

The North Atlantic Population Project

Progress and Prospects

STEVEN RUGGLES

EVAN ROBERTS

SULA SARKAR

MATTHEW SOBEK

*Minnesota Population Center
University of Minnesota*

Abstract. The North Atlantic Population Project (NAPP) is a massive database of historical census microdata from European and North American countries. The backbone of the project is the unique collection of completely digitized censuses providing information on the entire enumerated populations of each country. In addition, for some countries, the NAPP includes sample data from surrounding census years. In this article, the authors provide a brief history of the project, describe their progress to date and plans for the future, and discuss some potential implications of this unique data resource for social and economic research.

Keywords: census, database, demography, international, microdata

The North Atlantic Population Project (NAPP) emerged from a meeting of the International Microdata Access group hosted by Lisa Dillon at the University of Ottawa in April 1999. Steven Ruggles was excited to tell the participants about an agreement he had just reached with the Church of Jesus Christ of Latter-Day Saints (LDS) to gain access to their genealogical database of the entire census of the United States in 1880. Over 1,000 Mormon volunteers had devoted some 2,000,000 hours over 18 years to digitally transcribe the entire U.S. census, comprising more than 50 million records. The LDS was interested in creating a genealogical product on CD-ROM and sought academic collaborators to improve the data and speed up their release schedule. When Ruggles offered to assist the LDS in cleaning the data in exchange for permission to distribute the database freely for academic use, the LDS readily agreed.

As soon as Ruggles related this triumphant news, Dillon revealed that she, too, had been in touch with the LDS and that she anticipated an agreement that would give access to the 1881 census of Canada. Another participant, Matthew Woollard of the UK Data Archive, was already working on

the LDS transcription of the 1881 census of Britain. Gunnar Thorvaldsen of the Norwegian Historical Data Centre at the University of Tromsø then announced that his digital transcription of the 1865 and 1900 censuses of Norway—a project he had pursued for many years—was nearly complete.

The group quickly realized that they had an extraordinary opportunity to combine these massive data sets and create a crossnational database for the late-nineteenth century with almost 90 million observations. In June 2000, we met at the newly established Minnesota Population Center in Minneapolis to define the goals of the project and develop a plan of work. The participants agreed that we should not simply create compatible data sets, but rather should develop a single fully integrated database with common coding systems, constructed variables, documentation, and a single dissemination system.

The concept of NAPP is that a comparatively small investment in collaboration across countries and in shared processing and dissemination can leverage the resources in each country can ensure that the full potential of the data is realized. The project reduces barriers to comparative research by preserving data sets and making them freely available, converting them into a uniform format with standardized machine-actionable metadata, providing comprehensive documentation, and implementing powerful Web-based tools for dissemination and analysis. This is a rare success story of large-scale international collaboration. Producing a single coherent database with staff and funding scattered across a dozen institutions on two continents requires continuous intensive communication and negotiation.

Much of our initial work involved developing compatible coding systems, especially for occupations. We obtained a modest grant to fund our meetings, which we held in suitable North Atlantic sites in Iceland, Norway, England, and Canada. After considerable debate and some controversy, we settled on a modified version of the Historical International Standard Classification of Occupations (also known

Address correspondence to Steven Ruggles, Minnesota Population Center, 50 Willey Hall, 225 19th Ave. S, University of Minnesota, Minneapolis, MN 55455, USA. E-mail: ruggl001@umn.edu

as HISCO), and developed procedures to maximize consistency across countries (Roberts et al. 2003). The merged NAPP coding dictionaries are of unprecedented scale, since they include all alphabetic strings from all participating countries. The occupation dictionary, for example, so far includes 2,605,301 unique strings.

We released early versions of the 1880 U.S. and 1881 Canadian data sets in 2003 and data for 1881 England and Wales and 1900 Norway the following year. In 2006, we added complete data for 1865 Norway, 1881 Scotland, and samples of the Canadian censuses of 1871 and 1901. Since then, we have released a complete census of Sweden and samples of Norway in 1875, Britain in 1851, Mecklenburg, Germany, in 1819, and several U.S. data sets from between 1850 and 1910. We also plan to release additional data sets from Iceland and Canada later this year.

