The quality of constructed family and household relationships in African Census Samples

DRAFT: do not cite without permission of the authors

Sheela Kennedy
Robert McCaa
Matthew Sobek
Lara Cleveland

Minnesota Population Center

Paper prepared for the Sixth African Population Conference, Ouagadougou, Burkina Faso, 5-9 December 2011. The authors wish to acknowledge the statistical offices that provided the underlying data making this research possible: Central Agency for Public Mobilization and Statistics, Egypt; Ghana Statistical Services, Ghana; National Statistics Directorate, Guinea; National Bureau of Statistics, Kenya; National Statistical Office, Malawi; National Directorate of Statistics and Informatics, Mali; National Institute of Statistics, Rwanda; National Agency of Statistics and Demography, Senegal; Statistics Sierra Leone, Sierra Leone; Statistics South Africa, South Africa; Central Bureau of Statistics, Sudan; Bureau of Statistics, Tanzania; and Bureau of Statistics, Uganda.

In African countries, census data provide critical information on current and historical trends in households and family relationships. We use data from IPUMS-International and the African Integrated Census Microdata Series (AICMD), a freely available database of 52 million person records from 13 African countries, from the 1980s through the 2000s. Our paper assesses the quality of the data available in each of these censuses for constructing measures of spouse and parent-child relationship, household structure, and estimates of fertility. We consider the quality of age and sex reporting, as well as missing data. We assess the quality of the links created within African censuses and compare estimates of own child fertility in South Africa with other published estimates. We show that the IPUMS pointers perform well and are especially valuable given the complex family and household structure found in many African countries.

Introduction

Censuses are among the most widely used sources in population research. In countries with incomplete vital registration, censuses provide critical information on mortality and fertility (Hall, McCaa, and Thorvaldsen 2000; Preston, Heuveline, and Guillot 2000). In many countries, censuses are the earliest available data on populations, predating household surveys, and provide large sample sizes and coverage of men and women of all ages. Census microdata files typically have a hierarchical structure of individual persons nested within households. Although this structure can be used to identify family relationships, the process is inherently difficult. In most censuses, each person's relationship to a reference person in their household is known. Relationships between other household members are often ambiguous. To adequately determine family interrelationships analysts must use a number of variables in combination, including age, household relationships, fertility, marital status, and even proximity in the household roster. Creating consistent family interrelationship measures is especially important for comparative and historical research.

The IPUMS-International project--the world's largest collection of publicly available census microdata--has developed a series of family interrelationship variables that address this need.[1] These variables include "pointer" or locator variables which identify each person's spouse, mother, and father, if they reside in the household. The International Integrated Public Use Microdata Series consists of 397 million person records in 185 census samples from 62 countries (Minnesota Population Center 2011). Family relationship variables have been developed for 167 of these samples.[2] The freely available African Integrated Census Microdata Series (AICMD), maintained by IPUMS-International, provides census data for 13 African countries, and family relationship variables are available in all 25 census samples. Since the development of these family relationship variables in 2008, 40 percent of IPUMS data extracts have included one of the pointers or variables derived from them.

This paper has three goals. The most basic goal is to inform African population researchers about the availability of the IPUMS/AICMD database and the constructed family relationship variables. The second aim is to assess the quality of the family interrelationship variables constructed for African censuses. The quality of these pointers depend upon the complexity of households, the quality of age and sex reporting, and the amount of missing data for key variables. We conclude by demonstrating the usefulness of these variables in constructing estimates of own-child fertility in South Africa.

Background

IPUMS-International and African Integrated Census Microdata Series Projects

The African Integrated Census Microdata Series (AICMD) is a partnership between the Minnesota Population Center (MPC) and the African Centre for Statistics of the Economic

---

[1] The IPUMS-International data series and the African Integrated Census Microdata series are continually growing and evolving (McCaa, Esteve-Palos, Ruggles and Sobek 2006). The discussion in this paper pertains to the database and its constructed variables as of fall 2011 (Minnesota Population Center 2011).

[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.

Commission for Africa. The MPC founded the IPUMS-International project with the goal of preserving, harmonizing and disseminating international census microdata and documentation (Hall et al. 2000). The 2011 version of IPUMS-International, includes data from 62 countries, including 13 in Africa. Twenty-nine African countries—encompassing 700 million people, four-fifths of the continent with extant data—have entrusted microdata to IPUMS-International, ensuring continued expansion of the database. The complete database is available at https://international.ipums.org/, and can be accessed through the AICMD portal: http://ecastats.uneca.org/aicmd.

The database is designed to facilitate comparative research. Variables are harmonized so all samples use consistent but detailed codes. Comparability issues that cannot be adequately conveyed through variable labeling and coding schemes are described in the integrated documentation. Data are available at no charge through a web-based data extraction system. Researchers select the samples and variables they wish to include in their extract and then download the pooled microdata extract (i.e., all the data are in a single file) for analysis.

Individuals are organized into households in all 25 African samples, and family interrelationship variables have been created for these samples. The full list of these samples is shown in Table 1. The 2001 and 2007 South African samples included a question on the census enumeration form identifying the location (the line number) of each person's spouse and parents. These variables allow us to directly evaluate the IPUMS constructed family pointers.

[Table 1 about here]

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.[3] 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. 1984; 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, Retherford, and Choe 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 have since been developed, but have not been implemented widely (Zuberi and Sibanda 1999; Strong et al. 1989).

As part of the 1995 IPUMS-USA release of integrated microdata files for eleven U.S. censuses, family interrelationship variables were reconceived as a multi-purpose tool and made available to researchers in public use samples (Ruggles 1995; Ruggles et al. 2009). 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

[3] Examples of own-child matching programs are included in Cho, Retherford, and Choe (1986) and software is available online from the East-West Center (The East-West Center Research Program 1992).

descriptors. The linking algorithm had to be flexible enough to deal with differing variable availability and changing category detail across census years. 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; Hacker 2003).

