MINNESOTA POPULATION CENTER

MPC

UNIVERSITY OF MINNESOTA

Working Paper Series

The Development of Family Interrelationship Variables for International Census Data

Matthew Sobek and Sheela Kennedy
Minnesota Population Center

November 18, 2009

Working Paper No. 2009-02
https://doi.org/10.18128/MPC2009-02

Acknowledgements: Support for this research was provided by the National Science Foundation and the National Institute of Child Health and Human Development. Vladimir Vladykin programmed the computer algorithms identifying parents and spouses, and Monty Hindman provided invaluable technical and design assistance. ChangHwan Kim, Toni Lopez, Phuong Nguyen, Krishna Mohan Palipudi, and Le Wang were instrumental in developing and testing the linking rules. We are grateful to the following statistical agencies for providing the microdata making this research possible: Argentina, National Institute of Statistics and Censuses; Armenia, National Statistical Service; Austria, National Bureau of Statistics; Belarus, Ministry of Statistics and Analysis; Bolivia, National Institute of Statistics; Brazil, Institute of Geography and Statistics; Cambodia, National Institute of Statistics; Chile, National Institute of Statistics; China, National Bureau of Statistics; Colombia, National Administrative Department of Statistics; Costa Rica, National Institute of Statistics and Censuses; Ecuador, National Institute of Statistics and Censuses; Egypt, Central Agency for Public Mobilization and Statistics; France, National Institute of Statistics and Economic Studies; Ghana, Ghana Statistical Services; Greece, National Statistical Office; Guinea, National Statistics Directorate; Hungary, Central Statistical Office; India, Ministry of Statistics and Programme Implementation; Iraq, Central Statistical Office; Israel, Central Bureau of Statistics; Italy, National Institute of Statistics; Jordan, Department of Statistics; Kenya, National Bureau of Statistics; Kyrgyz Republic, National Statistical Committee; Malaysia, Department of Statistics; Mexico, National Institute of Statistics, Geography, and Informatics; Mongolia, National Statistical Office; Palestine, Central Bureau of Statistics; Panama, Census and Statistics Directorate; Philippines, National Statistics Office; Portugal, National Institute of Statistics; Romania, National Institute of Statistics; Rwanda, National Institute of Statistics; Slovenia, Statistical Office; South Africa, Statistics South Africa; Spain, National Institute of Statistics; Uganda, Bureau of Statistics; United Kingdom, Office of National Statistics; United States, Bureau of the Census; Venezuela, National Institute of Statistics; and Vietnam, General Statistics Office.

Abstract: Population microdata are typically organized into households, but household relationships are often ambiguous for persons outside the nuclear family. To facilitate comparative research on families and households, the Minnesota Population Center has developed consistent "pointer" variables identifying each person's mother, father and spouse for the International Integrated Public Use Microdata Series (IPUMS-International), a freely available database of 279 million person records from 44 countries. This paper documents the methodology used to identify the most likely parent-child and spouse pairings. We show that the IPUMS pointers agree with direct reports of family interrelationships more than 98% of the time, and highlight for researchers factors that affect the precision of these links. These new variables provide researchers with a common tool for studying family interrelationships, removing the possibility that differing results are an artifact of different linking procedures. A significant fraction of recent IPUMS data requests include the pointer variables, suggesting a growing body of population research will depend, at least in part, on the quality of these links.

Introduction

Census microdata are among the most widely used sources in population research. Microdata describe the characteristics of individuals and give researchers the freedom to calculate their own measures of demographic and social phenomena. In most census datasets individuals are organized into households, and the relationships among individuals are known. This hierarchical structure gives the data much of its power. Researchers can combine the characteristics of related and co-resident persons to create a wide range of new variables and measures, and can analyze their effects at the individual level. Constructed variables might include the age of a person's spouse, the school attendance of a father's children, or the number of own children present for each adult woman in the household.

The ability to create variables from multiple person records is essential for many analyses, but it is an inherently difficult task. Censuses typically identify each person's relationship to a reference person in their household, but the relationships to other persons are often ambiguous. Some people are grouped into residual "other relative" and "non-relative" categories, and even persons with a specified relationship like "grandchild" might have more than one potential mother or father. Adequately determining family interrelationships requires using a number of variables in combination and considering factors like persons' proximity within the household roster. Many researchers are unable to carry out these methods in a statistical package, but even those with sufficient skill to make the links will inevitably use differing methods, because of the many decisions embedded in such techniques. Consequently, there will always be uncertainty about the extent to which differing results between studies are artifacts of the linking process.

Consistent family interrelationship measures are especially critical for comparative research. This need is addressed by the IPUMS-International project, the world's largest collection of publicly available census microdata.[1] The International Integrated Public Use Microdata Series consists of 279 million person records in 130 census samples from 44 countries (Minnesota Population Center 2009). Family relationship variables recently have been developed for 115 of these samples.[2] These "pointers" are designed to produce a consistent, but versatile set of links between immediate family members. By capitalizing on the hierarchical structure of the data, the pointers give researchers the flexibility to define their own measures of family and household composition and to interrelate the characteristics of family members in complex ways. These new measures are already having an impact: since their development in 2008, 40 percent of IPUMS data extracts have included one of the pointers or variables derived from them.[3]

This paper has three primary purposes. The most basic goal is to inform population researchers about the availability of these measures. They can save scholars considerable time and effort or serve as a benchmark against which to assess their own linking procedures. The second aim is to help researchers understand the method by which the pointers were produced. A number of decisions were made in the development of these links, and some will affect particular types of analyses. Finally, we want to identify the factors influencing the comparability of the pointers across samples. This includes assessing how they perform in comparison to equivalent pointer information that was collected directly from census respondents in a number of countries.

[1] The IPUMS-International data series is continually growing and evolving. The discussion in this paper pertains to the database and its constructed variables as of fall 2009 (Minnesota Population Center 2009).
[2] Linking variables could not be constructed for some datasets because the person records were not organized into households or because they lacked a critical variable for making the links.
[3] Pointers have been available for a small subset of IPUMS samples since 2005, but they did not address many of the complexities of the data series and had other limitations, such as not making parental links between adults.

Background

IPUMS-International Project

The IPUMS-International project was developed at the Minnesota Population Center with the goal of preserving, harmonizing and disseminating international census microdata and documentation (Hall, McCaa, and Thorvaldsen 2000). The 2009 version of the data series includes data from 44 countries, and the project has agreements with an additional 38 countries to make their data available in the future. For most countries, data are available for multiple census years.

The database is designed for comparative research. Variables are harmonized across countries, so all samples use consistent codes. No information is lost. For more complex variables, the first digit or two are comparable across samples, while trailing digits retain information unique to particular samples. Integrated documentation describes the comparability issues that cannot be adequately conveyed through variable labeling and coding schemes. All data are available at no charge through a web-based data extraction system that provides pooled extracts containing only the samples and variables requested by researchers. Researchers download the microdata and analyze it themselves on their desktop. Over 3000 researchers from around the world have registered to use the database since its inception.

Individuals are organized into households in 115 samples from 42 countries, and family interrelationship variables have been created for these samples. The full list of these samples is shown in Appendix 1. In addition, 13 samples included a question on the census enumeration form that asks respondents to identify the location (the line number) of each person's spouse and parents. We use these census pointers to evaluate the IPUMS constructed family pointers.

Family Interrelationship variables

Locator variables or “pointers”—variables that identify each person’s mother, father, or spouse, if one is present in the household—are the basis of family interrelationship variables in IPUMS. Consider the 8-person household shown in Table 1. The relationship-to-household-head variable describes a number of family interrelationships. We know the head and spouse are parents of the three children and that the head and spouse are married to one another. For other household members, additional variables must be used to infer relationships, including marital status, the number of children-ever-born, and proximity to each other. The female child-in-law is almost certainly married to the preceding child, because both share the same marital status and because there are no other male children to whom she could be married. The grandchild, however, could be the son of the female child in position 3 (a single mother of one child). More likely, he is the son of the married child and child-in-law listed directly above him in the household.

[Table 1 about here]

Rather than forcing researchers to work through the complex logic to define these family interrelationships, IPUMS constructs the necessary variables using the same program for all samples. These pointer variables give the person number in the household of each individual's mother, father and spouse. Table 2 shows the constructed pointers for the same household described above. The variable SPLOC records the person number of each person’s spouse or partner. In this example, the head and spouse “point” to each other (receiving SPLOC values 2 and 1 respectively). The variables MOMLOC and POPLOC provide the person number of each individual’s parents—so the grandchild in position 7 points to his mother in position 6 and his father in position 5. When no spouse or no parents are identified, the pointer variables are given the value zero.