Table 1 summarizes the NAPP data sets in the current project phase and our future plans. The complete enumerations are shown in bold; the other data sets are nationally representative samples. The lower panel of table 1 shows our ambitious planned expansion, which will dramatically increase both the chronological and geographic scopes of the project. Over the coming seven years, we propose to add more than 250 million new person records to the database, more than tripling its size. This massive expansion capitalizes on an explosion of new genealogical data collection projects. The complete-count data from Canada and the United States derive from new LDS projects; the data for England, Wales, and Scotland are being created by the commercial genealogical firm Findmypast.com. In Denmark, Ireland, and Sweden, the data derive from genealogical projects carried out by each country's national archives.

In the current phase of the project, we have multiple complete enumerations from just two countries, Iceland and Norway. When the NAPP expansion is complete, we will have multiple complete enumerations from nine countries: Canada, Denmark, England and Wales, Iceland, Ireland, Norway, Scotland, Sweden, and the United States. For seven of these countries, we will have four or more complete enumerations.

In addition to the genealogical projects, the expansion will incorporate sample data that are being digitized for purely scientific purposes in Albania, Egypt, Germany, Norway, and Mexico. The new partners located on the periphery of the North Atlantic region are especially exciting because they offer different perspectives on the past. The observations from Albania and Egypt represent the first representative national historical microdata from the Islamic world. The sample from Mexico will be the most recent in the NAPP collection; it was taken in 1930, preceding the beginning of mass immigration of Mexicans into the United States.

Linked Data Sets

As soon as the complete-count census data became available, Joseph P. Ferrie and Jason Long got the idea of linking

TABLE 1. Major Components of the North Atlantic Population Project

Year	Country	Cases (in thousands)
Current database		
1852	Canada	170 ¹
1871	Canada	62 ²
1881	Canada	4,278
1891	Canada	350 ¹
1901	Canada	265 ¹
1851	England and Wales	376 ²
1881	England and Wales	26,125
1703	Iceland	50
1801	Iceland	47
1880	Iceland	72
1901	Iceland	78
1801	Norway	879
1865	Norway	1,702
1875	Norway	639 ¹
1900	Norway	2,294
1851	Scotland	22 ²
1881	Scotland	3,728
1900	Sweden	4,576
1850	United States	198 ²
1860	United States	354 ²
1870	United States	428 ²
1880	United States	50,486
1900	United States	5,220 ¹
1910	United States	1,271 ²
Total cases		103,670
Proposed additions		
1918	Albania	140 ¹
1852	Canada	2,436
1861	Canada	3,230
1871	Canada	3,689
1891	Canada	4,833³
1901	Canada	5,371³
1911	Canada	372 ¹
1787	Denmark	842
1801	Denmark	929
1845	Denmark	1,357
1880	Denmark	1,969
1890	Denmark	2,172
1848	Egypt	463 ¹
1868	Egypt	618 ¹
1851	England and Wales	17,298
1861	England and Wales	20,066
1871	England and Wales	22,712
1891	England and Wales	29,003
1901	England and Wales	32,528
1911	England and Wales	36,070
1835	Iceland	56
1845	Iceland	57
1870	Iceland	60
1901	Ireland	4,459
1911	Ireland	4,390
1819	Germany	78 ⁴
1867	Germany	112 ⁴
1930	Mexico	1,655 ¹

(Continued on next page)

TABLE 1. Major Components of the North Atlantic Population Project (Continued)

Year	Country	Cases (in thousands)
1875	Norway	1,813
1910	Norway	2,294
1851	Scotland	2,889
1861	Scotland	3,062
1871	Scotland	3,360
1891	Scotland	4,026
1901	Scotland	4,472
1911	Scotland	4,761
1880	Sweden	4,566
1890	Sweden	4,785
1910	Sweden	5,522
1850	United States	23,192
Total cases		261,707

