MINNESOTA POPULATION CENTER

UNIVERSITY OF MINNESOTA

Identifying Matrilineal Kin Networks in the United States 1900-1940

Matt A. Nelson†
University of Minnesota

August 2024

Working Paper No. 2024-02
DOI: https://doi.org/10.18128/MPC2024-02

†Address correspondence to Matt Nelson, nels5091@umn.edu) Research Scientist, Institute for Social Research and Data Innovation, University of Minnesota. Support for this work was provided by the Minnesota Population Center at the University of Minnesota (P2C HD041023).

Identifying Matrilineal Kin Networks in the United States 1900-1940

Matt A. Nelson (0000-0002-8849-4628, nels5091@umn.edu)
Research Scientist, Institute for Social Research and Data Innovation
University of Minnesota

National-level analysis of historical women's life courses and kinship networks is understudied due to data limitations. Previous analysis of patrilineal kin propinquity in the United States relied on surnames to identify the probability of non-random isonymy as a proxy for nearby kin. Two limitations of this approach include not knowing whether same surname matches were truly kin and losing matrilineal kinship networks due to women's changing surnames upon marriage. Using linked census data from IPUMS MLP, LIFE-M, and Census Tree for 1850-1940, this project aims to identify birth surnames for married women to better measure matrilineal kinship networks in the United States. Results using linked census data for Ohio and North Carolina suggest non-coresidential matrilineal kinship networks underwent a similar decline as proximate patrilineal kinship networks in this time period, but nuances suggest that we should not treat these declines similarly. Matrilineal and patrilineal kin propinquity differed in the life course patterns for currently married women. Patrilineal kin propinquity exhibited a U-shaped trend over the life course while matrilineal kin propinquity declined over the life course. While this paper focuses only on Ohio and North Carolina and does not directly identify kin, national-level analyses are possible and the data produced here will allow other research to easily identify relations from other censuses with data quality flags indicating how relationships are determined. This marks a revolutionary improvement in national-level analyses for women's life course research.

Keywords: census, women's life course, linked data, kinship

Introduction

Studying women's full life courses using historical demographic data in the United States has been complicated by a lack of data due to women's changing surnames preventing linking between pre- and post-marital states. Relatedly, we are unable to study the national prevalence of matrilineal kinship networks due to this data limitation. This is important for studies on fertility, mortality, migration and family structure as kinship networks influenced all of these demographic outcomes in many different spatial and temporal contexts (Dillon et al. 2024, Engelhardt et al. 2019, Hacker and Roberts 2017, 2019; Hacker et al. 2021, Harton et al. 2023). While recent research leveraged the IPUMS United States full count census data to estimate patrilineal kin propinquity in the United States between 1790-1940, it did not identify matrilineal kin proximity (Nelson 2020). This paper will use three different linked datasets to identify matrilineal kin, focusing on ever-married women who's surnames changed upon marriage.

The importance of this work for studying the life course of women cannot be understated. One of the largest data limitations in historical demography is the loss of information on women after marriage and the inability to study early life effects on later life outcomes at a national scale for these women. With linked census data, researchers can explore these kinship links to study intergenerational mobility, demographic outcomes, and socioeconomic inequalities for ever-married women. This paper discusses the creation of kinship links using linked census data, comparisons between different linked datasets, analysis of patrilineal and matrilineal kinship networks in Ohio and North Carolina for married women, and some limitations and biases in the data. While the analysis in this paper focuses on two states, the analysis can easily be scaled up to the national-level. These new kinship data marks a revolutionary improvement in national-level analyses for women's life course research in the United States in the early twentieth century.

Background

Demographic processes and kinship networks shape each other intergenerationally, specifically with fertility, mortality, migration, and marriage patterns determining the prevalence and extent of kinship networks, while at the same time kin provided child care, affected migration decisions, and exchanged economic resources which shaped demographic processes in turn (Daw, Verdery, Margolis 2016; Hacker et al 2021; Harton et al 2023, Nelson 2019; Newson et al. 2005, 2007; Newson & Richerson 2009). Due to data limitations, our understanding of these intricate relationships between demographic processes and non-coresidential kinship networks are not well understood for the nineteenth and early twentieth century United States. While local studies often contain an amazing amount of detail on kinship networks, these studies tend to be highly selected populations and difficult to generalize more broadly (Billingsley 2004, Nelson 2019). This is particularly true of migrants (harder to track over time) and women (historically have changed their surname upon marriage and also difficult to track over time).

Historical U.S. demographic research that does focus on the role of kin on demographic processes has typically focused on coresidential kin (Esteve & Reher 2021, 2024; Ruggles 1994, 2007, 2011). While studying coresidential kin is an important aspect of demographic outcomes, it excludes the locality of residence for married couples when they leave their households of origin. While United States residence patterns were typically defined by neolocal residency, individuals still often lived near patrilineal and matrilineal kin (Fawver 2006, Nelson 2020, Ruggles 2015). Relatedly, U.S. kinship networks were distinguished by their emphasis on multilineal kinship networks rather than patrilineal or matrilineal which is lost when we focus only on coresidential kin (Parsons 1943).