Family Interrelationship variables

Pointers, or locator variables, are variables that identify each person's mother, father, or spouse, if one is present in the household. These variables are the basis of family interrelationship variables in IPUMS. Table 2 displays the information used to construct pointers for an 8-person household. Most important is the relationship-to-household-head variable. From this, we know the head and spouse are married, and that they are the parents of the three children. For the remaining household members, additional demographic variables as well as proximity in the household must be used to infer relationships. The married male child in position 5 is almost certainly married to the female child-in-law directly following him, because there are no other children-in-law to whom he could be married and because of their relative positions in the household. The grandchild, listed directly below this couple, is most likely their son. He could, however, be the son of the female child in position 3 (a single mother of one child).

[Table 2 about here]

IPUMS identifies these relationships for researchers using the same program for all samples. The pointer variables "point" to each individual's mother, father and spouse, using their person numbers. The variable SPLOC reports the person number of each person's spouse or partner. In this example, the head and spouse receive SPLOC values 2 and 1 respectively. The variables MOMLOC and POPLOC record the person number of each individual's parents. The grandchild in position 7 points to his mother in position 6 and his father in position 5. The pointer variables are given the value zero when no spouse or no parents are identified.

Since consistent rules are applied across samples, households with similar characteristics in different samples will receive the same distribution of constructed pointers. Moreover, the pointer variables will be identical for every researcher who selects them for download from the IPUMS database. 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.

Once SPLOC, MOMLOC, and POPLOC are created, a number of additional family relationship variables are constructed, including the calculation of the number of children who are linked to particular woman, the age of her youngest child, the total number of families in a household, and the household characteristics. Each pointer variable is accompanied by a "rule" variable describing the criteria used to assign the spouse or parent link.

Users of the IPUMS database have access to these family interrelationship variables and can use them to identify whether a person lives with a spouse or parents or is raising own children. Using these variables, researchers can then construct their own measures of kin characteristics, family and household composition, fertility and marriage patterns, using the standard statistical packages. An important feature of the IPUMS web-based data extract system allows researchers to attach the characteristics of parents and spouses as new variables on each person's record.

The pointers constructed for IPUMS-International are loosely based on the IPUMS-USA pointers, but have been modified substantially to reflect the greater international variation in household and family characteristics. Documentation for each family relationship variable is available as part of the extract system. See Sobek and Kennedy (2009) for a detailed description of the IPUMS-I pointer construction rules.

The household presented in Tables 2 is of moderate size, provides detailed information on relationship to household head, and requires only one decision—a straightforward choice between the grandchild’s two possible mothers. Correctly identifying family relationships becomes more difficult when the relationship pairings are more ambiguous, when parental absence occurs commonly, or when there are multiple potential spouses and parents. The quality of the resulting pointers depends upon the complexity and quality of the input data. Thus, the primary goal of this paper is to examine how well the IPUMS pointers perform across African countries.

Methods

Our evaluation of the pointers constructed for the African censuses in the IPUMS-International database is primarily descriptive. We begin by comparing the IPUMS pointers constructed for the South African 2001 and 2007 censuses to the pointer data collected on the enumeration forms. Although census pointers are affected by reporting and imputation errors, we believe they provide a valuable tool for assessing the accuracy of the IPUMS pointers (Moultrie and Dorrington 2004). This analysis also highlights the types of families where errors are most likely to occur.

Given these findings, we then consider how often these situations are likely to occur. We consider the complexity of households and families and the availability of detailed relationship categories. We examine the degree of missing data for variables that are most important for accurate pointers construction. We systematically test the importance of fertility data for the construction of pointers by constructing an alternate set of pointers that ignore information on children ever born or surviving, in samples where these variables are available. Finally, we calculate standard measures of the quality of age and sex reporting. Note, some censuses undergo data editing prior to public release; thus, low levels of missing data may reflect data editing and imputation procedures rather than better reporting or data collection procedures.

We conclude our analysis by examining the result of own-child estimates of age-specific and total fertility rates for South Africa. We use the East-West Center's EASWESPOP fertility software and compare the results produced using the IPUMS pointers, the South African census pointers, and EASWESPOP’s MATCHTAB pointers (The East-West Center Research Program 1992). We also compare the estimates to other published estimates for South Africa which use different estimation techniques (Moultrie and Timæus 2003; Moultrie and Dorrington 2004); .

Results

South Africa: comparison between Census Pointers and IPUMS Constructed Pointers

Two African censuses, South Africa 2001 and 2007, directly recorded the line number of each person's spouse, biological mother, and biological father, if they were present in the household.[4] The degree of agreement between these census pointers and the IPUMS constructed pointers provides an estimate of the accuracy of the IPUMS pointers. Because the census pointers were supposed to identify only biological relationships, we exclude IPUMS pointers to step and adoptive children in our analysis.[5] Census pointers are subject to reporting errors and non-response, and disagreements between IPUMS pointers and census pointers could occur even if the IPUMS pointers make the correct links.

Table 3 demonstrates that the two sets of pointers agree in most instances. Inconsistencies can arise if there is disagreement between the two sets of pointers as to whether a person's spouse or parents are in the household, or if the two sets of pointers link to different spouses or parents. In both censuses, 99 percent of spousal links are in agreement. Disagreements are more common for the parent pointers but still unusual: 4 percent of links to mothers and less than 2 percent of links to fathers. If we restrict our analysis to children under 15, the percent disagreeing increases to 11 percent and 5 percent respectively. Note, in about 15 percent of these disagreements, the original MOMLOC pointer (which included likely stepmother relationships) matched the census mother pointers. This suggests that in some instances non-biological relationships are included in

[4] The 2001 census was a de facto census and enumerated only persons present on census night. The 2007 Community Survey enumerated all permanent household residents, even if they were not present on census night, as well as all persons temporarily present that night.

[5] Specifically, we exclude all links between children and fathers, when the child is explicitly identified as a step or adopted child of that parent, when the child's mother or father is dead, or when the age difference between the parent and child is implausible. We also exclude links between mothers and children when the mother reports no children ever born or surviving, or having no child present in the household.

the census pointers[6], or that there were reporting errors in age, mortality of parent, or mother's fertility histories.