¹Sample: 5% or greater.²Sample: under 5%.³LDS approval pending.⁴Duchy of Mecklenberg-Schwerin only, 20% sample.

existing census samples to the complete enumerations to study migration and economic mobility. Their strategy was simple: they searched the complete-count data sets of England and Wales in 1881 and the United States in 1880 for men recorded in nearby census samples, matching on the characteristics that would not be expected to change over time: name, birth year, birthplace, and (in the United States) race. When more than a single match for a sampled person was found in the complete enumeration for 1880 or 1881, that individual was dropped from the linked set.

Analyses of these linked data sets have produced striking new findings that challenge established theories of social and economic change. Throughout the past century, theorists have argued that rising residential mobility has had a variety of adverse consequences, such as the loss of family cohesion, social dislocation, disrupted schooling, and health impairment (e.g., Parsons and Bales 1955; Litwak 1960; Astone and McLanahan 1994; Nicolopoulou-Stamati 2005). Increasing residential mobility is widely cited in policy debates, including debates on child immunization law (Centers for Disease Control and Prevention 2001), state inheritance taxes (Cooper 2006), child support enforcement (Department of Health and Human Services 2002), and grandparent visitation (Piekarsky 2004). We now know that residential mobility in the United States has actually been declining for well over a century and was far higher in the nineteenth century than it is today (Ferrie 2005; Goeken and Hall forthcoming; Hall and Ruggles 2004; Long 2005).¹ Accordingly, literature that explains social change as a consequence of rising mobility is simply based on a false premise.

The findings on occupational mobility are even more compelling. Social scientists on both sides of the Atlantic have long debated trends and differences in opportunities for economic mobility. Nineteenth-century observers such as Alexis de Tocqueville (1862) frequently remarked on the high levels of economic opportunity in the United States compared with Europe. Social historians writing in the 1960s and 1970s disagreed, pointing to rigid class divisions and limited potential for upward mobility (Thernstrom 1964, 1973; Katz, Doucet, and Stern 1982). The linked data sets made it possible to assess long-run trends in intergenerational mobility. It turns out that in the nineteenth century, the United States was a far more fluid society than was Britain, and rapid upward mobility was common. Over the course of the past 150 years, however, occupational mobility in the United States has declined dramatically while it has increased in Britain; today, there is little difference in economic mobility between the two countries (Long and Ferrie 2007).

During the past five years, we have developed new record-linkage and data-mining technology to expand and improve on the simple linked samples created by Long and Ferrie (Ruggles 2006; Pamarthy 2007; Goeken, Huynh, Lynch, and Vick 2011). Our strategies build on new research in probabilistic record linkage and machine-learning technology. Like Long and Ferrie, the NAPP linked samples rely exclusively on matching characteristics that should not change over time. The NAPP samples, however, do not require exact matches of characteristics; we link records on a probabilistic basis, allowing for imperfect correspondence of names and ages. Thus, for example, to compare names we use a string comparison algorithm that computes a similarity measure between 0.0 and 1.0 based on the number of common characters in two strings, the lengths of both strings, and the number of transpositions, accounting for the increased probability of typographical errors toward the end of words (Porter and Winkler 1997). We also use the New York State Identification and Intelligence System and Double-Metaphone phonetic name coding, which provide multiple encoded strings corresponding to variant pronunciations (Philips 2000; Lait and Randell 1993).

We then implement a machine-learning tool known as a Support Vector Machine (SVM) to classify each possible link (Vapnik 1998; Christianini and Shawe-Taylor 2000; Abe 2005). Based on the hand-linked training data, the SVM calculates a confidence score for every potential match; when one and no more than one potential match exceeds the threshold, we establish a link. Once we have established the full set of links, we weight the cases to represent the potentially linkable population with respect to age, sex, birthplace, whether related to head, occupational group, and size of place in the terminal year.