[Table 2 about here]

Because the same rules are applied across samples, households with similar characteristics in different countries or different years of the same country will receive the same distribution of constructed pointers. Moreover, the pointer variables will be identical for every researcher who downloads IPUMS data. Once SPLOC, MOMLOC, and POPLOC are created, additional family relationship variables are constructed, including the identification of subfamilies, the calculation of the number of children who are linked to particular woman, and the number of families in a household.

All users of the IPUMS database have access to the family interrelationship variables and can easily identify whether a person lives with a spouse or parents or is raising own children. Researchers can then construct their own measures of kin characteristics, family and household composition, fertility and marriage patterns, using the web-based extraction system and standard statistical packages. A feature of the IPUMS data extract system lets researchers attach the characteristics of parents and spouses as new variables on each person's record; thus they never have to use the pointers to perform that matching procedure in a statistical package.

The family presented in Tables 1 and 2 is small, provides detailed relationship information, and requires only one decision—a relatively easy choice between the grandchild’s two possible mothers. Producing family pointers becomes substantially more difficult when the relationship pairings are more ambiguous, when parental absence or adoption occurs commonly, or when there are multiple potential spouses and parents. We needed a method that could identify family relationships simultaneously in a large number of countries that differ greatly in family and household structure and in the detail and quality of data.

Matching algorithms for international and historical census microdata

The origins of family interrelationship inference can be found in the “own-child” method of fertility measurement. First developed in the early 1960s and refined in later years, the own-child method estimates fertility using census data when birth registration data are incomplete or unavailable (Grabill and Cho 1965; Retherford and Cho 1978; Retherford, Cho, and Kim 1984; Luther and Cho 1988). Within each census household children are matched to mothers, using an algorithm that incorporates demographic data usually collected during census enumeration: relationship to household head, age, marital status, and the number of surviving children, when available.[4] Reverse survival methods are then used to estimate the number of children born in a particular year, as well as the population of women by age. From this, single-year age-specific fertility rates can be calculated for periods up to 15 years prior to census enumeration.

Own-child methods have been used widely to estimate international and historical fertility levels. Researchers continue to use these methods when birth registration data are not available, often to provide estimates of historical trends in fertility (Retherford et al. 2005; Hacker 2003). Comparisons have found that own-child matching yields similar population level fertility estimates as direct reports of mother-child relationships, even in samples with complex families, high rates of adoption, and a high rate of mismatches (Levin and Retherford 1982; Cho et al. 1986). Although individual-level errors tend to cancel out when aggregated, errors rates can be high at the extreme ends of the reproductive age range. More complex matching procedures

[4] Examples of own-child matching programs are included in Cho, Retherford, and Choe (1986).

have since been developed, but have not been implemented widely (Zuberi and Sibanda 1999; Strong et al. 1989).

With the 1995 release of integrated microdata files for eleven U.S. censuses, IPUMS-USA advanced the process of identifying family relationships (Ruggles 1995; Ruggles et al. 2009). Family interrelationship variables were reconceived as a multi-purpose tool and made available to researchers in public use samples. IPUMS-USA provided additional family pointers not included in own-child methods (links between spouses and between children and their fathers) and constructed additional family and household descriptors. The linking algorithm had to be flexible enough to deal with differing variable availability and changing category detail across census years. Each pointer variable was accompanied by a "rule" variable describing the criteria used to assign the spouse or parent link. The resulting family interrelationship variables have allowed researchers to study a variety of topics, including historical estimates of family and household composition and studies of family structure and child wellbeing (Ruggles and Brower 2003; Moehling 2004, 2007; Short, Goldscheider, and Torr 2006; McGarry and Schoeni 2000; Lichter, Qian, and Crowley 2008).

The IPUMS-USA parental pointers were deliberately conceived to include social parents, not simply biological ones. For one thing, it was not always possible to distinguish between the two, because of differing category and variable availability among samples. More importantly, for many research purposes, links identifying social and economic units are more useful than ones limited to biological connections. Because biological links are sometimes necessary, IPUMS-USA provides supplemental variables that identify whether a given mother or father is likely a stepparent.

IPUMS-International Pointer Design

General principles

We initially considered adapting the IPUMS-USA algorithms for the international IPUMS project, but soon determined that the U.S. model could offer only rough guidance. Like their U.S. counterparts, the international pointers are rule-based, evaluating individual pairings based on relationship to head, age, marital status, fertility (when available), and proximity in the household. The international samples simply had too much variation, however, to follow the specific rules developed for the U.S. censuses: in the reporting order of the enumerated persons, in the categories of the relationship-to-head variable, in the types of marital statuses, and in the quality of the data. The U.S. database's focus on social parentage, rather than strictly biological links, was retained in the international data series.

Perhaps the most important factor governing the development of international family interrelationships is the varying size and complexity of households around the world. Links between the spouse of the head and a child of the head, which are unambiguous in the United States, are less certain in samples with polygamy. Rising rates of non-marital fertility means that matching procedures cannot exclude never-married women when fertility data are unavailable. Likewise, the common presence of extended family members and nonrelatives in developing countries makes family interrelationships more uncertain for a higher proportion of individuals. To illustrate this diversity, Figure 1 presents data on the regional variation in the composition of children’s households. Only half of children in the IPUMS African samples live in a household containing only the head, at most one spouse, and children of the head, compared to over 80 percent of children in the U.S. and Europe. To compound the difficulty, many of the same samples with large numbers of complex households have relatively high rates of data errors in key variables like age, sex, and relationship. Substantial modifications were necessary to improve our links in these more complex households.

[Figure 1 about here]

Also important is variation in the data available to construct the pointers (partially described in Appendix 1, with more detail available on the IPUMS website).[5] Many samples, for instance, do not distinguish parents from parents-in-law or children from children-in-law, or they group grandchildren with other relatives. Data on children ever born or surviving—information that takes on considerable importance when relationship pairings are weak or when there are multiple potential parents—is often unavailable.

The relative position of individuals within households is a strong indicator of family interrelationships in many samples. Censuses commonly instruct enumerators to list household members in meaningful groupings—with spouses listed next to each other and with children directly following their parents. This order can be seen in the household presented in Table 2—the head is listed first, followed by his spouse, his unmarried children (listed by age), his married son's subfamily, and finally by the unrelated employee at the end of the household roster. The IPUMS linking program capitalizes on this common feature of censuses by searching first for adjacent persons. For spouses, adjacency refers to individuals next to one another in the household roster. A woman will link to an adjacent preceding man before linking to an adjacent subsequent man. Adjacency is defined differently for parental links. A potential parent is considered adjacent to a child so long as the parent precedes the child in the household and no persons separate the two except for a spouse and any children who are already linked to that

[5] For information on sample availability of key data used in pointer construction (e.g. relationship categories and fertility) see https://international.ipums.org/international/parrule_table.shtml. Detailed information on the comparability of IPUMS variables across samples is provided in variable descriptions, as well as enumeration text and instructions.

parent. By this definition, all three children in Table 2 are adjacent to the household head and to his spouse. If an adjacent spouse or parent cannot be found, the search continues from the top of the household. The meaningfulness of household order for family interrelationships varies across samples.

The IPUMS-International project emphasizes consistency across samples in the design of family interrelationship variables. Although some customization is necessary to handle particular situations, the same core conditions and basic linking methods are applied across all samples. Each household is evaluated individually. For each of the pointers, the program makes a series of passes looking for a spouse or parent. The strongest possible criteria are applied first to identify the most iron-clad links. Persons who are linked are removed from consideration by the subsequent, weaker passes that use more ambiguous criteria. The specific procedures for linking spouses and children are described below.

Location of Spouse -- SPLOC

The simplest of the family interrelationship pointers is the location-of-spouse variable (SPLOC) that identifies the person number within the household of each individual's co-resident spouse or partner. The spouse pointer is easier to construct than the parental pointers because we know the person's current marital status, spouses generally reside together, and most people only have one spouse. Nevertheless, there are various complications, and the quality of the links varies across samples because of differences among the key variables and in the organization of persons within households.

The basic algorithm for SPLOC restricts the allowable pairings based on age, sex, marital status, and relationship to the household reference person. A linked couple must be of opposite sex and both persons must be age 12 or older. Links can only be made between persons in the same subfamily in the small number of samples that report such subunits. Both persons in a couple must indicate that they are in a marital or consensual union.

Starting with the first record in a household, each person is evaluated using the strongest possible criteria to locate a probable spouse. The strongest criteria involve explicit relationship combinations such as head-to-spouse and parent-to-parent. Subsequent passes use progressively weaker rules to make links—generally based on more ambiguous relationship pairings. At the moment a person is linked they and their spouse are removed from further consideration, thus the order in which the passes are executed is determinative. In most households there is only one possible married couple, and the accuracy of the link is nearly certain. Where there are multiple equally valid potential spouse candidates, the persons' proximity within the household roster is used to choose among them. A separate variable indicates the specific set of conditions under which each link was made. Appendix 2 describes the allowable spouse pairings in each rule.