Nelson (2020) and Smith (1989) have established long-run declines in patrilineal kin propinquity in the United States since 1790, but even in 1900 nearly 30% of individuals lived near another family with the same surname non-randomly. This matters because geographic proximity has been shown to play important roles in reproductive outcomes, kin assistance, and migration decisions (Engelhardt et al. 2019; Hacker et al. 2021; Koylu et al. 2021). Unfortunately, this research on non-coresidential kin proximity experienced data limitations by focusing on the patrilineal line and on men, ignoring the bilateral descent kin structure in the United States due to data limitations. Matrilineal kin networks at a national scale have been vastly understudied due to these data limitations. With new linked census datasets however, we can now start overcoming some of these limitations from a data perspective.

DATA

To analyze matrilineal kinship networks, we need linked census data to identify the kinship networks for ever-married women. For this comparison of linked census data I use three datasets; the IPUMS Multigenerational Longitudinal Panel (IPUMS MLP), the Longitudinal Individual Family Electronic Micro-database (LIFE-M), and Census Tree. IPUMS MLP is a supervised learning method that probabilistically links census records. It links cases in two steps; first by linking individuals using a large number of features for the highest accuracy, and then in the second step linking remaining household members that are present in both censuses. The IPUMS MLP data used in this paper does not use administrative records to link cases, so one limitation is we only capture a small subset of potential matrilineal links (Helgertz et al. 2021, Ruggles et al. 2021).[1]

[1] The version of the data used here is version 1.1. Version 1.2 linked MLP data using administrative data from the Social Security Numident records were released in the spring of 2024 after most of the analyses of this paper were completed.

Data from the LIFE-M project (Bailey et al. 2022) and the Census Tree Project (Price et al. 2021, Buckles et al. 2023) are further used to validate the results. LIFE-M is limited to persons from administrative records in Ohio and North Carolina, but using administrative birth, marriage, and death records to link persons provides more potential links for matrilineal kin.[2] The individuals and families are linked over multiple steps to a variety of data sources using a supervised machine learning algorithm. The combination of a variety of sources allows for high accuracy in linking but also contains a wealth of information on various socioeconomic and demographic features of interest. Because LIFE-M focused on linking between administrative and vital records, sometimes the LIFE-M data only links to one census in the IPUMS full count data. To maintain similar data structures as IPUMS MLP and the Census Tree data, only LIFE-M data with links to two or more censuses is used here.

The Census Tree data were developed by combining multiple datasets to increase the number of links while adjudicating links between methods. The Census Tree data benefits from combining links from their XGBoost linking method with IPUMS MLP, the Census Linking Project (CLP), crowdsourced genealogical data from FamilySearch, and FamilySearch hints generated by a proprietary machine-learning algorithm. Because of the inclusion of genealogical data, Census Tree data allows for identifying additional matrilineal links beyond what IPUMS MLP currently identifies. The current limitation in Census Tree is selecting only non-conflicting links from their crosswalks[3] and that the provenances of some of the proprietary FamilySearch hints data is unknown. Links from the Census Linking Project (CLP) were not used because the

[2] Currently LIFE-M does not provide any links to the 1930 Census.
[3] Census Tree currently offers each census-to-census links for each year (e.g., 1920-1930, 1920-1940, and 1930-1940 links are all different datasets). If 1920 Person A links to 1930 Person B, and 1930 Person B links to 1940 Person C, but 1920 Person A links to 1940 Person D, I consider these links conflicting. While researchers can use the identified methods to potentially select which of the conflicting links to keep, for this project, all conflicting links are dropped from the analysis.

Census Linking Project only links men (Abramitzky et al. 2020). Census Tree links were selected by dropping CLP only links, and the remaining links were combined with all conflicting links removed.

METHODOLOGY

Identifying parents

Analyzing linked census records enables capturing a more representative sample of the population with matrilineal kin networks. Imagine a household with a married couple in 1900 that is linked to 1910 in Figure 1. In 1910, this married couple is now living with the wife's parents with the surname of Faber. Since the wife's record is linked to 1900, we can pull the Faber surname to the 1900 records, which allows us to identify potential matrilineal kin for these years in the absence of a matrilineal surname from 1900. We can also identify known kin with this approach. If the wife's parents lived in their own household in 1900 nearby the married couple and are linked to the 1910 census, we can directly identify these people are kin. Identifying known kin is currently limited to parents, siblings, and children, although in theory one could identify other kin such as grandparents, cousins, aunts and uncles, nieces, and nephews, etc. Currently this analysis does not consider stepparents, and cases with conflicting parental identification are dropped from the analysis.

Figure 1: Example of Linked Census Records and Identifying Matrilineal Kin

Table 1 provides the percentage count for each year of identified mothers in IPUMS MLP, Census Tree, and LIFE-M. As an example for MLP, in 1900 no mother is identified in any census for 42.5% of all linked census records. 47.2% of linked 1900 records are living with their mother in 1900, 9.6% are living with their mother in another census year, 0.7% of linked records had their mother identified via another relation (e.g. ego lived with a sibling in 1900 and their sibling lived with their mother in 1910). 0.6% of cases the mother was not identified because of conflicting links.[4] This could indicate stepparents and bad links. The numbers were similar for fathers.

