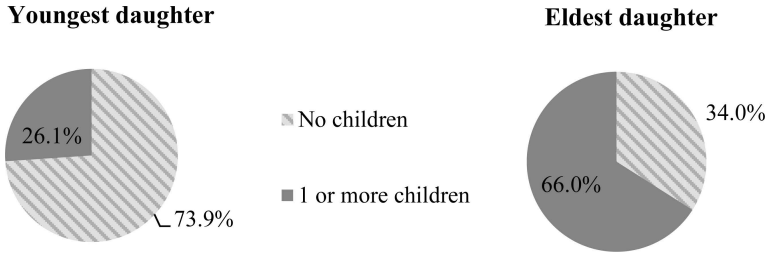
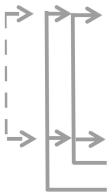


Fig. A2. Distribution of number of live births by age of female children in households with one grandchild and two female children. IPUMS CPS fertility supplements 1994, 1995, 1998, 2000, 2002, 2004, 2006, 2008, 2010, and 2012. $n = 1{,}293$. Calculated using supplement weights.

When there are both multiple "grandchildren" and multiple "children" in a household, it can be unclear if all the "grandchildren" are siblings (one "child" is the true parent for all "grandchildren") or cousins (multiple "children" are the true parents to the "grandchildren"). For example, in a household with two female "children" and two "grandchildren" it is unclear if or how the "grandchildren" should be split among the biological "children" of the "householder". Consider the household in Table A1.

It may be tempting to link the three-year-old "grandchild" with the 18-year-old "child" and the six-year-old "grandchild" with the 23-year-old "child". However, Fig. A3 shows that among households with two daughters and two grandchildren, it is more likely that the eldest daughter gave birth to two children and the youngest

Table A2
Parallel relationship values from the CPS and ATUS



Person number	Age	Sex	Relationship to CPS householder	Relationship to ATUS respondent	Marital status
1	27	F	Child	Self	Married
2	57	M	Householder	Parent	Married
3	57	F	Spouse	Parent	Married
4	25	F	Child	Brother/Sister	Never married
5	23	M	Child	Brother/Sister	Never married
6	30	M	Other relative	Spouse	Married
7	2	M	Grandchild	Child	Not in universe
8	1	F	Grandchild	Child	Not in universe

(caseid 20081211081682)

to none. That is, assigning both "grandchildren" to the eldest female "child" is more likely to be correct in this situation.

Using the linked ATUS-X and IPUMS-CPS datasets to determine the most probable parent

In addition to the fertility supplement in the IPUMS-CPS, the American Time Use Survey (ATUS) also offers insight into patterns of the most probable parent in ambiguous relationships. Notably, the IPUMS CPS fertility supplements only offer insight into which women are the likely parents in ambiguous situations; it offers no help in selecting between men and women or among male potential parents. For this question, we turn to the linked ATUS-X/IPUMS CPS datasets.

The ATUS surveys a subsample of responding CPS households after their last interview in the CPS. The CPS survey contains a variable showing how all household members are related to the CPS householder. The ATUS then randomly selects one household member who is at least 15 years old and asks them to complete a time diary of their daily activities. The ATUS includes a variable showing how all household members are related to the ATUS respondent – who is not necessarily the householder from the CPS. The reframing of family relationships to a different reference person provides additional information that can be used to determine the most probable parent and spouse in ambiguous situations.

Table A2 shows an example of how the parallel information from the ATUS can be used to determine parental and spousal relationships in ambiguous situations. In the example household, persons 7 and 8 are "grandchildren" of the CPS householder (person 2), and there are three "children" of the householder who are potential parents. There is also an "other relative" who is married, likely to one of the children. The ATUS randomly selected the first "child" of the CPS "householder" to be the ATUS respondent. The new ATUS-specified reference person provides clear and unique relationship values that indicate both "grandchildren" from the CPS (persons 7 and 8) are the children of person 1 (shown in solid arrows). It also shows that person 6 is the "spouse" of person 1 (shown in dashed arrows).