The biggest challenge in developing the spouse pointer was determining which relationship-to-head categories could link to one another. Theoretically the allowable pairings should be a straightforward inference from the relationship labels: spouse-to-head, child-to-child-in-law, etc. But matters are complicated considerably by differences in category availability, terminological slippage across samples, and data inconsistencies. For example, in some samples the "sibling" category may have included large numbers of siblings-in-law; or "spouse" might sometimes mean the wife of any household member rather than exclusively the wife of the head.[6]

---

[6] The samples for Iraq and India, for example, have large numbers of multi-spouse households, not because of polygamy, but because many daughters-in-law had a misreported relationship to the household head.

We required a method to systematically uncover these irregularities. Accordingly, for every sample we calculated the number of additional couples that would be created if we allowed any given pair of relationships to link. This "matchmaker" program produced a list of possible pairings in each sample that warranted further examination: those that involved a non-trivial proportion of the total married population and whose allowance would substantially reduce the spouse-absent rate of one of the involved relationship categories. Our analysis led to refinement of the basic list of acceptable pairings and to sample-specific customizations, such as allowing child-to-child links in samples where "child" apparently includes children-in-law.

The matchmaker method also exposed complications related to the reporting of marital status and cohabitation across samples. Close inspection revealed that obvious couples commonly gave different "in union" responses: for example a household head said he was legally married but his spouse reported being in a consensual union. This phenomenon was so widespread it proved necessary to globally allow mismatched statuses as long as both persons reported some kind of union and had appropriate relationship information. In selected instances relationship information can even override marital status: spouses can link to heads even if only one of them claims to be in a union, and unmarried partners can link to heads regardless of their marital statuses. We also uncovered widespread inconsistency among census respondents in the reporting of consensual partners of heads and family members as relatives or non-relatives. By definition, non-relatives should never be linked to relatives; but where consensual unions are concerned, that fundamental divide cannot be maintained. Consequently, after all other passes have been made "other relatives" and non-relatives can link to heads or any other family member, as long as both parties report being in consensual unions and are not already linked.

Polygamy poses a technical complication for the spouse identifier. Where polygamy was indicated, multiple females can link to one man; but he in turn can link to only the most proximate spouse, because the spouse pointer variable can only record a single person number. In samples in which only men are identified as being in a polygamous union, multiple women can link to a polygamous man as long as the women are in a marital union of some kind. Finally, some samples do not identify polygamous unions, although polygamy was widely practiced. We allow multiple female spouses to link to heads in those censuses. Polygamous marriages not involving the head and spouse cannot be identified, but they are much less common.[7]

Limited information on cohabitation in some samples poses the most serious comparability issue for the spouse pointer (see Appendix 1 for a summary of the availability of cohabitation data by country). Out of 115 samples, nine identify unmarried partners of household heads only in the relationship variable. Partners of persons other than the household head therefore cannot be identified; however, since these samples are exclusively from developed countries with relatively simple household structures, the great majority of consensual unions are undoubtedly recorded. More troubling are the 14 samples from censuses whose questionnaires specified only legal unions were to be reported. In some of these societies, consensual unions were probably rare, and in others it is possible that substantial numbers of de facto marriages were reported regardless of what the census instructions may have stipulated. Analysis of European countries that changed the legal-status requirement between censuses suggests the instructions had little impact on the overall distribution of responses; but there is no way to measure the effect or to be sure about countries that lack the data for such comparisons.

Location of Mother and Father -- MOMLOC/POPLOC

[7] Only 10 percent of polygamous marriages involve persons other than the head or spouse.

Links between children and parents occur after the creation of spousal links. Unlike the spouse pointer, there is no variable comparable to marital status that consistently indicates a person's eligibility to receive a parent link.[8] Consequently, all persons are considered eligible to receive parent links—i.e., to be "children"—with one exception: persons over age 15 unrelated or with an unspecified relationship to the household head.[9] All adults are eligible to be parents, although fertility plays an important role in evaluating parent-child pairings. Fertility data cannot be determinative for two reasons: first, because our parent pointers are designed to identify both biological and social parents; and because over one-quarter of IPUMSI samples contain no fertility data, while others limit this information to married or reproductive age women.

Like the spouse pointer, the parental linking algorithm works sequentially downwards through a household. Each person is evaluated in turn as a potential "child," and the program searches the household for a probable mother or father, based on relative ages, relationship to head, parent's marital status, mother's fertility, and proximity. The specific criteria used to evaluate a possible match depend on the child's relationship to the household head and fall under five broad rules. Appendix 3 describes the allowable pairings in each rule.

Most links are unambiguous—like a link between the household head and a child of household head—and 94% of all parent links fall under Rule 1, the strongest rule. Other links are less certain, such as links between children and grandchildren, or between nonrelatives of the head. As links become weaker, the criteria for matching become more stringent. For instance, adjacency is required for links involving nonrelatives of the head while additional age and

[8] For instance, data on mortality status of mother and father are available for only 16 samples. Further, these variables are also not comparable to marital status, because children with a deceased parent may live with a stepparent, while some children with living parents may live apart from their parents. These variables are used in the construction of variables indicating likely stepparent relationships.

[9] Empirical evidence from samples with census pointers indicated that among persons 16 and older, only about 1 percent of nonrelatives and 5 percent of relatives with an unspecified relationship to the head lived with a parent. We concluded that given the low numbers of matches that should be made and the ambiguity of these relationship categories, we could not successfully construct pointers for these individuals.

fertility requirements are implemented when linking children to polygamous spouses. Although the algorithm searches for both fathers and mothers simultaneously, within a given strength test links to potential mothers are evaluated before links to potential fathers. As soon as a link is made to a mother or father, a second link is automatically generated to that person's spouse or partner, and no additional attempts are made to find parents for that individual.

Once a link is made several variables are automatically generated. The first is a rule variable (PARRULE), which describes the specific conditions under which the parent pointers were produced. We also produce stepmother and stepfather variables to identify links that are definitely or probably not biological: including links to explicitly-identified adopted and stepchildren, links in excess of a woman's known fertility, and links that fall outside reproductive age ranges. Using the STEPMOM variable, researchers interested in fertility can select only those mother-child links which probably reflect biological relationships. It should be noted that there are many adopted and stepparents who cannot be identified with information available in the censuses; therefore, the IPUMS stepparent indicators will always under-represent their actual number in the population.

The specific rules implemented in IPUMS-international are loosely based on the algorithm developed for the United States (Ruggles 1995), but the logic had to be substantially revised for the international database. We used two approaches for evaluating changes to the algorithm. Whenever possible we relied on empirical evidence drawn from IPUMS samples. For example, we used information on the relative position of children and children-in-law in samples that distinguish these categories to develop a rule for samples that report only a combined child/child-in-law category. Thus only the first listed child/child-in-law receives a parental link to the household head and spouse, the second is presumed to be the in-law.[10] We also used information from the small number of samples that collected data on parent's location in the household as a question on the census form. If a particular specification produced links with a high rate of disagreement with the census pointers (described more below) we rejected the rule. Our final algorithm disallows links to never-married non-cohabiting men as potential fathers except when the relationships were unambiguous, because these links were invariably wrong.

When the above methods were not possible, we developed a system to systematically evaluate modifications to the linking algorithm. Each time we altered the program, we selected a random sample of about 500 households in which MOMLOC or POPLOC changed. These households were divided among several analysts who examined each household by hand, comparing the new pointers to the previous version and scoring each change as improved, worsened, or indeterminate. A change was accepted only if all analysts agreed that the modification resulted in a noticeable improvement.

A primary concern guiding the development of the pointers was to prevent all children in complex households from linking to a single parent when there were multiple legitimate candidates.[11] This is particularly salient where the ordering of the persons within households makes proximity an ineffective linking criterion, for instance when all grandchildren are grouped together instead directly following their parents. To address the problem, we rely heavily on reported children ever born and children surviving to determine how many children should link to a particular woman and, by extension, to her spouse or partner. We refer to this as the "child cap" for a parent or couple. In some contexts, the linking algorithm allows the cap to be

[10] Over 80 percent of married children are listed before their spouse in samples that distinguish children from children-in-law.

[11] Roughly 98 percent of persons under age 18 had at most one person who qualified as a possible mother. In 12 samples, however, more than 5 percent of children had 2 or more potential mothers, including 3 samples exceeding 10 percent of children.

exceeded, but only after other potential parents have received their share of eligible children. Thus, the child cap plays a powerful role in the allocation of children.