Because of the genealogical data, Census Tree captures more mothers beyond the household than MLP, but more conflicting mother links were identified as well. This could be due to several reasons. First, Census Tree has more links than MLP and on average links an individual more often between censuses, which raises the likelihood of a bad match being made. Second, since Census Tree combines multiple methods together, it's possible that because of the intersecting nature of links that if a bad match is made in one method, it's cascading into other methods causing further bad matches. By the 20th century, Census Tree links improve with far lower rates of conflicting mother identifications and significantly more mothers identified in different census years.

While the sample size of the LIFE-M data are significantly smaller and less representative than MLP and Census Tree, the use of administrative records dramatically improves the number of mothers identified from a different census year and has the unique feature of having no conflicting mother links in the data. Table 1 shows over 40% of linked individuals in 1940 LIFE-M data with their parents identified in the data from a different census year.

Table 1: Data Quality Flag for Identified Mothers by Linking Method, 1850-1940

This approach is repeated for each of the linked datasets. The next step is identifying the birth surnames for ever-married women. Because linking strategies do not require exact surname matches, one factor to account for is multiple surnames. For example, if you have a linked individual linked across five censuses, one linked record could in theory have five different

[4] Conflicting links did appear more likely in cases where an individual was linked over more censuses. Further investigation could focus on identifying which of the conflicting links are correct given a particular linked individual or identifying links that are likely incorrect for an ego based on the parental information.

surnames due to enumerator or data transcription error. To limit this issue, I compare each identified parental surname to each other by computing a Jaro-Winkler similarity score. A score of 1 means the strings match exactly with a lower number indicating less similarity between strings. Surnames that have a Jaro-Winkler similarity greater than 0.8 are considered to be the same surname while a score of less than 0.8 represents a different surname. Any linked records that have three or more different surnames across linked censuses are dropped from analysis. While some of these cases could represent legitimate surnames such as a mother remarrying, it could also represent an incorrectly linked record. While data from LIFE-M and Census Tree also go through a similar approach, we can directly observe name changes of many women, which allows for capturing a larger part of the potential universe as seen in Table 2.

Table 2: Sample Size Linked Ever Married Women by Dataset, 1850-1940

Table 2 shows the sample size of linked ever married women for each dataset and the sample size once parental identification is completed and conflicts are removed. Census Tree identifies the largest number of birth surnames for ever-married women, but in terms of the total universe, LIFE-M identifies the largest proportion of possible birth surnames. As an example, although MLP identifies more ever-married women links in 1940 relative to LIFE-M because of LIFE-M sampling restrictions (13.7 million to 340,000), because we directly observe name changes in the LIFE-M data, we capture a larger percentage of linked ever married women with potential matrilineal kin (11% vs 42%).[5]

When comparing the MLP and Census Tree mother links, we find that generally the methods either are in agreement or identify mothers individually. In general when a mother is

[5] Preliminary results suggest that the addition of links to the Numident records for IPUMS MLP will increase the proportion of identified birth surnames for women to approximately 29%, putting the method on par with Census Tree in terms of the proportion of identified birth surnames for ever-married women.

identified in one of the linked datasets, less than 1.5% of the time do the methods disagree. Moving forward in time, the two methods are more likely to identify a mother. In 1850 only 3.6% of mothers are identified who were linked from a different census year. By 1940, 42.5% of mothers who are not living with their children can be identified from the linked census data. Once you add in mothers who are living with their children, this increases to 76.5% of mothers identified in 1940.[6] At this stage, disagreeing parental links are simply discarded, although future work should look into conflicts to determine which links are generally correct.

Kin Propinquity in Ohio and North Carolina, 1900-1940

As a proxy for kinship, I calculate the probability of non-random isonymy for both patrilineal and matrilineal kin as defined by Nelson (2020). One important note here is patrilineal and matrilineal kin propinquity are defined at the married couple level, which means patrilineal identifys the husband's potential kin while matrilineal identifies the wife's potential kin. The analysis will focus on currently married women because LIFE-M primarily identifies currently married women and not as many widows, divorcees, etc. Further, the analysis only looks at North Carolina and Ohio to compare directly with LIFE-M, although national analysis can be performed using MLP and Census Tree data.[7] Finally, because LIFE-M does not link to the 1930 census, any results for LIFE-M and 1930 are simply the average trend between 1920 and 1940. Table 3 describes the sample sizes of currently married women by method in Ohio and North Carolina between 1900 and 1940. To control for representativity issues in the linked census data, I calculate inverse probability weights as described by Bailey et al. (2023) and Abramitzky et al. (2020). For inverse probability weighting, the propensity scores were modeled as a function of five-year age

[6] Results are not presented here for fathers but are essentially the same.
[7] Because LIFE-M is drawn from administrative and vital records from Ohio and North Carolina, estimates show that 89-95% of the LIFE-M cases in any given year reside within these states.

groups, marital status, state of residence, migration status (lived in state of birth, lived in region of birth, native-born, foreign-born), race (white, Black, American Indian, other), urban status, and farm residence status.