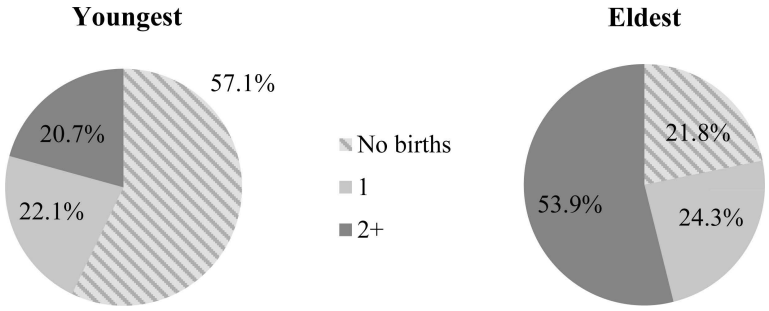


Fig. A3. Distribution of number of live births in households with two grandchildren and two female children, by age of female child. IPUMS CPS fertility supplements 1994, 1995, 1998, 2000, 2002, 2004, 2006, 2008, 2010, and 2012. $n = 550$. Calculated using supplement weights.

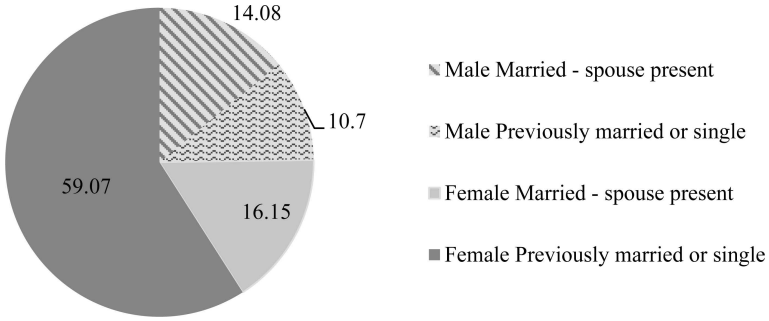


Fig. A4. Sex and marital status of those identified as a parent to a CPS grandchild using linked IPUMS CPS and ATUS-X datasets. ATUS-X/IPUMS CPS linked datasets 2003–2014. $N = 1{,}158$. Weighted by household size.

By using this dual relationship classification for all people listed as a grandchild in the CPS, we can observe patterns in who is most likely their parent. Figure A4 shows the confirmed parent of a "grandchild" when using the IPUMS CPS/ATUS-X linked data. Married men and women are about equally likely to be the parent. Among those who are not married, women are far more likely than men to be the parent: divorced, separated, and never-married women account for over half of the 1,158[10] identified parents of "grandchildren", but unmarried men are only 10.7%.[11]

In households with a previously married female "child", she is confirmed by the linked ATUS-X/IPUMS CPS data as the "grandchild's" parent 92% of the time. It is uncommon for a household to have a previously married male "child" and a "grand-

[10] The 1,768 grandchildren live in 1,158 households.

[11] Because the probability of a grandchild's parent being selected as the ATUS respondent depends on the number of eligible people in the household, all figures are weighted by the number of eligible respondents in the household.

child", but when this does occur, the male "child" is the parent to the "grandchild" 75% of the time. It is extremely rare for a household to have a "grandchild", a previously married male "child", and a previously married female "child", but in those cases we assign parent links to previously married female potential parents before previously married male potential parents.

We then face a particularly difficult decision: in the case where there is a previously married male potential parent and a single woman, who is more likely to be the true parent? The ATUS-X/IPUMS CPS linked datasets shows that among households with a grandchild and a single/never married daughter, she is the parent 84% of the time. There are very few households with a "grandchild", a single "daughter", and a previously married "son", so we cannot perform a direct comparison. In these rare cases, we chose to assign previously married male potential parents prior to single female potentially linked parents.[12] Because of the logic involved in assigning these types of parent links, both these steps have high values on the second digit of the parental rule (6 and 7, respectively).

Conclusions from IPUMS-CPS fertility supplement and linked ATUS-X/IPUMS-CPS datasets about grandchildren

Together, the patterns highlighted above yield clear rules about how to allocate "grandchildren" among "children" in a household. While a married couple living with one member's parents is uncommon, they are the most likely "children" to have had children, and are therefore the first choice for a "grandchild's" linked parents. If there is more than one married child and child-in-law couple, "grandchildren" are split among them by age. Because previously married women are more likely to be the true parent than previously married men, if there are no married couples present, "grandchildren" are split first among ever-married female "children". They are split among ever-married male "children" only if there are no ever-married female "children" present. If there are no ever-married "children", "grandchildren" are assigned to the eldest female "child". Single male "children" are the last linked parent selected among multiple potential parents.