Unfortunately, some censuses do not collect women's childbearing data and virtually no countries collect data for men. In these instances where we could not use empirical data to "cap" links to a potential parent, we needed some way to apportion children among potential parents. To do this, we calculate a child cap which our algorithm uses in place of known fertility. The caps are based on the five rules for linking children to parents. Children are allocated among parents in proportion to the total number of children eligible to link to each parent under a particular rule.[12] In addition, the caps are designed to increase the probability that we link to ever-married compared to never-married women,[13] in recognition of the higher fertility of married women.[14]

These caps play a critical role in determining whether a child should link to a particular parent, except when relationships are completely unambiguous. When a child links to a parent, the caps of the linked parent and their non-polygamous spouse/partner are reduced. Once a potential parent's cap is filled, we search for alternative parents with an available cap. In

---

[12] To calculate the cap, we first count a woman's "potential children", or the total number of children who meet the basic relationship and age requirements for a mother-child match. A 3-year old grandchild qualifies as a potential child of a 47-year old female child, but a 2-year old grandchild does not. Next, we calculate each woman's share of children as the ratio of her potential children to the sum of children who could potentially link under a particular rule. For instance if we calculated a child cap under Rule 2 (child-grandchild matches), we would divide each "parent's" potential matches by the sum of all potential matches between adult children and grandchildren. This proportion is then multiplied by the total number of children available to be matched under the rule (e.g. the total number of grandchildren in the household). Caps for men are calculated separately, but follow the same logic.

[13] Calculations for ever-married women exclude the potential children of never-married women when calculating the denominator. In essence, we divide all available children between the ever-married potential mothers in a household, ensuring that children will be more likely to link to the married women. Calculations for never-married women include the potential children of married potential mothers; as a result, never-married women have a reduced, but non-zero, probability of linking to children compared to co-resident ever-married women.

[14] The decision to privilege links to married persons, but to allow links to never-married individuals is supported by evidence from samples with fertility data. Take, for instance, households containing grandchildren as well as both ever-married and never-married adult female children of the head. Births were reported by ever-married children only in 75% of these households, by both ever-married and never-married children in 17% of households, and by never-married daughters only in 2%. (In the remaining 3% of households, no births were reported by children of the head, suggesting the mother of any grandchildren was absent.)

households with a small number of children to be linked and many potential parents, the estimated cap will tend to divide the children evenly among all potential parents (for instance, every parent links to one child), even when the household order suggests an uneven distribution is more accurate. We concluded that this was preferable to a situation in which unreasonably large numbers of children would link to just one mother.

In order to assess the performance of the calculated child caps, we set up an experiment. For the 76 samples containing fertility data for both married and unmarried women, we produced two sets of pointers: one using calculated caps and one using known fertility. For 98.5 percent of children under age 18, the two sets of maternal pointers agreed completely. Most disagreements occurred when children received links to two different mothers. The overall proportion of children who received a maternal link increased slightly, from 87.7 to 88.3 percent, when calculated caps were used in place of known fertility. The pointers constructed using calculated caps appear to overestimate teenage childbearing relative to fertility-based pointers (12.6 and 12.2% respectively) as well as births to never-married mothers (5.4% and 5.0%). Differences about twice as large are found in African samples, the region with the greatest household complexity. Thus, although pointers constructed without fertility data appear to overestimate early and nonmarital fertility, the differences are small in magnitude and predictable in direction.

Research with Census Pointers

It is difficult to assess the quality of the constructed family links without having some basis for comparison. Fortunately, a number of international censuses directly asked respondents for the line number on the census form of their mother, father or spouse.[15] These links were used

[15] The links for 2001 Spain were constructed by the census office using family names and other information not contained in the sample microdata.

for guidance during the development of the IPUMS pointers, and they provide a means to evaluate the final product.

Agreement between IPUMS and Census Pointers

Our analysis includes 13 IPUMS samples that contain census variables indicating the location of a spouse or parent. Although the samples over-represent Europe, they are nevertheless diverse, including both developed and developing countries, and having temporal depth. Thus they can provide information on the likely range of agreement between IPUMS pointers and unobserved family interrelationships in the full IPUMS database.

The rate of disagreement between the IPUMS pointers and the corresponding pointers from the censuses is presented in Table 3.[16] Overall, the spouse pointers agree 99.5% of the time, and the parental pointers more than 98.7%. The denominator for the mother and father statistics is all persons, because even adults are at risk of residing with parents. If one considers parental links only to persons under age 18, the rate of disagreement increases roughly doubles. Still, the absolute level of agreement is very strong, at over 97%.

[Table 3 about here]

The rate of disagreement varies across samples due to a variety of factors. The reporting order of persons within households often conveys significant information about family relationships, and the IPUMS linking algorithm is designed to be sensitive to that information. But some samples are less well ordered than others because of differing enumeration practices or post-enumeration data processing. Samples also vary in their rate of data errors in substantive variables. The linking process, which compares information from multiple records, will

[16] The analysis includes 500,000 persons per sample, so each is effectively weighted equally. The South African samples collected data on biological parents only; the comparative linking statistics therefore exclude step and adoptive children in those datasets.

moreover tend to illuminate data inconsistencies that are not evident in person-level tabulations. The category detail in the key variables also differs, producing more ambiguous situations for the pointer variable code to navigate in some samples. Finally, the census linking variables are recorded as numbers referring to other lines on the census form. Numeric data collected in this manner are especially error prone, and close examination of the data suggests these variables noisier than typical categorical variables.[17]

The linking success rate is also affected by the underlying social reality reflected in the data. Some situations and living arrangements are inherently more difficult for the pointer program to manage. Basically, the more complex the household structure, the more chance there is to make an error. At the sample level, the correlation between the discrepancy rate and the proportion of persons living in extended households is .89 for spouse links, and .86 and .83 for mother and father links (for persons under age 18). The relationship is still strong but somewhat weaker between mean household size and discrepancy rates, suggesting that household complexity is the salient issue. The samples with census pointers have smaller households on average than the full IPUMS database: 4.71 persons versus 5.39 persons per household. They also have fewer persons living in extended families: 29.8% compared to 33.5%. It is therefore likely that the constructed pointer variables for IPUMS as a whole are somewhat less accurate than the average rates suggested by Table 3.

A majority of mismatches between IPUMS and the census pointers involve situations where the census did not identify a parent or spouse, yet IPUMS linked to someone who met the necessary criteria. Such errors of commission are to some degree unavoidable. If there is any

[17] The census for 1998 Kyrgyzstan reported the location of spouse and parent, but we dropped the sample from the analysis because of the high rates of obvious data errors, including many links to nonexistent person numbers. The 2001 Armenian sample attempted to construct location of father by combining information on location of mother and of her spouse, but many married fathers were not successfully linked. A number of samples do not consistently link every child-of-head to the head/spouse; data errors that are excluded from the statistics reported in Table 4.

plausible parent or spouse in the household, the IPUMS program will link to them. For spouses, there is usually no way of knowing the correct partner is absent. For mothers and fathers there is sometimes supporting evidence on fertility history or parental mortality that suggests the biological parent is absent. But because the IPUMS pointers are intended to encompass social parentage--step and adopted children--the linking tends to be generous, even to the point of exceeding the known number of children a woman has borne. The constructed variables identifying probable step mothers and fathers allow researchers to exclude social parents from analyses requiring strictly biological links.

On the one hand, absent spouses and parents pose the most difficult situation for accurate linking. On the other hand, the lack of a spouse or parent sometimes indicates an error in the original census pointer data. Non-responses are indistinguishable in the microdata from an absent parent or spouse: both typically receive a code of zero in the data. In general, we would expect the non-response rates to be higher in less developed countries, but there is surprising variation in quality even among the developed countries. In any case, to the extent that there are missing data in the census pointers, the error rates suggested by Table 3 will be exaggerated.

Globally, less than 2 percent of persons live in a situation where there is more than one potential mother, father or spouse to whom they could conceivably link. Apart from the issue of absent persons, these complex situations pose the greatest challenge for the linking program; and in some African and Asian countries they can be several times more common than the world average. In such households, how frequently do IPUMS and the census point to different persons? Where there is a choice to make, IPUMS points to a different spouse 11% and a different father 15% of the time.[18] Mothers have a 26% discrepancy rate, driven substantially by

[18] To reduce the confounding effect posed by errors in the source data, these statistics exclude cases where either IPUMS or the census points to no one.

South Africa, where over one-third of the links are different. The mean rate for the other 10 samples is 14% for mothers. The error rate for South Africa may be indicative of factors that are likely to obtain elsewhere in Africa, but it could be an idiosyncrasy of the data collection practices in one country. In any case, the conclusion is that the pointers produce relatively high error rates in a small subset of households.