[Table 3 about here]

In Table 4 we detail pointer disagreement by relationship to household head. In most instances, links among heads, spouses, and children are highly accurate. Links between the head and spouse are virtually always accurate, and differences may reflect errors in reported relationship, sex, or age.[7] Spouse pointer disagreement is highest among children-in-law, at 6 percent. For all other categories, the IPUMS and census spouse pointers agree at least 97 percent of the time.

For parent pointers, several relationship categories stand out as having high disagreement rates among children under age 15 (over 15 percent for mother pointers): adopted children, grandchildren, unspecified other relatives, and nonrelatives. In each instance, there is ambiguity in the child's relationship to potential parents, and considerable chance of error. For biological children, just 3 percent of links are made in error and likely reflect the difficulty in determining whether a child of the household head is the biological or stepchild of his or her spouse.

[Table 4 about here]

In South Africa, polygamous marriages are uncommon and few husbands reside with multiple spouses. Typically, polygamous marriages can only be identified for the male household head. In polygamous marriages involving the head, however, we find that the error rate for links between children and mothers is over 40 percent. Thus, polygamy introduces considerable challenges in correctly linking children with the correct mothers.

[6] Statistics South Africa commonly imputed Mother Person Number in the 2001 census: for over 15% of infants and between 10-15% of children under age 5(Moultrie and Dorrington 2004).

[7] For instance, children-in-law are sometimes identified as a "spouse" in the relationship to household head field, or the head's age might be recorded as less than 12 years and will not be linked to a spouse under IPUMS rules.

Overall, however, we find close agreement between the constructed IPUMS pointer and the census pointers. Because the agreement rate for spouse pointers is extremely high, in the sections that follow we will focus primarily on the factors which are likely to affect the accuracy of parent-child pointers.

Complexity of family and household relationships

As shown above, when relationships are largely unambiguous, as with children of the household head, errors are uncommon. Nearly all other links involve some ambiguity. A child and a grandchild of a household head may co-reside in the same household, but not be parent and child. In many samples, grandchildren are commonly grouped into an "other relative" category, adding greater uncertainty to these links. Typically very little is known about non-relatives, greatly increasing the likely errors that are made.

Children's living arrangements in Africa are more complex than observed in other regions of the world (See Figure 1). Only about half of children in Africa reside in a household with no other persons than the head, at most one spouse, and other children of the household head. In contrast, over 80 percent of children in the U.S. and Europe, 70 percent of children in Asia, and 65 percent of children in Latin America reside in equally simple households. African household complexity introduces greater uncertainty in family relationships; *however, this same uncertainty makes the IPUMS pointers especially valuable for African population researchers.*

[Figure 1 about here]

Within Africa, the percent of children in complex families varies widely. Table 5 presents information on the relationship between children and the household head. About 95 percent of children under age 15 are children of the household head in Egypt, compared to less than 60 percent in South Africa. Most samples fall in the middle of this range, between 70 and 90 percent. Linking children in South Africa is more complicated than elsewhere in most African samples, and we expect pointer accuracy in Africa as a whole to exceed the levels observed for the 2001 and 2007 South African samples.

[Table 5 about here]

Many children of the head, however, live in households with two or more possible mothers: when their fathers are in polygamous marriages. Polygamous marriages occur in all of the samples included in AICMD; however, they are particularly common in Guinea, Mali, and Senegal, where between one-quarter and one-third of all children under age 15 are the child of a polygamous head. In most samples, however, at least 60 percent of children under the age of 15 have a largely unambiguous relationship with their parents.

The detail of relationship categories for constructing pointers also varies considerably across samples (see Table 6).[8] All samples identify children of the household head. However, in most African samples, the category "child" included some children-in-law. In these samples, IPUMS allows marriages (spousal links) between two children of the head. Only about half of the samples explicitly identify grandchildren instead of categorizing them as other relatives, adding additional uncertainty in constructing parent pointers. In samples without a grandchild code, parent pointers were allowed to link other relatives with children and children-in-law (if identified).

[Table 6 about here]

Children ever born and children surviving

When a child could link to more than one potential parent, the IPUMS pointers rely on women's fertility histories (children ever born and children surviving) to identify the most likely

[8] In Senegal 1988, detailed information on relationship to household head is available only for members of the primary family. Persons in subfamilies (for instance married children of the head, or mother and siblings of the head) are identified only as "other relative" or "other relative or non-relative".

mother. In samples with fertility data, children only link to women reporting no children born or surviving when the relationship is unambiguous (e.g. head, spouse, or child) or the woman is ever-married or cohabiting and there isn't a better match to a woman with surviving children. In such cases, they are flagged as likely children. When fertility data are unavailable, we must rely more heavily on position in the household and age-differences to identify the most likely parents for each child.

Links in samples with fertility data will be more accurate than those without them, but the actual impact is small. In the Appendix, we describe the results of an experiment, where we constructed pointers for samples with children ever born or surviving, but ignored this information. In African samples with fertility data, we found that just 3 percent of maternal pointers changed as a result of ignoring reported fertility. On average, however, links were more likely to occur at the youngest or oldest ages and to mothers with no children when fertility was unavailable. Although the impact is relatively small, caution should be taken when assessing small differences between censuses with and without explicit fertility data.

Of the 25 IPUMS-Africa samples, 21 collected explicit data on children ever born or children surviving (see Table 6). Table 7 presents estimates of the percent of women ages 15 to 64 with no information on children ever born in censuses that requested fertility data. The most common source of missing data occurs when the fertility questions were asked only of married women or women of reproductive ages. Data can also be missing as a result of non-response among eligible women.

[Table 7 about here]

Quality of age, sex, and marital status data

Age, sex, and marital status are key variables used in the construction of the pointers. Reporting is nearly complete in all samples (see Table 7); thus, we do not expect missing data to pose a problem for pointer construction.[9]

We also calculated three standard measures of the quality of age and/or sex reports: the Whipple Index, Myer's blended index, and the U.N. Age-Sex Accuracy Index. The Whipple index measures digit preference for 0's and 5's, and values range from 100 (no concentration) to 500 (only 0's and 5's reported). Myer's index examines concentration on any terminal digits. Values range from 0 (no heaping) to 90 (all ages reported with same terminal digit). The U.N. age-sex accuracy index is a summary measure of the quality of age and sex reporting, based on changes in sex ratios by age. Results are presented in Table 8 for all samples. Age heaping is evident in most of the samples, although the quality of age and sex reporting generally improves in the more recent samples.