The linked NAPP data sets have several advantages over the linked sets produced by Long and Ferrie. First, our technology has allowed us to substantially reduce the rate of false links, which has significant implications for the rate of

geographic and social mobility (Goeken and Hall forthcoming). Second, the NAPP weighted samples are also significantly more representative of the linkable population than were the early linked samples. Finally, unlike Long and Ferrie (2007), we link women (as long as they do not marry and change their surname in the interval), and we provide users with full information on all members of the linked individuals' households.

We have released 28 sets of linked individuals and households for the United States and Norway, and we expect to complete linked sets for Britain and Iceland soon. The linked sets we have constructed to date match a census sample to a complete census enumeration. As we obtain multiple complete digital censuses for each country in coming years, we plan to link individuals through multiple censuses, thereby constructing longitudinal panels with multiple observations of each individual and family. To accomplish this, we will link successive pairs of censuses using the same general strategies we developed to link census pairs.

Linking multiple complete census enumerations will require significant innovations to accommodate the massive increase in the scale of processing. The complete enumerations contain 50–100 times the number of records as the samples we have been working with so far, and instead of just linking pairs of censuses, we will be linking up to seven censuses per country. This means the number of comparisons we must make could rise by a factor of approximately 500,000. Our current procedures already press at the limits of available computing capacity. We process the data on a high-performance computer at the Minnesota Supercomputing Institute for Advanced Computational Research (MSI). MSI has provided us with substantial amounts of supercomputing time at no cost to the project and will continue to do so in the next project phase. It would be impossible, however, to ramp up our supercomputer usage by 500,000-fold. We will therefore pursue a cluster of strategies to improve the computational efficiency of our procedures.

Spatiotemporal Analysis

NAPP is making a strategic contribution to demographic infrastructure by providing a baseline for study of changes in the demography and health of European and North American populations. In each country, NAPP provides the earliest national census microdata available. The NAPP database should not be regarded as a description of the population and the economy before it was affected by modern transformations. It is impossible to pinpoint a static time that precedes the modern world. In the period in which these censuses were taken, Europe and North America were already undergoing accelerating change. Indeed, the development of sophisticated statistical data in the eighteenth and nineteenth centuries can be seen as a byproduct of changes that were already underway. Nonetheless, these materials provide the

earliest detailed statistical picture of Western society ever available. In many cases, that picture was taken before most people's lives were transformed by urbanization, industrialization, and the shift to wage-labor employment.

The landscape of scientific research on the human population is shifting. It is no longer sufficient just to study the relationships among variables at a particular moment in time. Researchers around the world now recognize that to understand the large-scale processes that are transforming society, we must investigate long-term change. The goal of this project is to provide the kind of data that make such analysis possible.

A hint of the possibilities for comparative spatiotemporal analysis is provided by Ruggles's (2009, 2010) work on the Northwest European family system. For the past four decades, demographers have argued that historic Northwest Europe and North America had a unique weak family system characterized by neolocal marriage and nuclear family structure (e.g., Laslett 1972; Hajnal 1982; Hareven 1994). This interpretation has become a central component of theories of economic development and gender relations, and analysts argue that the distinctive northwest European family has important implications for present-day demographic behavior (Macfarlane 1978, 1986; Cain and McNicoll 1988; Reher 1998; Hartmann 2004; Thornton 2005). NAPP data allow us for the first time to place historical Northwest Europe and North America in the context of comparative data from around the world. The results provide no support for the theory of a distinctive northwest-European preference for nuclear families: with respect to residence with kin and in intergenerational families, nineteenth-century Europe and North America closely resemble other places with similar demographic composition (Ruggles 2009). When it comes to joint families in which an elder resides with multiple married offspring, however, there are systematic regional and temporal patterns that cannot be so easily explained by compositional factors (Ruggles 2010).