Table 3: Sample Size Linked Currently Married Women with Surname at Birth by Dataset in Ohio and North Carolina, 1900-1940

In general, the three methods have different strengths to the underlying populations. Census Tree performs particular well in the occupational structure of husbands, with the largest different between the full underlying population related to overidentifying farmers and under identifying laborers. LIFE-M and Census Tree both do well in identifying rural farm, rural non-farm, and urban status of linked cases. While all three linking methods underidentify non-whites in the linked data, Census Tree performs slightly better than MLP, while LIFE-M only identifies 1-2% of the linked population as non-white compared to the true population rate of nearly 10%. LIFE-M also overidentifies younger persons in the data, likely due to their reliance on birth, death, and marriage records for linking cases. Finally, while MLP still underidentifies interstate migrants and foreign-born relative to the full population, MLP performs far better than both Census Tree and LIFE-M in identifying migrants in their linking algorithm. These are all important features to consider as will be shown in the results section.

RESULTS

Aggregate kin propinquity rates

The kin propinquity rates for patrilineal kin for each linking method can be compared directly to the full population of currently married women in Ohio and North Carolina as a baseline to determine the general accuracy of the weighted results for patrilineal kin. I find that weighting the data does change the results slightly. When unweighted, LIFE-M and Census Tree generally showed higher patrilineal kin propinquity rates than MLP. Once weighted, the differences between methods are minor and generally reflect the full population. When comparing the results of patrilineal kin propinquity for each method to the cross-sectional patrilineal kin propinquity rate for all currently married women in Ohio and North Carolina, on average MLP performs the closest to the population rate.

While all three methods show a general decline in matrilineal kin, their rates do appear to differ from each other, even when the results are weighted. This suggests that some of the biases of the underlying linked data could be driving these results. Census Tree data was more likely to have non-migrants in their data relative to MLP, and non-migration is associated with higher kinship network density which could explain why Census Tree has higher rates of matrilineal kin propinquity relative to MLP (Nelson 2020). For LIFE-M, because the populations tended to be younger relative to MLP and Census Tree, younger persons were associated with lower kin propinquity typically, which could explain why their kin propinquity rates in some cases are lower than MLP and Census Tree. Prior to 1930, patrilineal kin propinquity was a few percentage points higher than matrilineal kin propinquity. After 1930, both MLP and Census Tree show higher matrilineal kin propinquity. LIFE-M differs from MLP and Census Tree and shows significantly lower matrilineal kin propinquity relative to patrilineal kin propinquity for the entire time frame.

Figure 2: Patrilineal and Matrilineal Kin Propinquity Rates by Method for Currently Married Women Age 15+ Ohio and North Carolina, 1900-1940

Five-year age group kin propinquity rates

As seen in Nelson (2020), patrilineal kin propinquity follows a family life cycle interpretation. Patrilineal kin propinquity declines from age 15 to approximately age 40, then increases afterwards to age 70. This likely reflects the earliest years of declining kin availability (mortality of older generations, migration patterns of younger generations) and later years of increasing kin availability (declining mobility of older generations, children leaving household of origin). LIFE-M results differ slightly from MLP and Census Tree with a low point of patrilineal kin propinquity in 1910 at age 50 rather than 40. Older age group for LIFE-M in 1910 are not presented due to sample size limitations. Similar patterns can be seen in 1940.

Figure 3: Patrilineal and Matrilineal Kin Propinquity 5-Year Age Group Rates by Method for Currently Married Women Age 15+ Ohio and North Carolina, 1910 & 1940

Matrilineal kin propinquity on the other hand does not follow the same pattern as patrilineal kin propinquity. This can potentially be attributed to changing surnames, since the women's children will have the patrilineal surname and not the matrilineal surname, meaning any children who live nearby when a person is elderly will have the patrilineal surname. Related to that is kin availability; as time goes on, fewer kin with the matrilineal surname will survive while the patrilineal surname is reproduced via sons and unmarried daughters. While the trends for all three methods are generally consistent, Census Tree shows slightly higher kin propinquity rates in all age groups relative to MLP and LIFE-M.

When comparing the three methods to the full cross-sectional patrilineal kin propinquity results, we find that they all describe the same trend generally speaking, although the rates differ by a larger margin. Census Tree performed best in 1910 (on average 2.5 percentage points difference from full population compared to 3.6% and 2.6% for MLP and LIFE-M respectively). MLP performed best in 1940 (on average 1.5% point difference from full population compared to 4.4% and 3.4% for Census Tree and LIFE-M respectively).

LIMITATIONS

The primary considerations researchers should consider with this research are the representivity of linked census data, the indirect measure of kinship presented here, and the cross-sectional analysis in this paper. While linked census data corrects for some of the population representativity issue, this data is not perfectly representative of the true population, particularly the genealogical data (Buckles et al. 2023, Helgertz et al. 2022, Koylu et al. 2021, Price et al. 2021, Stelter & Alburez-Gutierrez 2022). This is especially true for studying racial inequalities, where linkage rates for non-whites are lower. The linked data are selective of the wider population, and researchers should consider how the linked population (and by extension their kinship networks) are representative of the population at large.

Based on the strengths and weaknesses of each linking method, even when weighting for the underlying demographic and socioeconomic characteristics of the data, some of these results could be driven by these biases. As an example, Census Tree is less likely to identify migrants and more likely to identify farmers, both of which were groups with higher kin propinquity rates. While the general kin proximity trends between methods are the same, the higher Census Tree rates could be due to this underlying bias. Conversely, LIFE-M tends to overidentify younger persons, who tend to have lower kin propinquity than individuals over the age of 40.