[Table 8 about here]

Own-child fertility estimates: comparison of IPUMS and South African census pointers

Figure 2 plots own-child estimates of the Total Fertility Rate (TFR) produced using the 2001 South African census from 1987 to 2001. Point estimates for other studies using census and survey data are also plotted. Because of underreporting of infants is common in censuses, we focus on TFR values beginning one year before the census.

[Figure 2 about here]

Using the IPUMS pointers, we estimate that the TFR in South Africa declined from 4.0 in 1987 to 2.6 in 2000. These results are indistinguishable from those produced using the South African census pointers. In addition, they are generally consistent with TFR calculated using different methods and/or data sets. In 2001, Moultrie and Dorrington (2004) estimate that the

---

[9] In some instances, this may reflect data editing which occurs after census enumeration.

TFR had fallen to 2.84 nationally, quite close to our own-child estimates. Moultrie and Timæus (2003), using the 1996 Census estimate that the TFR was 3.23 in that year; our own-child estimate is again quite close—3.3 for 1996 from the 2001 census.

As Levin and Retherford (1982) and our own analysis have demonstrated, inferred pointers tend to link too frequently at the extremes--to younger or older mothers. Because the TFR is the sum of age-specific fertility rates, choosing the wrong mother is unlikely to have a large impact on TFR. Age-specific fertility rates, however, may be biased to extreme maternal ages. To examine this possibility, we calculated ages-specific fertility rates for 1996 using own-child methods and data from the 2001 census. Figure 3, present the results from this analysis and compares it to Moultrie and Timæus (2003) estimates. The IPUMS and census pointers yield extremely similar estimates of fertility at each age: at ages 15-19, age-specific fertility rate are 7 births per 1000 higher based on IPUMS pointers than census pointers; difference at all other ages are less than 3 per 1000. Compared to the Moultrie & Timæus (2003) age-specific rate estimates, the own-child fertility curve is shifted to the right, with lower age-specific fertility rates at young ages and higher age-specific fertility rates at older ages. These differences cancel out, yielding a similar estimate of TFR. Overall, we find very little difference in estimated age-specific and total fertility rates between the own-child estimates produced using IPUMS and census pointers, and own-child TFRs based on the 2001 South African census perform well compared to Moultrie and colleagues' estimates. However, age-specific fertility rates estimated using the own-child method are consistently shifted to older ages compared to the estimates of Moultrie and colleagues produced using questions on childbearing in the past year.

[Figure 3 about here]

Conclusion

This paper assessed the quality of household and family relationship variables in the African Integrated Census Microdata Series (AICMD) and IPUMS-International database. We show that there is close agreement between the IPUMS pointer and the spouse and parent line numbers reported during census and survey enumeration in South Africa in 2001 and 2007. We examine the detail and quality of important input variables: including household complexity, availability of fertility data, and standard indices of age and sex reporting errors. We concluded with a comparison of own-child estimates of fertility in South Africa produced using IPUMS and census pointers and with previously published estimates.

Our analysis demonstrates the success of the IPUMS constructed pointers in South Africa, a country with highly complex families. We anticipate that the IPUMS pointers will perform as well, if not better, in the other African samples. Given the complexity of many households in Africa, the availability of consistent AICMD/IPUMS family relationship variables for 25 African census samples from the 1980s through the present is a valuable resource for African Population researchers.

References

Cho, L. J, R. D Retherford, and M. K Choe. 1986. *The own-children method of fertility estimation*. East-West Center, East-West Population Institute.

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(4):605-620.

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

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(3):451–472.

McCaa, Robert , Albert Esteve-Palos, Steven Ruggles and Matthew Sobek. 2006. "Using integrated census microdata for evidence-based policy making: the IPUMS-International global initiative," *African Statistical Journal*, 2:83-100.

McGarry, Kathleen, and Robert F. Schoeni. 2000. "Social security, economic growth, and the rise in elderly widows' independence in the twentieth century." *Demography* 37(2):221–236.

Minnesota Population Center. 2011. "Integrated Public Use Microdata Series, International: Version 6.1." Retrieved November 23, 2011 (https://international.ipums.org/international/citation.shtml).

Moehling, Carolyn M. 2004. "Family structure, school attendance, and child labor in the American South in 1900 and 1910." *Explorations in Economic History* 41(1):73–100.

Moehling, Carolyn M. 2007. "The American Welfare System and Family Structure: An Historical Perspective." *Journal of Human Resources* 42(1):117-155.

Moultrie, Tom A., and Rob Dorrington. 2004. *Estimation of fertility from the 2001 South Africa Census data*. Cape Town: University of Cape Town.

Moultrie, Tom A., and Ian M. Timæus. 2003. "The South African fertility decline: Evidence from two censuses and a Demographic and Health Survey." *Population Studies* 57:265-283. Retrieved November 20, 2011.

Preston, Samuel, Patrick Heuveline, and Michel Guillot. 2000. *Demography: Measuring and Modeling Population Processes*. 1st ed. Wiley-Blackwell.

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(3):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(4):537–558.

Ruggles, Steven. 1995. "Family Interrelationships." *Historical Methods* 28(1):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(1):73-101.

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(4):617–629.

Sobek, Matthew, and Sheela Kennedy. 2009. *The Development of Family Interrelationship Variables for International Census Data*. Minneapolis, MN: University of Minnesota Retrieved November 20, 2011 (http://www.pop.umn.edu/sites/www.pop.umn.edu/files/Working%20Paper%202009-02.pdf).

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.

The East-West Center Research Program. 1992. *EASWESPOP: Fertility Estimate Programs, Version 2.0*. Honolulu, HI: East-West Center.