While not shown here for brevity, the patterns with respect to marital status and sex that inform parent-child links among "children" and "grandchildren" hold among other relationship categories as well. This priority ranking is employed within all relationship clarity levels of the family interrelationship variables.

Determining age difference between parent and child

In order to reduce incorrect parent locations, we take into account the age difference between a potential parent and their child. There are biological constraints that

[12]While this choice is difficult, we are rarely forced to make it. In the 2007–2015 ASEC samples, 0.77% of grandchildren live in a household with both a single daughter and a previously married son and no married children. This is a total of 290 grandchildren in 177 households.

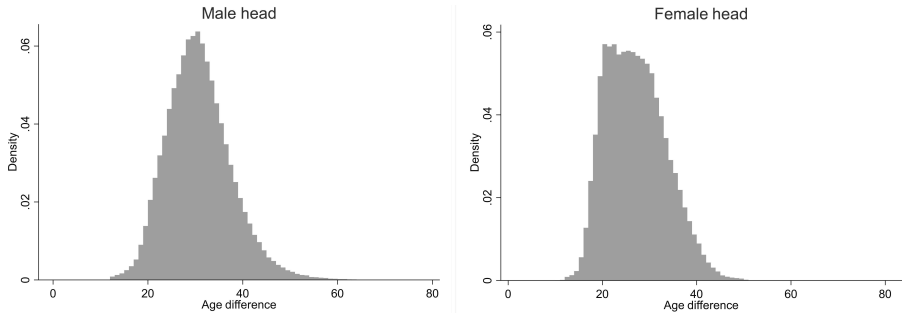


Fig. A5. Age difference between householder and child of householder in the 2010 ACS.

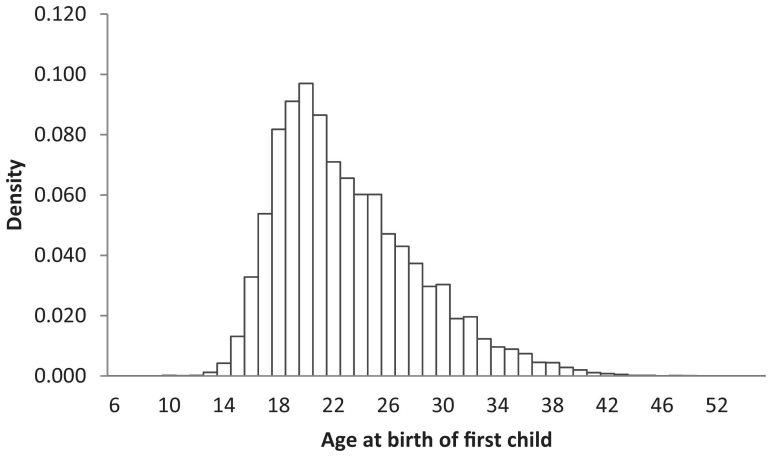


Fig. A6. Mother's age at birth of first child in pooled 2000, 2005, and 2010 NHIS.

make giving birth to a child more likely during certain age periods. For example, average age at menarche in the United States is 12.5 years [17].

To determine acceptable parent-child age differences for linking purposes, we examine two patterns. First, we examine the age difference between the householder and the child of the householder in the ACS. Because this relationship is direct and unambiguous, we use these patterns to determine the appropriate age cutoffs for parental relationships in more ambiguous settings. Second, we observe age at first birth from the NHIS to test the lower bound of the age cutoff.

As shown in Fig. A5, women who are parents are concentrated between 15 and 44 years older than their children. Figure A5 demonstrates that men have a longer right tail; a small but notable portion of male parents are 50 or more years older than their children. Reflecting these patterns of age difference, the linking protocol requires potential parents identified through less direct relationship pairings (links formed in levels 2 and higher in Table 4) to be have a specific age difference from

their potential children: women must be between 15 and 44 years older than their child while men identified as parents must be between 15 and 60 years older than their children. Those identified with direct relationships (level 1 in Table 4 – "child" to "householder" for example) face no age difference restrictions.