The example analysis in this paper (kin propinquity) is not a direct measure of kinship but a measure of non-random isonymy. Since researchers can now capture some direct kinship links in the linked census data, research should push for more direct measures when possible. Another limitation of the kin propinquity measure is that urban enumeration districts were geographically smaller than rural enumeration districts and it is likely that kin propinquity in urban areas is biased downwards (Nelson 2020). Finally, while the kinship links in this linked data would allow for some reconstruction of life courses for individuals, this research paper focused on cross-sectional analysis. Future research would benefit from analyzing this data in a longitudinal format and tracking how kinship networks changed over the life course for individuals.

DISCUSSION

This paper establishes how to use linked census data to identify familial relations beyond the household and use that information to study kinship networks. While patrilineal and matrilineal kin propinquity both declined during this time period, the declines were not similar over the life course, suggesting that these kinship networks potentially operated differently. Further, the linked datasets do show slightly different results in the absolute values of kin propinquity, suggesting that researchers need to consider the strengths and weaknesses of each linked data set when determining which linking method to use in their analyses.

Perhaps most importantly, this work shows how researchers can better study kinship structures using linked census data, specifically for ever-married women whose kinship networks are often lost when they change their birth surname to the husband's surname. While the results for this paper only described Ohio and North Carolina, the analysis can be scaled up to a national-level using either IPUMS MLP or Census Tree data. Future research on demographic and socioeconomic outcomes need to consider the effects of both coresidential kin and non-coresidential kin living nearby. This linked data has the potential to revolutionize national-level analyses of women's life courses, and this paper is just one example of how we can do that.

Acknowledgements

This research is supported in part by funding from the Minnesota Population Center (P2C HD041023) and by grants from the Eunice Kennedy Shriver National Institute for Child Health and Human Development (R01AG057679, R01HD043392, & R01AG041831). I would like to thank audiences from the Social Science History Association 2022 and 2023 Annual meetings, and the workshop group for this special issue for their feedback and suggestions.

Disclosure: The authors report there are no competing interests to declare.

BIBLIOGRAPHY

Abramitzky, R., Boustan, L., Eriksson, K., Pérez, S., & Rashid, M. (2020). Census Linking Project: Version 2.0 [dataset]. https://censuslinkingproject.org

Bailey, M., Cole, C., & Massey, C. (2020). Simple strategies for improving inference with linked data: A case study of the 1850–1930 IPUMS linked representative historical samples. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *53*(2), 80–93. https://doi.org/10.1080/01615440.2019.1630343

Bailey, M.J., Lin, P.Z., Mohammed, A.R.S., Mohnen, P., Murray, J., Zhang, M., and Prettyman, A. LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2022-10-05. https://doi.org/10.3886/E155186V4

Billingsley, C. E. (2004). *Communities of kinship: Antebellum families and the settlement of the cotton frontier*. University of Georgia Press.

Buckles, K., Haws, A., Price, J., & Wilbert, H. "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project." Working Paper, 2023.

Daw, J., Verdery, A. M., & Margolis, R. (2016). Kin count (s): Educational and racial differences in extended kinship in the United States. *Population and Development Review*, *42*(3), 491-517.

Dillon, L., Chernenko, A., Dribe, M., Engelhardt, S., Gagnon, A., Hanson, H.A., Meeks, H., Quaranta, L., Smith, K.R., & Vézina, H., (2024), "Did Grandmothers Enhance Reproductive Success in Historic Populations?: Testing Evolutionary Theories on Historical Demographic Data in Scandinavia and North America," in *Human Evolution Demography*, ed. Oskar Burger, Ronald Lee, and Rebecca Sear, 475-502, https://doi.org/10.11647/obp.0251.20.

Engelhardt, S.C., Bergeron, P., Gagnon, A., Dillon, L., & Pelletier, F., (2019), "Using Geographic Distance as a Potential Proxy for Help in the Assessment of the Grandmother Hypothesis" *Current Biology* 29: 651-656.

Esteve, A., & Reher, D. S. (2021). Rising global levels of intergenerational coresidence among young adults. *Population and Development Review*, *47*(3), 691–717.

Esteve, A., & Reher, D. S. (2024). Trends in Living Arrangements Around the World. *Population and Development Review*.

Fawver, K. (2006). Gender and the structure of planter households in the eighteenth-century Chesapeake: Harford County, Maryland, in 1776. *Early American Studies*, *4*(2), 442–470.

Hacker, J.D., & Roberts, E. (2017), "The Impact of Kin Availability, Parental Religiosity, and Nativity on Fertility Differentials in the Late 19th-Century United States," *Demographic Research*, 37:34, 1049-1080.

-- (2019), “Fertility Decline in the United States, 1850-1930: New Evidence from complete-count datasets,” *Annales de Démographie Historique* 138:2, 143-177.