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. IPUMS/AICMD samples

Census		Sample %	Person records
1996	**Egypt**	10	5,902,243
2006		10	7,282,434
2000	**Ghana**	10	1,894,133
1983	**Guinea**	10	457,837
1996		10	729,071
1989	**Kenya**	5	1,074,098
1999		5	1,407,547
1987	**Malawi**	10	798,669
1988		10	991,393
2008		10	1,343,078
1987	**Mali**	10	785,384
1998		10	991,330
1991	**Rwanda**	10	742,918
2002		10	843,392
1988	**Senegal**	10	700,199
2002		10	994,562
2004	**Sierra Leone**	10	494,298
1996	**South Africa**	10	3,621,164
2001		10	3,725,655
2007		2	1,047,657
2008	**South Sudan***	7	542,784
2008	**Sudan***	17	5,066,511
1988	**Tanzania**	10	2,310,424
2002		10	3,732,735
1991	**Uganda**	10	1,548,460
2002		10	2,497,449

Source: http://ecastats.uneca.org/aicmd/en-us/samples.aspx

*Person records for South Sudan and Sudan computed from Sudan sample.

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 census pointer and IPUMS pointers in South Africa

		All persons		Child Age <15	
Census	Spouse*	Mother	Father	Mother	Father
2001	0.48%	4.69%	1.88%	10.74%	4.55%
2007	0.28%	4.83%	1.27%	11.84%	3.16%

*The denominator for spouses includes only person in a union.

Table 4. Household relationship and disagreement in percent between census pointers and IPUMS pointers in the South Africa 2001 census sample

		All persons		Age <15	
Relationship	Spouse*	Mother	Father	Mother	Father
Head	0.21	0.19	0.06		
Spouse/partner	0.40	0.16	0.05		
Biological child	0.20	2.76	0.64	2.98	0.65
Adopted child	0.35	18.77	3.11	20.92	4.34
Stepchild	0.14	8.36	1.30	8.83	1.58
Grandchild or great grandchild	0.12	20.76	7.98	23.11	8.88
Parent	2.24	0.00	0.00		
Parent-in-law	2.52	0.22	0.02		
Child-in-law	6.17	0.28	0.11	1.16	0.39
Sibling	0.30	1.14	0.50	2.14	1.35
Sibling-in-law	1.59	0.98	0.40	1.79	0.97
Grandparent	1.26	27.89	6.19		
Other relative, not elsewhere classified	1.23	10.68	2.99	21.59	6.07
Group quarters	1.31	0.16	0.07	1.18	0.49
Non-relative, n.e.c.	2.30	2.49	1.00	15.66	5.54

*The denominator for spouses includes only person in a union.

Table 5. Household relationships of children age 0-14 (%)

	Children of the household head			
Sample	**All**	**Non-polygamous father**	**Polygamous father**	**Ambiguous relationships**
Egypt 1996	94%	93%	1%	6%
Egypt 2006	97%	96%	1%	3%
Ghana 2000	61%	56%	5%	39%
Guinea 1983	77%	46%	30%	23%
Guinea 1996	71%	37%	33%	29%
Kenya 1989	79%	77%	3%	20%
Kenya 1999	76%	75%	1%	24%
Malawi 1987	74%	73%	2%	25%
Malawi 1998	78%	76%	2%	22%
Malawi 2008	81%	79%	2%	19%
Mali 1987	83%	54%	29%	16%
Mali 1998	86%	59%	27%	14%
Rwanda 1991	84%	84%	0%	15%
Rwanda 2002	79%	79%	0%	21%
Senegal 1988	62%	36%	25%	38%
Senegal 2002	60%	38%	22%	40%
Sierra Leone 2004	64%	64%	0%	36%
South Africa 1996	63%	62%	0%	36%
South Africa 2001	58%	58%	0%	41%
South Africa 2007	55%	55%	0%	45%
Sudan 2008	83%	79%	4%	17%
Tanzania 1988	72%	66%	6%	28%
Tanzania 2002	76%	74%	3%	23%
Uganda 1991	71%	66%	5%	29%
Uganda 2002	81%	79%	3%	18%
Total	77%	72%	4%	23%

Table 6. Availability of important variables used in pointer construction

Sample	Availability of Relationship to Household Categories			Fertility data
	Grandchildren identified	Parents or parents-in-law identified	Child combined with children-in-law*	
Egypt 1996	Yes	Neither identified	No	No
Egypt 2006	Yes	Parents only	No	No
Ghana 2000	Yes	Combined category	Yes	Yes
Guinea 1983	No	Neither identified	Yes	No
Guinea 1996	Yes	Combined category	Yes	Yes
Kenya 1989	No	Parents only	Yes	Yes
Kenya 1999	No	Parents only	Yes	Yes
Malawi 1987	No	Neither identified	Yes	Yes
Malawi 1998	No	Neither identified	Yes	Yes
Malawi 2008	No	Neither identified	Yes	Yes
Mali 1987	No	Parents only	No	Yes
Mali 1998	No	Parents only	No	Yes
Rwanda 1991	Yes	Parents only	Yes	Yes
Rwanda 2002	Yes	Parents only	Yes	Yes
Senegal 1988	No	Parents only	No	No
Senegal 2002	Yes	Parents only	Yes	Yes
Sierra Leone 2004	Yes	Combined category	Yes	Yes
South Africa 1996	Yes	Parents only	Yes	Yes
South Africa 2001	Yes	Both identified	No	Yes
South Africa 2007	Yes	Both identified	No	Yes
Sudan 2008	Yes	Combined category	No	Yes
Tanzania 1988	No	Neither identified	Yes	Yes
Tanzania 2002	Yes	Parents only	Yes	Yes
Uganda 1991	No	Parents only	Yes	Yes
Uganda 2002	No	Combined category	Yes	Yes

*Indicates that the category "child" of the head explicitly or implicitly includes children-in-law.