The nation-level Ruggles studies only suggests the power of these data; it does not exploit their full potential. NAPP will allow a new class of studies of the spatial organization of human activity, new measures of residential segregation, and innovative analyses of the effects of local context at multiple scales on individual behavior. Most historical studies of the impact of local conditions on individual behavior have focused on a single community or on a small number of communities (e.g., Anderson 1972; Hareven 1978; Katz 1975; Foster 1974; Modell 1978; Ruggles 1987; Janssens 1993). Because they are limited to single communities, such studies cannot generalize about the effects of context on behavior. NAPP provides data on the entire populations of tens of thousands of communities. That contextual data may be combined with the linked longitudinal panels to assess the ways in which individual and family transitions were conditioned by local, regional, and national opportunities and constraints.

Complete-count data are vital for spatial and multilevel analysis. Economic development across Europe and North America in the long nineteenth century was highly uneven; in some places, daily life changed little from the preceding centuries, and in others, it was transformed beyond recognition. This great variation allows us to assess the impact of local economic and demographic characteristics on individual behavior, thereby offering the potential for understanding the consequences of early industrial and commercial development.

The existing NAPP data are invaluable, but because they usually provide just a single detailed snapshot, they lack the power to reveal processes of change. The future availability of multiple complete-count cross-sections for the population of the North Atlantic world in the nineteenth and early twentieth centuries will open up vast new terrain in the study of industrialization, urbanization, demographic transition, and mass international population movements. NAPP will be a vital resource for studying the transformation of nuptiality, female labor-force participation, life-cycle service, mass education, indigenous populations, oldest-old populations, international migrations, and life-course transitions to adulthood. For each of these topics, the collection will provide information on the changing interrelationships among variables that cannot be obtained from any other source. The new data will permit much more subtle and comprehensive analysis that will help us to understand the sources of past geographic and economic mobility and to explain why mobility declined in the twentieth century. The database will enable us to begin to disentangle period and cohort changes in life-course processes, and will open exciting new opportunities for multilevel multivariate analyses in a key period of social and economic transition.

NAPP Articles in this Issue

The two articles that immediately follow offer a detailed explication of the techniques used for record linkage in NAPP. Ron Goeken, Lap Huynh, T. A. Lynch, and Rebecca Vick describe the process of creating the linked representative samples for the United States in the late nineteenth and early twentieth century. Rebecca Vick and Lap Huynh describe protocols for the name standardization for record linkage and quantifies the effects of standardization on linkage rates in the United States and Norway.

NAPP microdata offer unprecedented opportunities for spatial analysis. Sula Sarkar and Patricia Kelly Hall demonstrate the use of Geographic Information Systems (GIS) to explore the internal and international migration of adult male workers in nineteenth century Europe and North America, offering guidance to researchers who would like to incorporate NAPP maps into their analysis.

The final two articles discuss major new initiatives that are building on NAPP to make additional resources available to researchers. Gunnar Thorvaldsen describes an excit-

ing new initiative, the Longitudinal Population Register for Norway. The Norwegian NAPP data will serve as the base for building a longitudinal population register that can be linked with the existing population register for the period from 1964 to the present. This ambitious project will eventually link individuals across all available church and vital statistics records, providing the longest-running longitudinal demographic database ever constructed, and opening extraordinary opportunities for studying the demographic transformation of Norway over the past two centuries. John Logan, Jason Jindrich, Hyoungjin Shin, and Weiwei Zhang describe the Urban Transition Historical GIS Project, a path-breaking initiative to map the enumeration districts of 39 major cities in 1880. Used in conjunction with the NAPP 1880 census microdata, the project will allow unprecedented comparative study of the structure of nineteenth-century cities, and will permit the development of new measures of residential and socioeconomic segregation.