Figure A6 shows requiring mothers to be at least 15 years older than their children is consistent with patterns of fertility: very few women report their first birth prior to age 15.

Appendix 2

The ATUS does not ask marital status and instead uses the marital status from the last CPS sample (ATUS respondents are a subsample of outgoing CPS respondents). That is, a person who was married between the CPS and the ATUS interview will report having a spouse, but their marital status will be out of date. For this reason, in the ATUS the link between a head and spouse will occur in both level 1 (those with a "Married, spouse present" marital status) and in the 2nd level (those with any other marital status).

Table A3
Spouse links for IPUMS time use ATUS-X and IPUMS USA

	ATUS-X		IPUMS USA (ACS and census 1970 and later)	
1st level links	Respondent	Spouse	Householder	Spouse
SPRULE value: 1-	Parent	Parent	Child-in-law	Child, step-child, adopted child
	Housemate	Housemate	Sibling-in-law	Sibling
	Roomer	Roomer	Parent	Parent
	Non-relative	Non-relative	Parent-in-law	Parent-in-law
			Housemate	Housemate
			Roomer	Roomer
			Non-relative	Non-relative
			Aunt/uncle	Aunt/uncle
			Roomer/boarder	Roomer/boarder
			Partner/friend	Partner/friend
			Partner/roommate	Partner/roommate
2nd level links	Respondent	Unmarried partner, spouse	Householder	Partner
SPRULE value: 2-				
3rd level links	Other relative	Other relative, grandchild, child, brother/ sister	Other relative	Sibling-in-law, other relative, grandchild, nephew/niece, cousin, child, step-child, adopted child, sibling
SPRULE value: 3-	Non-relative	Roomer, housemate	Non-relative	Roomer, ward, housemate, partner/roommate
			Unknown	Unknown

Table 3, continued

	ATUS-X		IPUMS USA (ACS and census 1970 and later)	
4^{th} level links SPRULE value: 4-	Householder	Other relative, non-relative	Other relative	Grandparent, aunt/uncle, parent, householder
	Child	Child	Child	Child
	Grandchild	Grandchild	Grandchild	Grandchild
	Unknown	Roomer, housemate	Nephew/niece	Nephew/niece
	Brother/sister	Brother/sister	Cousin	Cousin
			Unknown	Roomer, ward, housemate
			Sibling-in-law	Sibling-in-law
			Sibling	Sibling
5^{th} level links SPRULE values: 5-			Householder	Non-relative, other relative

Table A4
Parental links for IPUMS time use ATUS-X and IPUMS USA

	ATUS-X		IPUMS USA (ACS and census 1970 and later)	
Direct links Parental rule value: 1-	Child, non-household child	Respondent	Child, adopted child	Householder
	Respondent	Parent	Householder	Parent
	Brother/sister	Parent	Step-child	Householder
			Sibling	Parent
			Spouse	Parent-in-law
2nd level links Parental rule value: 2-	Grandchild	Child, non-household child	Grandchild	Child, adopted child, stepchild
	Child, non-household child	Spouse, partner	Parent	Grandparent
			Child, adopted child, stepchild	Spouse, partner
			Sibling-in-law	Parent-in-law
			Cousin	Aunt/uncle
			Nephew/niece	Sibling, sibling-in-law
3rd level links Parental rule value: 3-	Housemate	Housemate	Housemate	Housemate
	Roomer	Roomer	Roomer	Roomer
	Other relative	Other relative, sibling	Other relative	Other relative, niece/nephew, sibling, sibling-in-law
	Non-relative	Partner, non-relative	Non-relative	Non-relative, partner, partner/roommate
	Unknown	Unknown	Grandchild	Child-in-law
			Partner/roommate	Partner/roommate, employee
			Employee	Employee
			Roomer/boarder	Roomer/boarder
			Partner/Friend	Partner/friend, employee
			Unknown	Unknown

Appendix 3

As shown in the following graph, among those "Other relatives" who are 19 to 22 that IPUMS links to a mother, the majority match their self-reported mother. This decreases with age: those who are 22 years old are only correct 58% of the time. However, because this is over 50%, we are more accurate linking these "Other relatives" than not linking.