Table 7. Percent missing for key variables used in pointer construction

Sample	Children ever born*	Relationship	Age	Sex	Marital Status
Egypt 1996	N/A	0	0	0	0
Egypt 2006	N/A	0	0	0	0
Ghana 2000	1	0	0	0	0
Guinea 1983	N/A	0	0	0	0
Guinea 1996	9	0	0	0	4
Kenya 1989	18	0	0	0	1
Kenya 1999	0	0	0	0	0
Malawi 1987	2	0	0	0	2
Malawi 1998	0	0	0	0	0
Malawi 2008	0	0	0	0	0
Mali 1987	5	0	2	0	2
Mali 1998	4	0	0	0	1
Rwanda 1991	6	0	0	0	2
Rwanda 2002	9	1	0	0	5
Senegal 1988	N/A	1	0	0	1
Senegal 2002	21	0	0	0	0
Sierra Leone 2004	0	0	0	0	0
South Africa 1996	16	2	1	0	1
South Africa 2001	12	0	0	0	0
South Africa 2007	14	0	0	0	0
Sudan 2008	33	0	0	0	0
Tanzania 1988	5	0	0	0	0
Tanzania 2002	1	0	0	0	0
Uganda 1991	3	0	0	0	0
Uganda 2002	5	1	0	0	0

*N/A indicates that children ever born is not available for this sample.

Table 8. Indices of quality of age and sex reports

Sample	Whipple Index[10]	Myer's Index[11]	UN Age Sex Ratio[12]
Egypt 1996	212.3	17.7	36.4
Egypt 2006	196.3	15.7	21.8
Ghana 2000	183.8	15.3	32.2
Guinea 1983	217.0	21.1	65.8
Guinea 1996	207.0	19.3	56.4
Kenya 1989	147.9	7.8	20.5
Kenya 1999	149.4	7.5	23.6
Malawi 1987	138.7	7.1	47.0
Malawi 1998	147.9	10.9	28.1
Malawi 2008	120.5	5.3	30.7
Mali 1987	185.8	14.9	35.7
Mali 1998	180.9	15.5	36.5
Rwanda 1991	100.5	1.5	28.1
Rwanda 2002	106.9	2.4	27.3
Senegal 1988	101.4	9.8	39.9
Senegal 2002	185.3	15.3	35.5
Sierra Leone 2004	242.7	24.4	47.8
South Africa 1996	100.7	2.3	18.3
South Africa 2001	97.1	1.4	19.4
South Africa 2007	96.1	1.2	18.1
Sudan 2008	241.9	23.3	63.8
Tanzania 1988	189.7	16.9	46.3
Tanzania 2002	157.9	13.1	30.4
Uganda 1991	165.5	12.8	35.0
Uganda 2002	134.6	7.7	41.1

[10] Whipple index: Identifies heaping on 0's and 5's. Values range from 100 (no concentration) to 500 (only 0s and 5s reported). Can be categorized as highly accurate (<105), fairly accurate (105-109.9), approximate (110-124.9), rough (125-174.9) and very rough (>=175) data.

[11] Myer blended method: identifies concentration on any digits. Values range from 0 (no heaping) to 90 (all ages reported with same terminal digit).

[12] UN age-sex accuracy index: Summary measure of the quality of age and sex reporting, based on changes in sex ratios by age. Can be categorized as accurate (<20), inaccurate (20-40) and highly inaccurate (>40).

Figure 1. Children's household composition

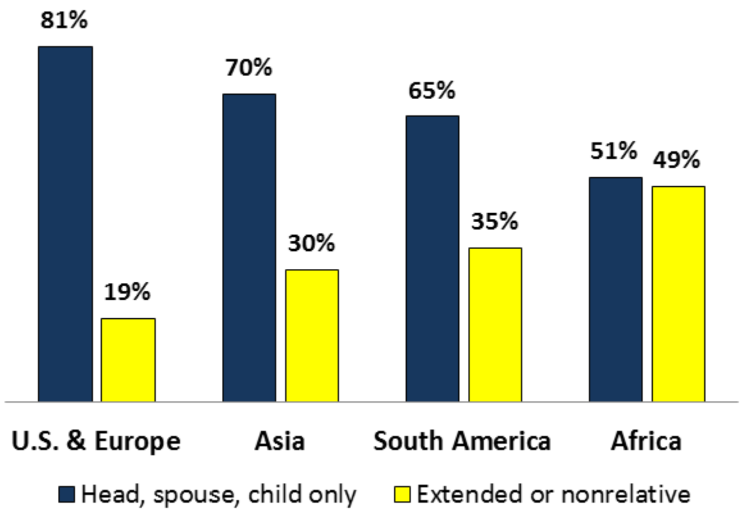
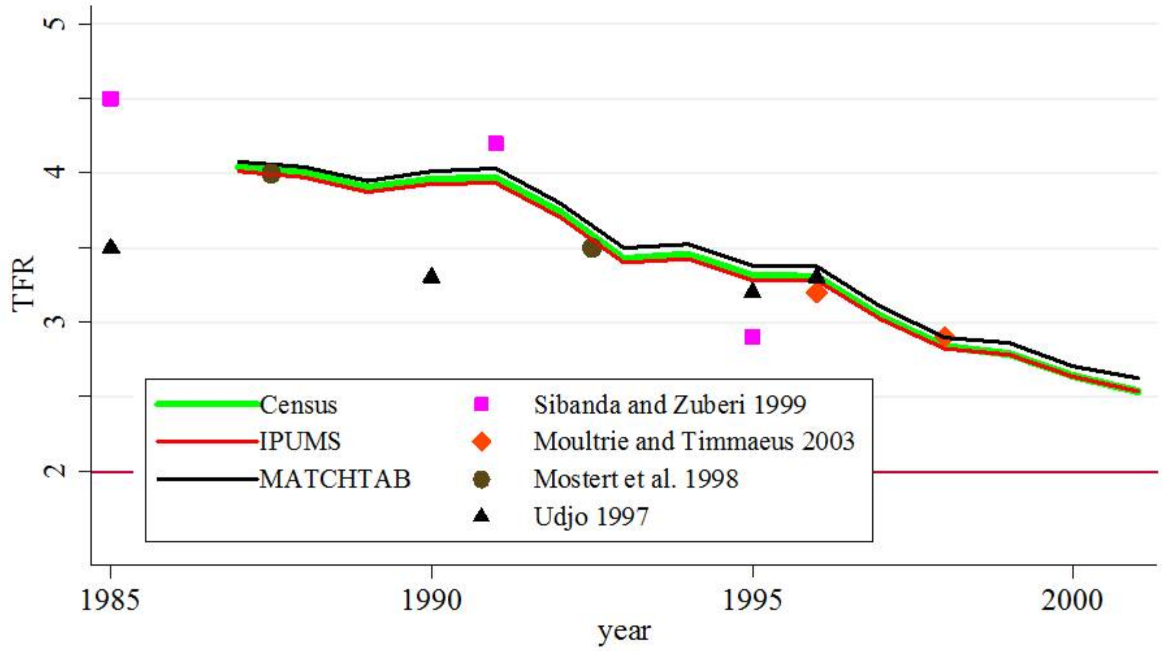