As the articles in this issue suggest, old census data are not of purely historical interest; they are essential tools for basic social research and policy analysis. Models and descriptions based on historical experience underlie both theories of past change and projections into the future. The NAPP data provide a unique laboratory for the study of economic and demographic processes. This kind of empirical foundation is essential for testing social and economic theory. The massive structural shifts of the long nineteenth century still resonate today. Revealing the causal mechanisms underlying these shifts is crucial for understanding the forces that are now shaping twenty-first century society.

NOTE

1. We can see the recent decline using survey data (e.g., Wolf and Longino 2005), but surveys cannot reveal the long-run trends.

REFERENCES

- Abe, S. 2005. *Support vector machines for pattern classification*. London: Springer-Verlag.
- Anderson, M. 1972. *Family structure in nineteenth century Lancashire*. Cambridge: Cambridge University Press.
- Astone, N. M., and S. S. McLanahan. 1994. Family structure, residential mobility, and school dropout: A research note. *Demography* 31:575–584.
- Cain, M., and G. McNicoll. 1988. Population growth and agrarian outcomes. In *Population, food and rural development*, ed. R. D. Lee, W. B. Arthur, A. C. Kelley, G. Rodgers, and T. N. Srinivasan, 101–117. Oxford, England: Clarendon.
- Centers for Disease Control and Prevention. 2001. *Morbidity and Mortality Weekly Report*, 50 (RR17): 1–17.
- Christiani, N., and J. Shawe-Taylor. 2000. *An introduction to support vector machines*. Cambridge: Cambridge University Press.
- Cooper, J. A. 2006. Interstate competition and state death taxes: A modern crisis in historical perspective. *Pepperdine Law Review* 3:835–81.
- de Tocqueville, A. 1862. *Democracy in America*. Trans. H. Reeve. Cambridge, MA: Sever and Francis.
- Department of Health and Human Services. 2002. *Essentials for attorneys in child support enforcement*. Washington, DC: Administration for Children and Families, Office of Child Support Enforcement.
- Ferrie, J. 2005. The end of American exceptionalism? Mobility in the U.S. since 1850. *Journal of Economic Perspectives* 19:199–215.
- Foster, J. O. 1974. *Class struggle and the industrial revolution: Early industrial capitalism in three English towns*. London: Weidenfeld and Nicolson.