Hacker, J. D., Helgertz, J., Nelson, M. A., & Roberts, E. (2021). The Influence of Kin Proximity on the Reproductive Success of American Couples, 1900–1910. *Demography*, *58*(6), 2337–2364. https://doi.org/10.1215/00703370-9518532

Harton, M.-È., Hacker, J. D., & Gauvreau, D. (2023). Migration, Kinship and Child Mortality in Early Twentieth-Century North America. *Social Science History*, *47*(3), 367–395. Cambridge Core. https://doi.org/10.1017/ssh.2023.11

Helgertz, J., Price, J., Wellington, J., Thompson, K. J., Ruggles, S., & Fitch, C. A. (2022). A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *55*(1), 12–29. https://doi.org/10.1080/01615440.2021.1985027

Helgertz, J., Ruggles, S., Warren, J.R., Fitch, C., Goeken, R., Hacker, J.D., Nelson, M., Price, J., Roberts, E., & Sobek, M. (2021). IPUMS Multigenerational Longitudinal Panel: Version 1.1 [dataset]. *Minneapolis, MN: IPUMS*, https://doi.org/10.18128/D016.V1.1.

Koylu, C., Guo, D., Huang, Y., Kasakoff, A., & Grieve, J. (2021). Connecting family trees to construct a population-scale and longitudinal geo-social network for the U.S. *International Journal of Geographical Information Science*, *35*(12), 2380–2423. https://doi.org/10.1080/13658816.2020.1821885

Nelson, M. A. (2019). Agricultural family and neighbor labor in nineteenth-century Minnesota. *Historicka Demografie*, *43*(2), 211–231.

Nelson, M.A. (2020). The Decline of Patrilineal Kin Propinquity in the United States, 1790-1940. *Demographic Research*, *43*. https://www.demographic-research.org/volumes/vol43/18/

Newson, Lesley, and Peter J. Richerson. “Why Do People Become Modern? A Darwinian Explanation.” *Population and Development Review* 35, no. 1 (March 1, 2009): 117–58. https://doi.org/10.1111/j.1728-4457.2009.00263.x.

Newson, L., Postmes, T., Lea, S. E. G., & Webley, P. (2005). Why Are Modern Families Small? Toward an Evolutionary and Cultural Explanation for the Demographic Transition. *Personality and Social Psychology Review*, *9*(4), 360–375. https://doi.org/10.1207/s15327957pspr0904_5

Newson, L., Postmes, T., Lea, S. E. G., Webley, P., Richerson, P. J., & Mcelreath, R. (2007). Influences on communication about reproduction: The cultural evolution of low fertility. *Evolution and Human Behavior*, *28*(3), 199–210. https://doi.org/10.1016/j.evolhumbehav.2007.01.003

Parsons, T. (1943). The kinship system of the contemporary United States. *American Anthropologist*, *45*(1), 22–38.

Price, J., Buckles, K., Van Leeuwen, J., & Riley, I. (2021). Combining family history and machine learning to link historical records: The Census Tree data set. *Explorations in Economic History*, *80*, 101391. https://doi.org/10.1016/j.eeh.2021.101391

Ruggles, S. (1994). The transformation of American family structure. *The American Historical Review*, *99*(1), 103–128.

Ruggles, S. (2007). The Decline of Intergenerational Coresidence in the United States, 1850 to 2000. *American Sociological Review*, *72*(6), 964–989. https://doi.org/10.1177/000312240707200606

Ruggles, S. (2011). Intergenerational Coresidence and Family Transitions in the United States, 1850–1880. *Journal of Marriage and Family*, *73*(1), 136–148. https://doi.org/10.1111/j.1741-3737.2010.00794.x

Ruggles, S. (2015). Patriarchy, power, and pay: The transformation of American families, 1800–2015. *Demography*, *52*(6), 1797–1823.

Ruggles, Steven, Nelson, Matt A., Sobek, Matthew, Fitch, Catherine A., Goeken, Ronald, Hacker, J. David, Evan, R., & Warren, J. Robert. (2024). *IPUMS Ancestry Full Count Data: Version 4.0* (4.0) [Dataset]. https://doi.org/10.18128/D014.V4.0

Stelter, R., & Alburez-Gutierrez, D. (2022). Representativeness is crucial for inferring demographic processes from online genealogies: Evidence from lifespan dynamics. *Proceedings of the National Academy of Sciences*, *119*(10), e2120455119. https://doi.org/10.1073/pnas.2120455119