The census pointers can also reveal the specific relationship categories that pose the most difficulty for the linking program. For the spouse links, the greatest number of errors involves children linking to the wrong child-in-law. The error rate is only about 3 percent, but these are large categories. It's not uncommon for there to be more than one possible child to whom a child-in-law might link, or for the spouse in such situations to be absent. The parental linking errors are dominated by grandchildren linking to children, with the great majority stemming from the South Africa samples. In South African households there are often strings of children followed by strings of grandchildren, and it is difficult to accurately assign people to the correct mother.

There are at least two additional factors that the census pointer data do not help us address. Polygamy is legal in many African and Muslim countries, and polygamous households are especially challenging for determining family interrelationships. But among the samples with census pointer variables, only South Africa identifies polygamy, and it has too few such cases to generalize. The samples with census pointers also do not let us measure the effect of de facto versus de jure census practices. The de facto censuses enumerated persons where they happened to be at the moment of the census, while the de jure censuses recorded people at their usual place of residence. The de facto censuses should have higher rates of absent spouses and parents, but there is insufficient diversity in the samples with census pointers to explore the issue.

Agreement on the characteristics of linked persons: IPUMS and Census pointers

The previous analysis suggests that individual level disagreements between the IPUMS and census pointers occur rarely, except in complex households. These disagreements, however, may balance out in the aggregate, if the IPUMS links are representative overall of actual family relationships. For instance, although we may link a grandchild to the wrong parent, it is likely that the true parent, a sibling, will be similar in age, education, or even marital status. In the analysis that follows, we examine how well our spouse and parent links reproduce the characteristics of spouse and parent-child pairings.

Table 4 presents data on age differences between spouses identified by the census pointers and by IPUMS pointers. Recall that the agreement between these pointers is quite high. Not surprisingly, the two distributions of age-differences are virtually identical. A comparable analysis for individual samples yields similar results: in total, only three IPUMS estimates differ from the census estimates by 1 percent, and there is no clear country or age-pattern of discrepancies.

[Table 4 about here]

Census and IPUMS estimates of the age difference between children and mothers are shown in Figure 2. These estimates include likely step or adoptive relationships, but can be interpreted roughly as mother's age at birth. Across samples, we find that the two sets of estimates are extremely similar. The IPUMS pointers produce higher estimates of early motherhood: age differences less than 20 years (11.5% compared to 11.1% using census pointers) while ages-differences between 20 and 34 years are slightly underestimated. Nonetheless, the magnitude of disagreement is substantially less than the differences in actual pointer-links (shown earlier in Table 3). Thus, although the IPUMS pointers appear to overestimate the proportion of children living with a mother, the distribution of mother-child ages is affected only slightly.

[Figure 2 about here]

Overestimation of young motherhood, relative to census pointers, occurs in all samples. The magnitude is quite small, except in South Africa where the Census and IPUMS pointers differ by about 1 percentage point. Motherhood at ages 20-34 is consistently lower in the IPUMS estimates than in census estimates, with the largest deviations observed in South Africa, while no pattern emerges at ages 35 and older.

Figure 3 presents estimates of the proportion of children ages 0-17 who live in a two-parent, one-parent, or no-parent family. This analysis is restricted to the 9 samples with both maternal and paternal census pointers. By both methods, we estimate that two-thirds of children live with two parents (receive both a mother and a father pointer), although the IPUMS estimate is marginally higher. The IPUMS pointers appears to overestimate residence with a lone mother (21% compared to 19% using census pointers), and to slightly underestimate children with no parents (11% and 13%). In additional analyses, we found that IPUMS estimates of the proportion of children with a never-married mother exceed census estimates by less than 1 percentage point (7.7% compared to 7.0%) in samples with maternal pointers. South Africa, with a high rate of parent absence and complex household structure, is largely responsible for the overall disparity rates: differences between IPUMS and census estimates in that country are about 2-3 times larger than average. Note that most of the samples with census pointers also collect data on mother's fertility, so these differences result primarily from IPUMS links to single or never-married women who have borne children.

[Figure 3 about here]

Despite their limitations, the census pointers offer the best available evidence on the strengths and limitations of the IPUMS pointers. Most of the factors that complicate accurate linking are correlated at the country level. The less developed countries tend to have larger and more complex households, less consistent enumeration practices, and more data errors during data collection and processing. On the positive side of the ledger, most developing countries with large, complex households have fertility and often parental mortality data, which help significantly in making the parental links.

Discussion

This paper describes the development of family interrelationship variables for the IPUMS-International project, facilitating comparative analysis of families and households in 42 countries. The value of the IPUMS family variables lies not only in the size of the database, but also in the comparability of the pointer variables across samples. Despite large differences in census enumeration practices and in family and household structure, the IPUMS-International pointers provide a consistent set of tools for researchers wishing to identify family relationships. Pointers are difficult for individual researchers to construct, and they would invariably do so in different ways. By using consistent pointers available to everyone, researchers can replicate each other's results and be certain they are measuring the same phenomena. Moreover, by identifying family relationships, the IPUMS pointers allow researchers to easily create their own measures of family and household composition or measures of family-level characteristics.

The primary goal of this paper is to provide population researchers with guidance in using the IPUMS family interrelationship variables. The building blocks of these variables are the location of spouse and location of parent variables. Using information on relationship to head, age, sex, marital status, fertility, and relative position in the household, links are created between persons and their likely spouses or likely parents. We document the methodology for producing links, revising earlier international and U.S. projects to address the scale and diversity of the IPUMS project. The linking algorithm is the same for all samples, although samples vary in the availability of variables and in variable detail.

The census pointers, direct reports of spouse and parent location collected during enumeration, provide valuable information on the strengths and limitations of the IPUMS pointers. Across samples with empirical census pointers, the IPUMS and census pointers are in close agreement: 99.5% for the spouse pointer and 98.7% for parent pointers, although disagreement rates are higher for individual countries. The agreement rate falls notably when there are multiple spouse or parent candidates, but fewer than one-in-fifty persons face such a choice. The characteristics of spouse and parent-child pairings produced by the IPUMS pointers resemble closely those of census pairings. Estimates of the age-differences between spouses are virtually indistinguishable in all samples, while estimates of early motherhood are only slightly higher in the IPUMS samples. The IPUMS estimates of residence with a lone mother or with an unmarried motherhood are also higher, on average, than census pointer estimates. Additional analyses indicate pointers constructed in samples without fertility data will slightly overestimate early and unmarried motherhood. Even in the most challenging situations, limited data or complex household structure, the IPUMS pointers perform well.

The major factors that affect our ability to correctly identify spouse and parent-child pairs include: the meaningfulness of household order; the availability of detailed relationship categories, such as grandchildren or in-laws; the availability of childbearing data; and the size and complexity of households structures. Differences in these factors will influence the comparability of the pointers across countries and within countries over time. In general, we expect the impact to be quite small, but researchers must evaluate the likely effect for their individual projects. For example, in samples without fertility data, overall estimates of nonmarital fertility are likely to track closely the true population levels, but a non-trivial minority of the children who receive links to unmarried mothers may do so erroneously. To assist researchers, the IPUMS project documents the differences among samples in the available raw materials for the construction of these links. This allows researchers to make informed decisions when their object of study might be especially susceptible to particular limitations in the underlying data.

There is already considerable demand among population researchers for this data. Since the pointers were released in June 2008, 40 percent of IPUMS data extracts (2359 of 5924) have included one or more of the pointers or variables derived from them. That is roughly equal to the number of extracts that requested the variable “children ever born.” Studies using the IPUMS pointers include research on intergenerational co-residence, changes in family size and children’s resources, measurement of single-parent families, and the study of children and international migration (Ruggles and Heggeness 2008; Lam and Marteleto 2008; Bryant 2008; Heggeness 2009). The IPUMS-International database is expected to double in size over the next five years. The family interrelationship pointers are a key component of the IPUMS, and the value of these consistently derived measures will only increase as the data series expands.