- Goeken, R., and P. K. Hall. Forthcoming. New findings on internal migration using linked records. *Historical Methods*.
- Goeken, R., L. Huynh, T. A. Lynch, and R. Vick. 2011. New methods of census linking. *Historical Methods* 44:7–14.
- Hajnal, 1982. Two kinds of preindustrial household formation system. *Population and Development Review* 8:449–94.
- Hall, P. K., and S. Ruggles. 2004. “Restless in the midst of their prosperity”: New evidence of the internal migration patterns of Americans, 1850–1990. *Journal of American History* 91:829–46.
- Hareven, T. 1978. The dynamics of kin in an industrial community. *American Journal of Sociology* 84:S151–S182.
- Hartman, M. S. 2004. *The household and the making of history: A subversive view of the Western past*. Cambridge: Cambridge University Press.
- Janssens, A. 1993. *Family and social change: The household as a process in an industrializing community*. Cambridge: Cambridge University Press.
- Katz, M. B. 1975. *The people of Hamilton, Canada West: Family and class in a mid-nineteenth-century city*. Cambridge, MA: Harvard University Press.
- Katz, M. B., M. J. Doucet, and M. J. Stern. 1982. *The social organization of early industrial capitalism*. Cambridge, MA: Harvard University Press.
- Lait, A. J., and B. Randell. 1993. An assessment of name matching algorithms. Department Technical Report Series No. 550. Department of Computing Science, University of Newcastle upon Tyne, England. <http://homepages.cs.ncl.ac.uk/brian.randell/home.informal/Genealogy/NameMatching.pdf>
- Laslett, P. 1972. Introduction: The history of the family. In *Household and family in past time*, ed. P. Laslett and R. Wall, 1–73. Cambridge: Cambridge University Press.
- Litwak, E. 1960. Geographic mobility and extended family cohesion. *American Sociological Review* 25:385–94.
- Long, J. 2005. A tale of two labor markets: Intergenerational occupational mobility in Britain and the U.S. since 1850. Working Paper #W11253, National Bureau of Economic Research. <http://www.nber.org/papers/w11253>
- Long, J., and J. P. Ferrie. 2007. The path to convergence: Intergenerational occupational mobility in Britain and the U.S. in three eras. *Economic Journal* 117:C61–C71. <http://www3.interscience.wiley.com/cgi-bin/fulltext/117984644/PDFSTART>
- Macfarlane, A. 1978. *The origins of English individualism: The family, property and social transition*. New York: Cambridge University Press.
- . 1986. *Marriage and love in England: Modes of reproduction 1300–1800*. Oxford: Blackwell.
- Modell, J. 1978. Patterns of consumption, acculturation, and family income strategies in late nineteenth-century America. In *Family and Population in Nineteenth-Century America*, ed. T. K. Hareven and M. A. Vinovskis, 206–40. Princeton, NJ: Princeton University Press.
- Nicolopoulou-Stamati, P. 2005. Effects of mobility on health. In *Environmental health impacts of transport and mobility*, eds. P. Nicolopoulou-Stamati, L. Hens, and C. V. Howard, 1–7. Dordrecht, the Netherlands: Springer.
- Pamarchy, K. 2007. A machine learning framework for record linkage in census data. *Master of Science Report*, University of Minnesota, Minneapolis, MN, May.
- Parsons, T., and R. F. Bales. 1955. *Family, socialization and interaction process*. Glencoe, IL: Free Press.
- Philips, L. 2000. The double-metaphone search algorithm. *C/C++ User’s Journal* 18:38–43.
- Piekarski, C. 2004. The effect of an increasingly mobile society on Kentucky’s grandparent visitation statute: The ability of courts to enforce their orders. *Brandeis Law Journal* 42:693–710.
- Porter, E. H., and W. E. Winkler. 1997. Approximate string comparison and its effect on an advanced record linkage system. Census Bureau Research Report RR97/02. U.S. Bureau of the Census, Washington, DC. <http://www.fcsm.gov/working-papers/porter-winkler.pdf>
- Reher, D. S. 1998. Family ties in Western Europe: Persistent contrasts. *Population and Development Review* 24:203–34.
- Roberts, E., S. Ruggles, L. Dillon, O. Gardarsdottir, J. Oldervoll, G. Thorvaldsen, et al. 2003. The North Atlantic Population Project: An overview. *Historical Methods* 36:80–88.
- Ruggles, S. 1987. *Prolonged connections: The rise of the extended family in nineteenth-century England and America*. Madison: University of Wisconsin Press.
- . 2006. Linking historical censuses: A new approach. *History and Computing* 14:213–24. <http://www.hist.umn.edu/ruggles/Articles/linking.pdf>
- . 2009. Reconsidering the northwest European family system: Living arrangements of the aged in comparative historical perspective. *Population and Development Review* 35:249–73. <http://www3rd.interscience.wiley.com/cgi-bin/fulltext/122456480/PDFSTART>
- . 2010. Stem families and joint families in comparative historical perspective. *Population and Development Review* 36:563–77.
- Thernstrom, S. 1964. *Poverty and progress: Social mobility in a nineteenth century city*. Cambridge, MA: Harvard University Press.
- . 1973. *The other Bostonians: Poverty and progress in the American metropolis, 1880–1970*. Cambridge, MA: Harvard University Press.
- Thornton, A. 2005. *Reading history sideways: The fallacy and enduring impact of the developmental paradigm on family life*. Chicago: University of Chicago Press.
- Vapnik, V. N. 1998. *Statistical learning theory*. New York: Wiley Interscience.
- Wolf, D. A., and C. F. Longino. 2005. Our “increasingly mobile society?” The curious persistence of a false belief. *Gerontologist* 45:5–11.