Fig. 2. Estimates of Total Fertility Rates: 2001 Census of South Africa

Computed with 3 Types of Pointers: Census, IPUMS, and MATCHTAB



Source: IPUMS-International: 10% sample of 2001 census of South Africa

All 3 Closely Approximate Authoritative Estimates

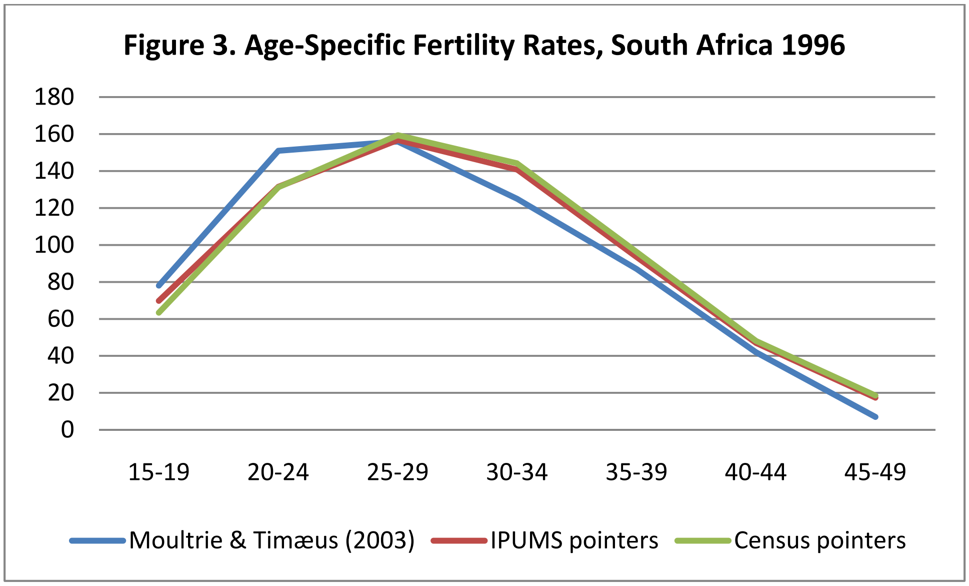


Figure 3. Age-Specific Fertility Rates, South Africa 1996
180
160
140
120
100
80
60
40
20
0
15-19
20-24
25-29
30-34
35-39
40-44
45-49
Moultrie & Timæus (2003)
IPUMS pointers
Census pointers

Appendix. Evaluating the child cap

We use fertility caps to distribute children among women when number of children surviving or children ever born is known. When fertility data are unavailable, we calculate a cap in order to prevent large numbers of children linking to any given individual.[13] A likely consequence of the cap calculation is that we will overestimate the proportion of children living with a mother, as well as the proportion of children born to a never-married mother. Take for instance a household with one never-married female child age 18 and one grandchild age 1. If we know that the child has no children ever born, our algorithm will not link the grandchild to the child. In the absence of fertility data, however, our algorithm links the grandchild and child via MOMLOC.

In order to assess the scale of this problem, we constructed new parent locators that ignored reported fertility for 76 samples that collected fertility data for never-married and ever-married women. The degree of similarity between the IPUMS pointers and the experimental pointers provides an estimate of the degree to which we are successfully able to reconstruct likely family relationships in the absence of fertility data. Likewise, differences between IPUMS and experimental estimates provide an estimate of the degree of uncertainty introduced when constructing pointers for samples without fertility. Although on average we expect links using fertility data to outperform links that ignore fertility, in some cases, either because of errors in fertility data, or because of step-parenthood or adoption, the opposite may be true.

Table A1 displays the proportion of individuals who experienced a change in MOMLOC or POPLOC as a result of our experiment. The value of MOMLOC changed for less than 1% of all persons and for 1.5% of children under age 18. About half of this change can be attributed to individuals who linked to different mothers in the IPUMS and experimental pointers. The remainder is primarily attributable to individuals receiving a mother for the first time under the experiment, and, less commonly, from persons losing a mother under the experimental algorithm. As a result, our estimate of the proportion of children under 18 who lived with a mother increased by 0.5 percentage points when fertility is ignored. Regional differences in the proportion of children experiencing changes in parent pointers are also shown in Table A1. Three percent of children in African samples received a new MOMLOC, and about one-third of this is due to children newly receiving a mother under the experiment. By contrast, just 0.3% of children in the U.S. and Europe experienced a change of any kind.

[Table A1 about here]

However, changes were concentrated in only a few relationship categories. Grandchildren were the largest group, accounting for nearly 60% of all children under age 18 with a new MOMLOC value. The

[13] For instance, consider a household that is ordered by relationship to head and by age and with no data on women's fertility. The household enumeration lists a head, his spouse, two married daughters and 5 grandchildren. In the absence of a mechanism to control the distribution of parent-child links, if the second daughter was old enough to be the parent of all 5 grandchildren, then she would receive all 5 links. . Alternatively, if the first listed grandchild was too old to be the child of the second daughter, then potentially all 5 grandchildren would link to the first daughter.

next largest group is the broad category of "other relatives" which accounts for over 30% of MOMLOC changes. For both of these relationship categories, approximately 10 percent of children under 18 received a different MOMLOC value under the experiment than under the IPUMS algorithm.