Appendix 1. IPUMS samples with family interrelationships variables

Country	Census years	Key variables for pointer construction[a]		
		Cohabitation data[b]	Fertility data[c]	Polygamy[d]
Argentina	1970, 1980, 1991, 2001	**	**	
Armenia	2001	**	**	
Austria	1971, 1981, 1991, 2001	**	*	
Belarus	1999	**	**	
Bolivia	1976, 1992, 2001	*	**	
Brazil	1960, 1970, 1980, 1991, 2000	**	**	
Cambodia	1998		**	
Chile	1970, 1982, 1992, 2002	**	**	
China	1982, 1990		*	
Colombia	1973, 1985, 1993, 2005	**	**	
Costa Rica	1973, 1984, 2000	**	**	
Ecuador	1974, 1982, 1990, 2001	**	**	
Egypt	1996			Yes
France	1962, 1968, 1975, 1982, 1990, 1999	*		
Ghana	2000	**	**	Yes
Greece	1971, 1981, 1991, 2001	*	*	
Guinea	1983, 1996	*	*	Yes
Hungary	1970, 1980, 1990, 2001	**	**	
India	1983, 1987, 1993, 1999			
Iraq	1997		*	Yes
Israel	1972, 1983, 1995		*	
Italy	2001	*		
Jordan	2004			Yes
Kenya	1989, 1999		**	Yes
Kyrgyz Republic	1999	**	**	
Malaysia	1970, 1980, 1991, 2000		*	
Mexico	1970, 1990, 1995, 2000	**	*	
Mongolia	1989, 2000	*	*	
Palestine	1997		*	Yes
Panama	1960, 1970, 1980, 1990, 2000	**	*	
Philippines	1990, 1995, 2000	**	*	
Portugal	1981, 1991, 2001	*		
Romania	1977, 1992, 2002	*	**	
Rwanda	1991, 2002	**	**	Yes
Slovenia	2002	*	**	
South Africa	1996, 2001, 2007	**	*	Yes
Spain	1991, 2001	*	*	
Uganda	1991, 2002		*	Yes
United Kingdom	1991	*		
United States	1960, 1970, 1980, 1990, 2000, 2005	*	*	
Venezuela	1971, 1981, 1990, 2001	**	*	
Vietnam	1989, 1999	**	*	

Note: *=data are incomplete or available in only some years, ** = complete data available in all years, blank = no data available.
[a]For additional details on data limitations, see https://international.ipums.org/international/parrule_table.shtml.
[b]Consensual unions are explicitly identified, although in some samples only for the head and his/her unmarried partner. Note that many samples without explicit cohabitation data treated “de facto” marriages as marriages in marital status and relationship to head variables. Consequently, we are able to identify people in informal unions in most samples, even if the type of union is unknown.
[c]Fertility data are available in at least some years; some samples may collect data for only married or reproductive age women.
[d]In polygamous samples, the household head is allowed to link to multiple wives.

Appendix 2. Rules for SPLOC construction

Rule	Individual's relationship to head	Partner's relationship to head	Age difference	MARST match	Require adjacency	Notes
Rule 1: Strong relationship pairing, couple adjacent						
	Head	Spouse, unmarried partner	No	No	Yes	1,2
	Parent	Parent	No	No	Yes	1
	Parent-in-law	Parent-in-law	No	No	Yes	1
	Child	Child-in-law	No	No	Yes	1
	Sibling	Sibling-in-law	No	No	Yes	1
	Aunt	Uncle	No	No	Yes	1
	Grandparent	Grandparent	No	No	Yes	1
	Nephew/niece	Nephew/niece-in-law	No	No	Yes	1
	Cousin	Cousin-in-law	No	No	Yes	1
Rule 2: Strong relationship pairings, couple not adjacent						
	See Rule 1	See Rule 1	No	No	No	3
Rule 3: Weak relationship pairing, couple adjacent						
	Other relative	Other relative	Yes	Prefer	Yes	1,4
	Non-relative	Non-relative	Yes	Prefer	Yes	1,4
	Child	Other relative, other relative/non-relative	Yes	Prefer	Yes	1,4
	Sibling	Other relative	Yes	Prefer	Yes	1,4
	Grandchild	Grandchild	Yes	Prefer	Yes	1,4
	Grandchild	Other relative	Yes	Prefer	Yes	1,4
	Nephew/niece	Nephew/niece	Yes	Prefer	Yes	1,4
	Cousin	Cousin	Yes	Prefer	Yes	1,4
	Unknown	Head, child, unknown	Yes	Prefer	Yes	1,4
Rule 4: Weak relationship pairing, couple not adjacent						
	See Rule 3	See Rule 3	Yes	Prefer	No	3,4
Rule 5: Weak consensual union pairings						
	Head	Non-relative	Yes	Exact	Preferred	5
	Any relative	Non-relative	Yes	Exact	Preferred	5
	Head	Spouse	Yes?	No	Preferred	6
Rule 6: Sample-specific pairings						
	Child	Child	Yes	Exact	Preferred	7

Notes: Persons must be at least age 12 and married or in-union to be eligible. See note 6 for exception.

1. A woman will link to an adjacent preceding man before linking to an adjacent subsequent man.

2. In polygamous unions (where the man is identified as polygamous), each female spouse identifies the husband's person number in SPLOC. The husband in polygamous unions points to the first female spouse that links to him.

3. A woman will link to the most proximate preceding man before linking to the most proximate subsequent man.

4. A woman can be no more than 20 years older or 35 years younger than a potential partner. Couples with exact marital status matches will link before couples with mismatched marital status.

5. Both partners must be in a consensual union. See note 4 for age differences.

6. The head or spouse is out-of-union while the other is cohabiting. See note 4 for age differences.

7. Marriages between two children are allowed in a small number of samples, typically those that do not include a category for child-in-law. Exact marital status matches are required and this rule is implemented last. . See note 4 for age differences.

Appendix 3. Rules for MOMLOC/POPLOC construction

Rule	Child's relationship to head	Parent's relationship to head	Age difference	Fertility limits	Require adjacency	Notes
Rule 1: Links involving Head and Spouse						
	Child	Head, spouse, unmarried partner	10-69	no	no	
	Child	Spouse/partner of polygamous head	10-54	weak	no	1
	Child/child-in-law	Head, spouse, unmarried partner	10-69	no	no	2
	Child/grandchild	Head, spouse, unmarried partner	10-44	no	no	3
	Head, sibling	Parent, parent/parent-in-law, parent/grandparent	10-69	no	no	
	Spouse, sibling-in-law	Parent-in-law	10-69	no	no	
	Sibling/sibling-in-law	All parent categories	10-69	no	no	
Rule 2: Links between grandchildren and children						
	Grandchild	Child, child/child-in-law	15-44	yes	no	4
Rule 3: Links between other specified relatives						
	Nephew/niece	Sibling, sibling/sibling-in-law	15-44	weak	no	4
	Nephew-in-law/niece-in-law	Sibling-in-law, sibling/sibling-in-law	15-44	weak	no	4
	Grandchild, great-grandchild	Grandchild	15-44	weak	no	4
	Cousin	Aunt/uncle	15-44	weak	no	4
Rule 4: Links involving other unspecified relatives and other relatives/non-relatives						
	Head	Other relative	>=20	strict	no	5, 6
	Other relative, other rel/non-rel	Child	15-44	strict	no	5, 7
	Other relative, other rel/non-rel	Unmarried partner	15-44	strict	no	5
	Other relative, other rel/non-rel	Other relative	15-44	strict	no	5
	Other relative, other rel/non-rel	Grandchild	15-44	strict	no	5
	Other relative, other rel/non-rel	Sibling, sibling-in-law	15-44	strict	no	5
	Other relative, other rel/non-rel	Other relative/non-relative	15-44	strict	no	5
Rule 5: Links between people unrelated to the head						
	Any non-relative age 0-15	Unmarried partner	15-44	strict	yes	5
	Any non-relative age 0-15	All other non-relatives	15-44	strict	yes	5

1. When the household head is polygamous, we narrow the allowable age difference between a potential mother and child and give priority to women who have not exceeded their child cap. Children who do not link to any mother, are linked to the head and to his first spouse.

2. When two children/children-in-law are linked by SPLOC, the first listed receives parent link

3. Applies only to France 1962-1975

4. In samples with childbearing data, a potential mother must be ever-married or in a consensual union or have ever given birth. A potential father must be ever-married or in a consensual union.

5. Number of links cannot exceed a woman's observed number of children-ever-born or constructed child cap. A potential father must be ever-married or in a consensual union, and number of links is limited based on spouse/partner's childbearing history.

6. Allowed only in samples without a parent relationship code

7. Allowed only in samples without a grandchild relationship code

References

Bryant, John. 2008. “Children and International Migration.” Pp. 177-194 in *Situation Report on International Migration in East and South-East Asia*. Bangkok: International Organization for Migration, Regional Office for Southeast Asia.

Cho, Lee-Jay, Robert D. Retherford, and Minja Kim Choe. 1986. *The own-children method of fertility estimation*. East-West Center, East-West Population Institute.

Grabill, Wilson H., and Lee-Jay Cho. 1965. “Methodology for the measurement of current fertility from population data on young children.” *Demography* 2:50–73.

Hacker, J. D. 2003. “Rethinking the "early" decline of marital fertility in the United States.” *Demography* 40:605-620.

Hall, Patricia Kelly, Robert McCaa, and Gunnar Thorvaldsen. 2000. *Handbook of international historical microdata for population research*. Minnesota Population Center.

Heggeness, Misty. 2009. “(Mis)Measuring Lone-Mother Families.” in *Annual Meeting of the Population Association of America*. Detroit.