Table 1: Data Quality Flag for Identified Mothers by Linking Method, 1850-1940

			MLP			
Year	Not identfied	In Census Year	In different census year	With other relation	Conflicting links	*N*
1850	40.8%	54.8%	2.5%	0.3%	1.6%	5,960,607
1860	41.4%	54.3%	2.0%	0.9%	1.5%	12,322,075
1870	42.7%	53.0%	2.0%	1.3%	1.1%	18,294,071
1880	48.0%	48.9%	0.8%	1.7%	0.6%	20,487,959
1900	42.5%	47.2%	9.6%	0.7%	0.1%	35,612,447
1910	40.0%	46.9%	12.5%	0.6%	0.0%	52,360,117
1920	40.1%	48.7%	10.7%	0.5%	0.0%	62,928,092
1930	41.3%	47.2%	11.1%	0.4%	0.0%	74,786,099
1940	43.1%	40.5%	16.2%	0.2%	0.0%	52,136,562
			Census Tree			
Year	**Not identfied**	**In Census Year**	**In different census year**	**With other relation**	**Conflicting links**	***N***
1850	44.6%	47.1%	2.9%	0.8%	4.7%	9,161,700
1860	43.3%	46.1%	2.4%	3.1%	5.1%	13,904,000
1870	41.8%	47.5%	2.5%	3.3%	4.9%	19,065,739
1880	41.7%	50.8%	1.2%	3.8%	2.6%	25,434,658
1900	29.6%	50.6%	16.6%	2.1%	1.1%	44,487,930
1910	25.8%	47.5%	24.5%	1.5%	0.7%	59,725,949
1920	22.9%	49.3%	26.4%	1.1%	0.4%	73,850,562
1930	21.9%	46.9%	30.2%	0.8%	0.2%	90,124,319
1940	21.3%	34.7%	43.5%	0.5%	0.1%	76,776,953
			LIFE-M			
Year	**Not identfied**	**In Census Year**	**In different census year**	**With other relation**	**Conflicting links**	***N***
1850						
1860						
1870						
1880	24.4%	75.0%	0.6%	0.0%	0.0%	107,467
1900	44.8%	46.9%	8.3%	0.1%	0.0%	488,802
1910	58.1%	24.6%	17.3%	0.1%	0.0%	626,022
1920	65.6%	6.5%	27.9%	0.0%	0.0%	758,910
1930						
1940	55.2%	2.4%	42.4%	0.0%	0.0%	688,854

Table 2: Sample Size Linked Ever Married Women by Dataset, 1850-1940

	Linked Ever Married Women			Linked Ever Married Women with surname at birth		
Year	MLP	LIFE-M	Census Tree	MLP	LIFE-M	Census Tree
1850	1,196,803		1,822,439	52,805		88,026
1860	2,355,420		2,839,760	87,901		296,762
1870	3,554,992		3,369,698	140,415		432,864
1880	4,603,309	10,695	5,819,067	149,166	233	689,737
1900	8,138,480	117,251	10,246,262	609,703	20,243	2,478,805
1910	11,643,408	228,428	14,088,009	1,013,067	53,933	3,685,669
1920	14,559,344	369,661	18,148,653	1,403,014	119,618	5,109,017
1930	18,255,178		23,578,991	1,773,775		6,584,738
1940	13,750,112	342,374	23,348,208	1,567,530	143,607	7,012,624

Table 3: Sample Size Linked Currently Married Women with Surname at Birth by Dataset in Ohio and North Carolina, 1900-1940

Year	MLP	LIFE-M	Census Tree
1900	44,790	19,068	235,071
1910	73,929	52,496	357,253
1920	104,273	116,877	486,290
1930	130,194		622,656
1940	113,747	132,537	653,057

Figure 1: Example of Linked Census Records and Identifying Matrilineal Kin

1900 Data

Serial	Pernum	Given name	Surname	Relationship to Head	ID
506	1	T A	Schrincher	Head	11893
506	2	E B	Schrincher	Wife	11894
510	1	Aaron	Faber	Head	11904
510	2	Pearl	Faber	Wife	11905

1910 Data

Serial	Pernum	Given name	Surname	Relationship to Head	ID
247	1	Thomas A	Schrincher	Head	11893
247	2	Edith B	Schrincher	Wife	11894
247	3	Priscilla J	Schrincher	Child	
247	4	Aaron	Faber	Father-in-law	11904
247	5	Pearl	Faber	Motehr-in-law	11905

1900 Data w/Kinship IDs

Serial	Pernum	Given name	Surname	Relationship to Head	ID	Mother ID	Father ID
506	1	T A	Schrincher	Head	11893		
506	2	E B	Schrincher	Wife	11894	11904	11905
510	1	Aaron	Faber	Head	11904		
510	2	Pearl	Faber	Wife	11905		

Figure 2: Patrilineal and Matrilineal Kin Propinquity Rates by Method for Currently Married Women Age 15+ Ohio and North Carolina, 1900-1940