Table A2 presents data by child age on the proportion of children with a mother under the IPUMS algorithm and under the assumption that fertility is unknown. Under experimental conditions, we consistently link at a higher rate than when we use known fertility, but the differences are small, ranging from about 1% for children under 5 to less than 0.3% percent at ages 10 and older. The largest differences are found in the African samples, where the proportion of children linked to a mother increased by nearly 1.6% overall, and by 2.5 percent among children under 5 years of age.

[Table A2 about here]

Table A3 presents estimates of the age difference between mother and children who are linked by MOMLOC under the IPUMS algorithm and the experimental algorithm. The percentage of children with a teen mother increased slightly, from 12.2% to 12.6%, while the proportion of children with mothers ages 25 and older decreased accordingly. Estimates of teenage childbearing increased by 0.8 percentage point in the African samples, by 0.3 percentage points in Asia and Central and South America, and only negligibly in the U.S. and Europe. If we concentrate on grandchildren, the category most affected by the experiment, we find that the proportion of grandchildren with a teenage mother increased from 25% to 28%.

[Table A3 about here]

Table A4 presents data on persons under age 18 who are linked to never-married mothers by the IPUMS algorithm and the experimental algorithm. Overall, the percentage of children with never-married mothers increased from 5.0% to 5.4%. Increases occurred at all ages, with the largest increases observed among children under age 5 (7.1% to 7.6%). Most of this increase can be attributed to changes in African samples, where the proportion of children linked to an unmarried mother increased from 8.3% to 9.0%. We observe small decreases in our estimate of never-married mothers in only 4 of the 76 samples included in our experiment. This suggests that in samples where fertility is unknown, the IPUMS algorithm will result in a small overestimate of unmarried parenthood.

[Table A4 about here]

Table A1. Percent of persons receiving a different mother or father pointer under the fertility experiment

	% change in parent pointers					
	All persons	Children < 18				
Type of change	All regions	All regions	Africa	Asia	Central & South America	U.S. & Europe
MOMLOC value	0.6%	1.5%	3.1%	0.8%	1.3%	0.3%
Has a mother	0.3%	0.7%	1.5%	0.4%	0.7%	0.1%
POPLOC value	0.2%	0.4%	0.5%	0.5%	0.5%	0.1%
Has a father	0.1%	0.3%	0.4%	0.3%	0.4%	0.1%

Note: analyses are restricted to sample that collected fertility data for women regardless of age or marital status. In the fertility experiment, fertility caps were calculated and used in place of known fertility.

Table A2. Percent of children linked to a mother under the IPUMS algorithm and under the assumption that fertility is unknown

	All regions		Africa		Asia		Central & South America		U.S. & Europe	
Child Age	Fertility cap	Calculated cap	Fertility cap	Calculated cap	Fertility cap	Calculated cap	Fertility cap	Calculated cap	Fertility cap	Calculated cap
0-4	92.4%	93.3%	85.8%	88.0%	96.4%	96.9%	93.1%	93.8%	96.3%	96.5%
5-9	89.8%	90.3%	78.7%	80.0%	95.5%	95.8%	90.8%	91.3%	96.3%	96.4%
10-14	86.3%	86.6%	72.9%	73.5%	93.1%	93.3%	86.9%	87.2%	94.9%	95.0%
15-17	77.7%	77.9%	63.2%	63.4%	86.6%	86.7%	77.1%	77.2%	88.5%	88.5%
Total	87.7%	88.3%	77.3%	78.5%	93.6%	93.9%	88.4%	88.8%	94.4%	94.5%

Note: analyses are restricted to sample that collected fertility data for women regardless of age or marital status. In the fertility experiment, fertility caps were calculated and used in place of known fertility.

Table A3. Age differences between mothers and children, under the IPUMS algorithm and under the assumption that fertility is unknown

	All regions		Africa		Asia		Central &South America		U.S. & Europe	
Age Difference	Fertility cap	Calculated cap	Fertility cap	Calculated cap	Fertility cap	Calculated cap	Fertility cap	Calculated cap	Fertility cap	Calculated cap
15-19	12.2%	12.6%	15.7%	16.5%	5.6%	5.9%	13.7%	14.0%	8.3%	8.4%
20-24	28.8%	28.8%	25.6%	25.6%	30.3%	30.2%	28.2%	28.1%	34.2%	34.1%
25-29	26.5%	26.4%	23.5%	23.2%	30.6%	30.4%	25.4%	25.3%	31.1%	31.1%
30-34	17.3%	17.2%	16.7%	16.5%	18.3%	18.2%	17.3%	17.2%	17.0%	16.9%
35-39	15.1%	15.1%	18.5%	18.3%	15.2%	15.2%	15.4%	15.4%	9.4%	9.4%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Note: analyses are restricted to sample that collected fertility data for women regardless of age or marital status. In the fertility experiment, fertility caps were calculated and used in place of known fertility.

Table A4. Percent of children linked to a never-married mother under the IPUMS algorithm and under the assumption that fertility is unknown

	All regions		Africa		Asia		Central & South America		U.S. & Europe	
Child age	Fertility cap	Calculated cap	Fertility cap	Calculated cap	Fertility cap	Calculated cap	Fertility cap	Calculated cap	Fertility cap	Calculated cap
0-4	7.1%	7.6%	10.4%	11.5%	1.3%	2.1%	7.2%	7.5%	7.1%	7.4%
5-9	4.8%	5.2%	7.9%	8.7%	0.7%	1.0%	5.2%	5.5%	3.6%	3.7%
10-14	3.8%	4.1%	6.7%	7.1%	0.5%	0.6%	4.3%	4.6%	2.3%	2.3%
15-17	3.3%	3.4%	6.0%	6.2%	0.4%	0.5%	3.8%	3.9%	1.9%	1.9%
Total	5.0%	5.4%	8.3%	9.0%	0.8%	1.1%	5.4%	5.7%	3.8%	3.9%

Note: analyses are restricted to sample that collected fertility data for women regardless of age or marital status. In the fertility experiment, fertility caps were calculated and used in place of known fertility.