Lam, David, and Leticia Marteleto. 2008. “Stages of the Demographic Transition from a Child’s Perspective: Family Size, Cohort Size, and Children’s Resources.” *Population and Development Review* 34:225-252.

Levin, Michael J., and Robert D. Retherford. 1982. “The effect of alternative matching procedures on fertility estimates based on the own-children method..” Pp. 11–17 in *Asian and Pacific Census Forum*, vol. 8.

Lichter, Daniel T., Zhenchao Qian, and Martha L. Crowley. 2008. “Poverty and economic polarization among America’s minority and immigrant children.” Pp. 118-143 in *Handbook of families and poverty: Interdisciplinary perspectives*, edited by D. Russell Crane and Tim B. Heaton. New York: Sage.

Luther, Norman Y., and Lee-Jay Cho. 1988. “Reconstruction of birth histories from census and household survey data.” *Population Studies* 42:451–472.

McGarry, Kathleen, and Robert F. Schoeni. 2000. “Social security, economic growth, and the rise in elderly widows' independence in the twentieth century.” *Demography* 37:221–236.

Minnesota Population Center. 2009. *Integrated Public Use Microdata Series — International: Version 5.0*. Minneapolis: University of Minnesota.

Moehling, Carolyn M. 2004. “Family structure, school attendance, and child labor in the American South in 1900 and 1910.” *Explorations in Economic History* 41:73–100.

Moehling, Carolyn M. 2007. "The American Welfare System and Family Structure: An Historical Perspective." *Journal of Human Resources* 42:117-155.

Retherford, Robert D., and Lee-Jay Cho. 1978. "Age-parity-specific birth rates and birth probabilities from census or survey data on own children." *Population Studies* 32:567–581.

Retherford, Robert D., Lee-Jay Cho, and Nam-Il Kim. 1984. "Census-derived estimates of fertility by duration since first marriage in the Republic of Korea." *Demography* 21:537–558.

Retherford, Robert D., Minja Kim Choe, Jiajian Chen, Li Xiru, and Cui Hongyan. 2005. "How far has fertility in China really declined?." *Population and Development Review* 31:57–84.

Ruggles, Steven. 1995. "Family Interrelationships." *Historical Methods* 28:52-58.

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, and Misty Heggeness. 2008. "Intergenerational coresidence in developing countries." *Population and Development Review* 34:253-281.

Ruggles, Steven et al. 2009. *Integrated Public Use Microdata Series: Version 4.0 [Machine-readable database]*. Minneapolis, MN: Minnesota Population Center [producer and distributor].

Short, Susan E., Frances K. Goldscheider, and Berna M. Torr. 2006. "Less help for mother: the decline in coresidential female support for the mothers of young children, 1880-2000." *Demography* 43:617–629.

Strong, Michael A. et al. 1989. *User's Guide: Public Use Sample, 1910 United States Census of Population.* Ann Arbor: University of Michigan, Inter University Consortium for Political and Social Research.

Zuberi, Tukufu, and Amson Sibanda. 1999. *Fertility Differentials in sub-Saharan Africa: Applying Own-Children Methods to African Censuses*. Philadelphia: University of Pennsylvania.

Table 1. Example of census household

Person Number	Relationship	Age	Sex	Marital status	Children ever born
1	Head	73	Male	Married	n/a
2	Spouse	62	Female	Married	6
3	Child	38	Female	Single	1
4	Child	30	Female	Cohabiting	0
5	Child	32	Male	Married	n/a
6	Child-in-Law	30	Female	Married	1
7	Grandchild	6	Male	Single	n/a
8	Employee	16	Female	Cohabiting	Unknown

Table 2. Example of census household with constructed pointers

Person number	**Relationship**	**Age**	**Sex**	**Marital status**	**Children ever born**	**SPLOC**	**MOMLOC**	**POPLOC**
1	Head	73	Male	Married	n/a	2	0	0
2	Spouse	62	Female	Married	6	1	0	0
3	Child	38	Female	Single	1	0	2	1
4	Child	30	Female	Cohabiting	0	0	2	1
5	Child	32	Male	Married	n/a	6	2	1
6	Child-in-Law	30	Female	Married	1	5	0	0
7	Grandchild	6	Male	Single	n/a	0	6	5
8	Employee	16	Female	Cohabiting	Unknown	0	0	0

Table 3. Disagreement between IPUMS and Census Pointers (%)

		All persons		Age < 18	
Census	Spouse	Mother	Father	Mother	Father
Armenia 2001	1.29	1.09		2.62	
Belarus 1999	0.16	0.28		0.43	
Brazil 1991		0.46		1.30	
Portugal 1981	0.32	1.11	0.39	1.06	0.46
Portugal 1991	0.15	1.92	0.63	1.27	0.61
Portugal 2001	0.23	0.61	0.31	1.39	0.91
Romania 1977	0.36	0.43	0.21	0.62	0.44
Romania 1992	0.54	0.36	0.29	1.09	0.93
Romania 2002	0.11	0.20	0.17	0.69	0.63
South Africa 2001	1.21	4.87	1.88	9.96	4.08
South Africa 2007	0.83	3.89	1.28	8.88	2.92
Spain 1991	0.10				
Spain 2001	0.20	0.30	0.23	0.55	0.51
TOTAL	0.46	1.29	0.60	2.49	1.28

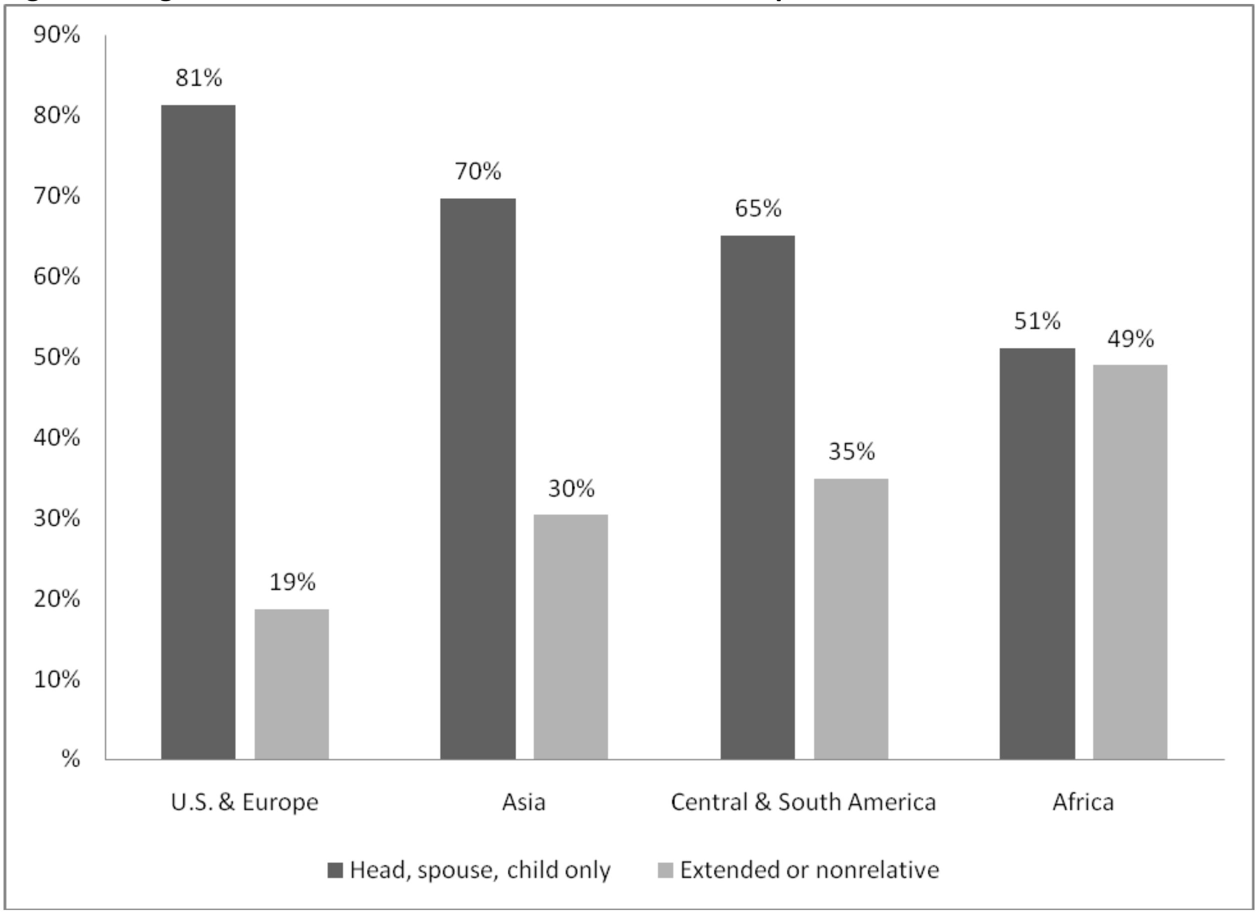
Samples are weighted equally.