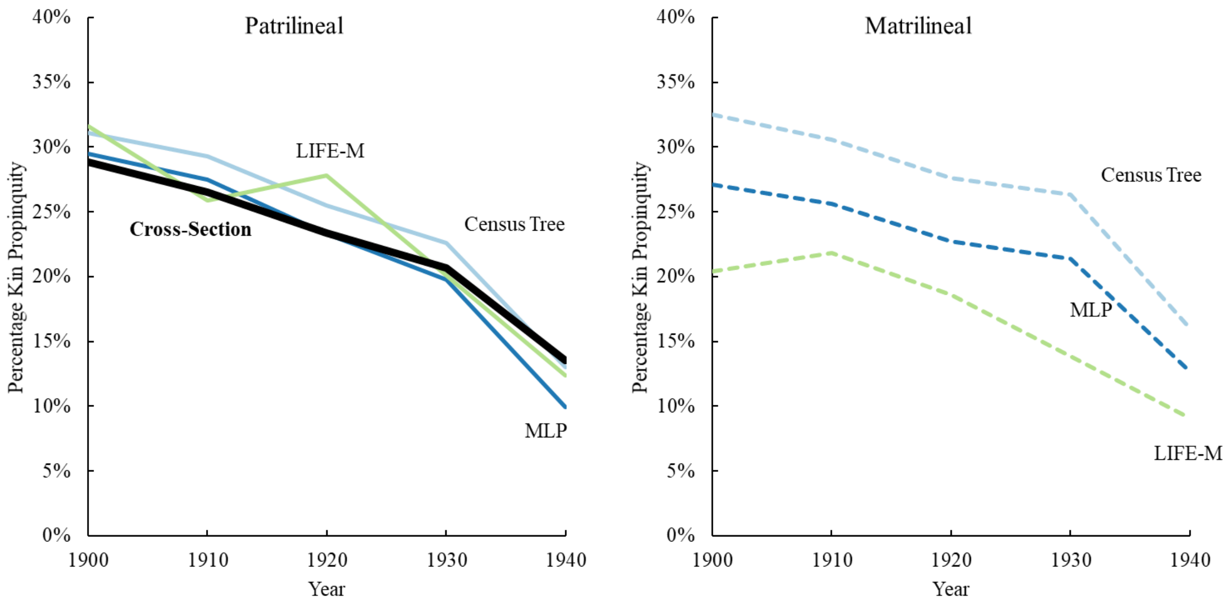
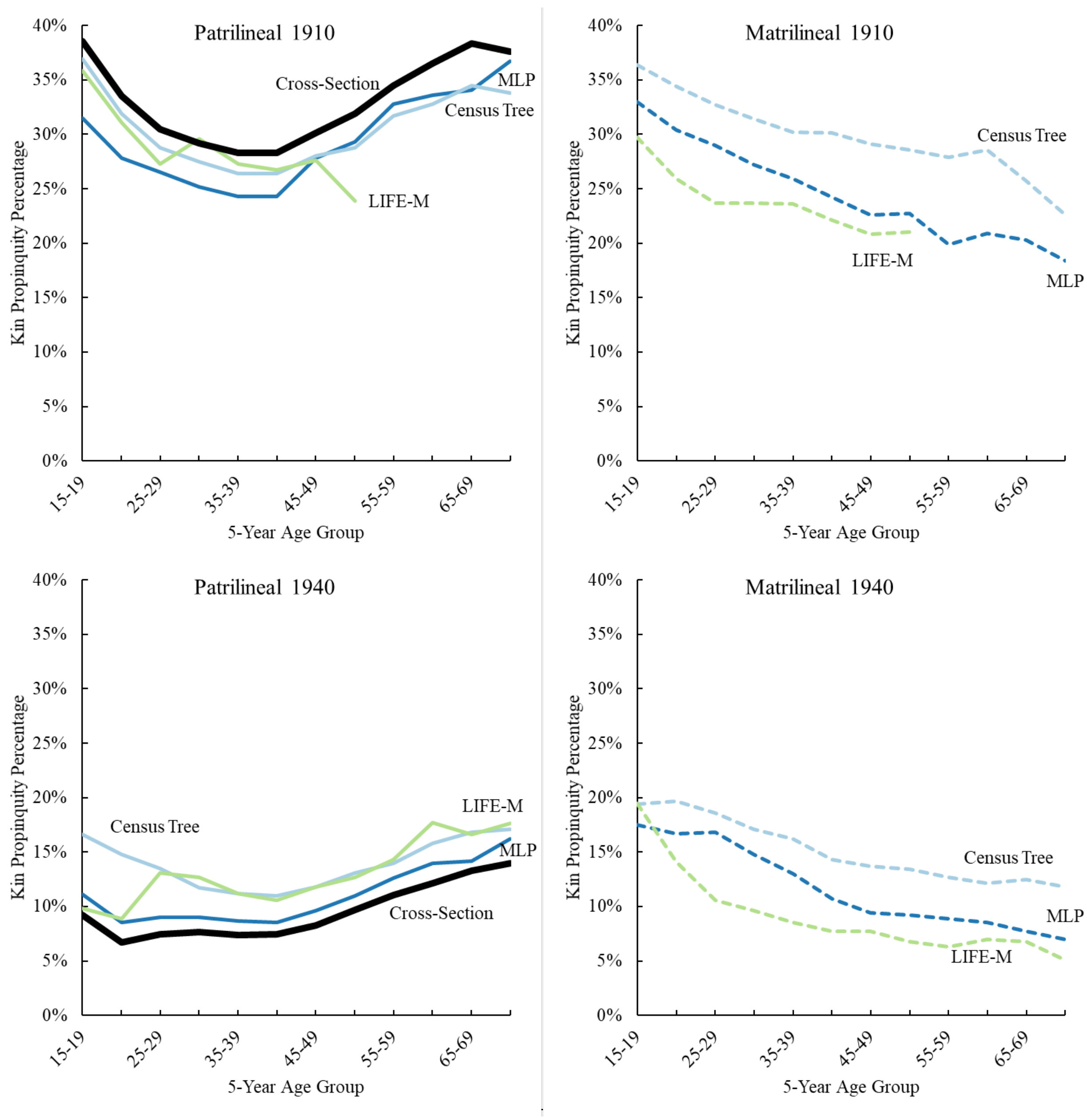


Figure 3: Patrilineal and Matrilineal Kin Propinquity 5-Year Age Group Rates by Method for Currently Married Women Age 15+ Ohio and North Carolina, 1910 & 1940



List of Tables and Figures