The denominator for the spouse column is persons in a union.

Table 4. Census and IPUMS pointer estimates of the age difference between spouses

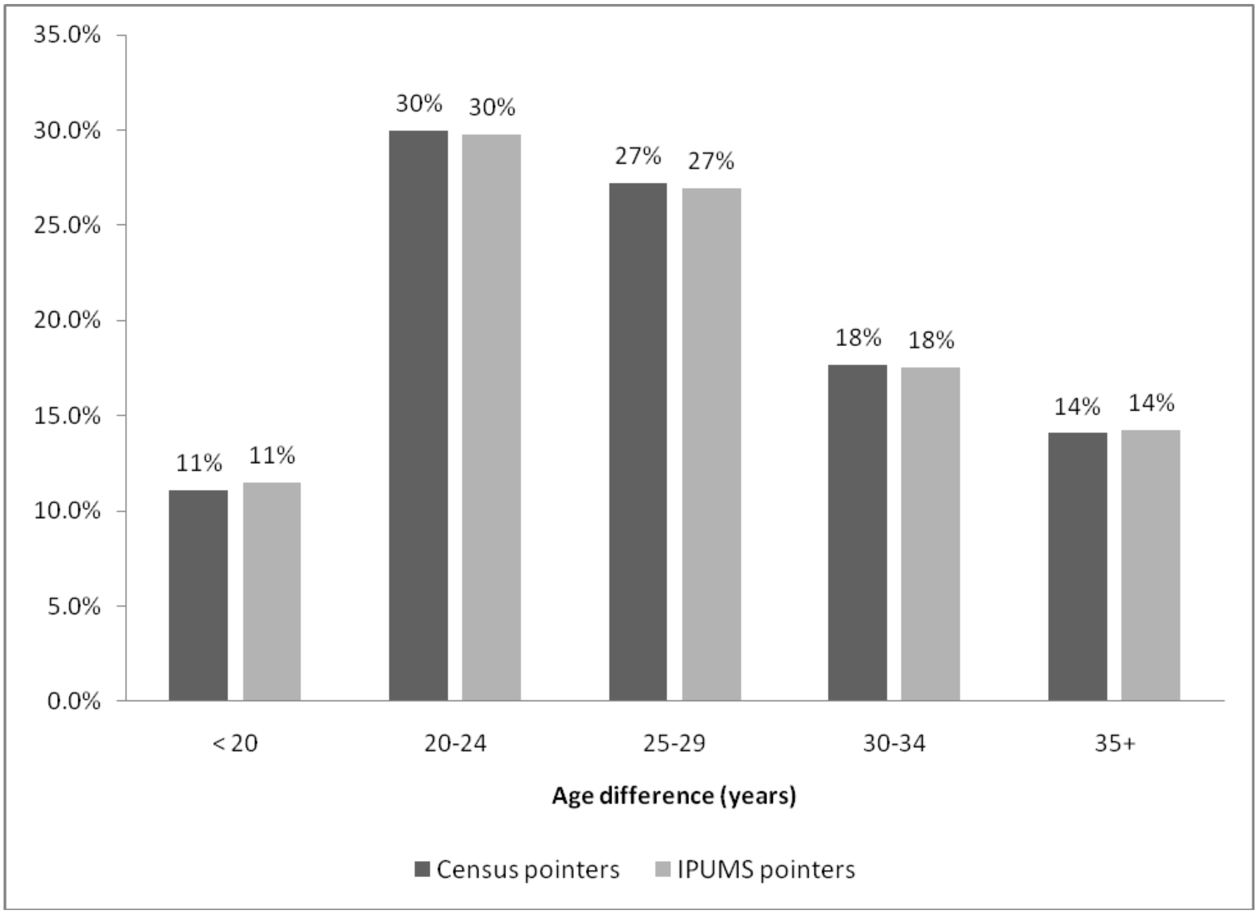
Age difference	Census	IPUMS	Absolute Difference	IPUMS/ Census
Wife older by				
2+ yrs	9.6%	9.6%	0.0%	0.998
Similar ages (0-1 yr)	24.0%	24.0%	0.0%	0.999
Husband older by				
2-4 yrs	32.1%	32.1%	0.0%	1.001
5-9 yrs	26.4%	26.4%	0.0%	1.001
10+ yrs	7.9%	7.9%	0.0%	1.001

Figure 1. Regional differences in children's household composition



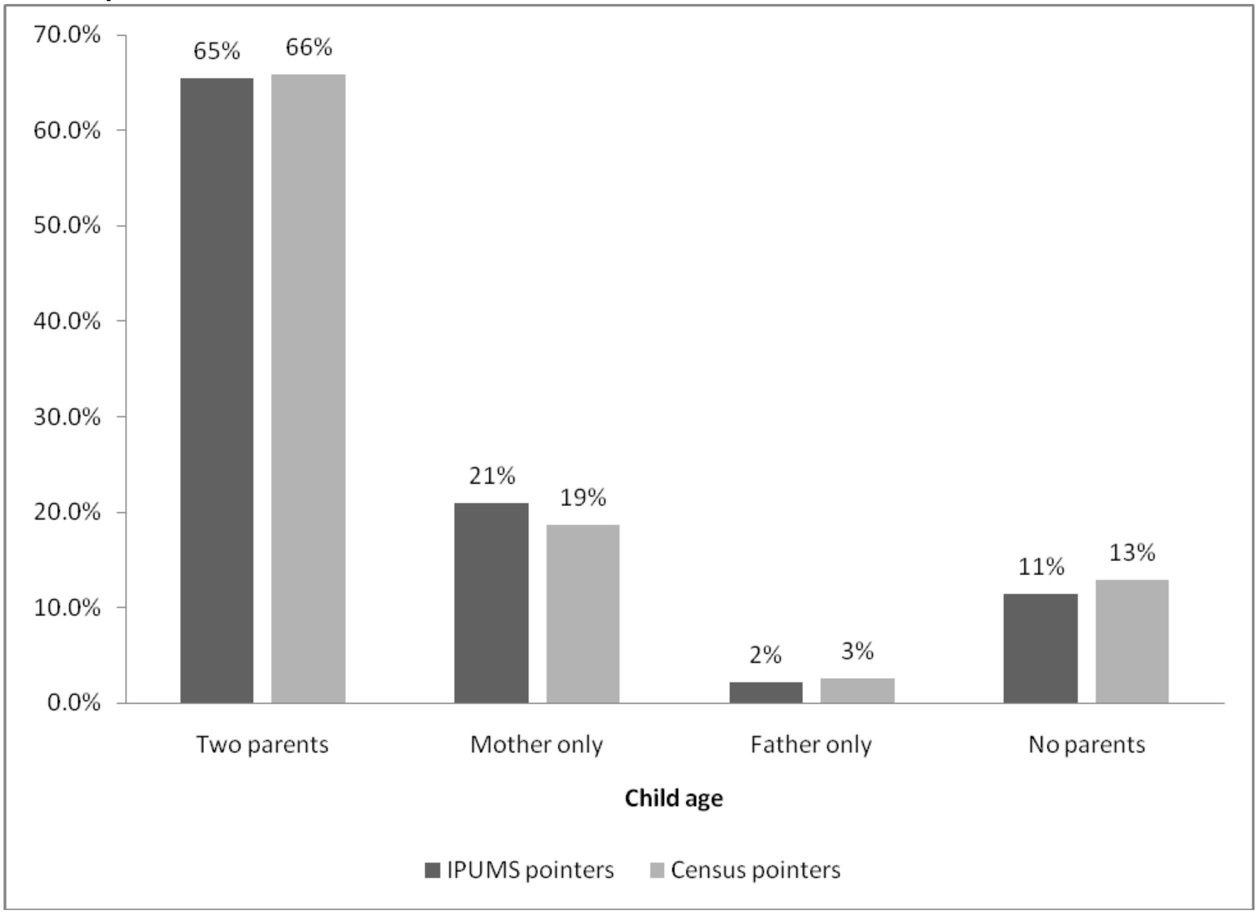
Notes: Estimates are reported for children ages 0-17. Polygamous households are included with "extended families".

Figure 2. Census and IPUMS estimates of the age difference between mothers and children



Notes: Estimates are reported for children ages 0-17 in samples with maternal census pointers.

Figure 3. Census and IPUMS estimates of children living in a two-parent, one-parent, and no-parent families



Notes: Estimates of children's family structure are reported for children ages 0-17 in samples with census pointers for both mothers and fathers.