Assessing Immigrant Naturalization: Longitudinal Research, Findings, and Challenges

Karen A. Woodrow-Lafield
University of Maryland, Maryland Population Research Center
0124N Cole Student Activities Building (#162), College Park, MD 20742
kwoodrow@umd.edu

Abstract

Linked administrative records on immigrants and naturalizations constitute a dataset for modeling naturalization over duration in lawful permanent residence and examining the role of covariates related to admission characteristics that may reflect social capital and human capital. This presentation reviews several recommendations and research programs in establishing such data for research. This paper describes creation of the first multi-cohort immigration-to-naturalization dataset for those individuals admitted for lawful permanent residence in 1978-1991 with naturalization outcomes observed through 1996. Several studies show the importance of origin country and admission characteristics in explaining differential timing of naturalization. Studies show that understanding naturalization outcomes involves not only considering observed heterogeneity on demographic, origin, and admission characteristics but also controlling for unobserved heterogeneity. The discussion reviews improvements in federal statistics on immigrant naturalization and makes recommendations for research on naturalization as an aspect of immigrant incorporation.

Introduction

Naturalization is the second major transition that foreign-born persons make that is sanctioned by the U.S. government with the first transition as their immigrant status, being admitted for lawful permanent residence. Becoming a U.S. citizen may be an indicator of immigrant adaptation, family integration, and settlement. The general, statutory requirements for U.S. naturalization are being at least eighteen years old, having continuous lawful permanent residence for five years, demonstrating English proficiency, having knowledge of the U.S. government and U.S. history, and possessing good moral character. Most studies on immigrant naturalization have been based on analyses of census or survey data that lacked the timing of naturalization and receiving lawful permanent residence and for which the universe was imprecise. The main uses of naturalization statistics have been descriptive.

This project sought to develop statistical models of the occurrence and timing of naturalization with existing administrative records by comprehensively linking for the first time existing administrative records on immigrants and naturalizations. Linked administrative records on immigrants and naturalizations constitute a dataset for modeling naturalization over duration in lawful permanent residence with covariates related to admission characteristics that may reflect social capital and human capital. These models were borrowed from biostatistics by demographers in the 1970s. Statistical methods for longitudinal analysis have been applied in studying marriage, divorce, fertility, labor force transitions, educational attainment, and retirement. These methods have not been applied fully for understanding immigrant adaptation because survey data on post-immigration experiences have been lacking. For studying naturalization, traditional cross-sectional sources, censuses and national surveys, are limited by universe discrepancies (inclusion of unauthorized residents and other ineligibles), and current without initial characteristics. Until introduction on the American Community Survey 2008, these surveys obtained current naturalization status without date of naturalization, but these surveys are still lacking immigration status and years of lawful permanent residence. The primary aim is to study the process of naturalizing in time, not just changes in citizenship status from one point in time to another.

Titled “Models of the Occurrence and Timing of Naturalization” and also designated “The Immigration-to-Naturalization Project,” the project was funded by NIH (HD R01-37279) over a four year period, and a competing continuation application will seek further funding. This project was responsive to a call for more research on U.S. immigration (Program
Announcement-95-036) that, specifically, encouraged "(3) maximum use of existing data on immigrants or the foreign born for analyses, and the linking of such data to administrative records ... to obtain a more accurate profile of immigrant experiences." The initial aims of this project on the occurrence and timing of naturalization were: (1) to utilize immigrant and naturalization administrative records as a unique multi-cohort retrospective data source on the immigration-to-naturalization transition for U.S. immigrants admitted for lawful permanent residence over 1978-1991 followed until 1996, as data were available in 1999 and prior to the 1992-effective changes in immigrant admissions under the Immigration Act of 1990 (IA1990); (2) to describe the timing and occurrence of immigrant naturalization by characteristics at admission, especially country of birth; (3) to develop statistical models of the occurrence of naturalizing with cohort covariates and characteristics at admission, and gain understanding of influences associated with greater likelihood of having naturalized or naturalizing more quickly for some immigrants than others. This grant application was initiated in response to recommendations for linking immigrant cohorts with naturalizations by experts (Levine, Hill, and Warren 1985). Those led to release of two public use datasets in the early 1990s by the former Immigration and Naturalization Service (INS), but progress was limited by difficulties with information technology. This project expanded the scope of cohort analyses of immigration and naturalization beyond the cohorts (1971, 1977, and 1982) for which statistics were previously shown in government reports (INS 1996) or in scholarly literature. Jasso and Rosenzweig (1986, 1990) found origin and admission criteria differentials in naturalization for 1971 immigrants after a decade, supporting micro-level themes of labor force attachment and family reunification incentive as explaining naturalization. For men and women, naturalization was higher for employment-sponsored immigrants and refugees than for those who came as family members of aliens or citizens.

These studies addressed a major gap in naturalization statistics and contribute to understanding naturalization as a process. This administrative records approach has typical limitations, and survey and mixed methods research provide rich perspectives on motivations, barriers, and patterns of naturalization and political participation. Political incorporation, naturalization of immigrants plus substantive citizenship, can be understood as processes of structured mobilization taking place within macro-level contexts, taking account of motivations and barriers (Bloemraad 2006). A longitudinal survey of immigrants, e.g., the New Immigrant Survey (Jasso, Massey, Rosenzweig, and Smith 2003), gives the capability for analyzing causal relationships between socioeconomic measures and naturalizing. However, these immigration-to-naturalization data allow analyses covering the entire population and subgroups of lawfully admitted immigrants in their transition to naturalization over time.

Prior Research and Background

The 1990 census was the first since the 1920 census to show more noncitizens than naturalized citizens, a ratio of three aliens for every two naturalized citizens, whereas the ratio had been three naturalized citizens for each alien from 1950-1980. With the passing of many European immigrants of earlier decades, higher lawful permanent admissions, higher admissions and stays of nonimmigrants, and the persistence of an unauthorized population contributed to a dramatically altered citizenship composition. The proportion naturalized was about 38-40 percent in 1990-2007, continuing a downward trend since 1950 when more than three-quarters of the foreign-born were naturalized citizens. The number of aliens increased to 3.5 million (1970), 7.0 million (1980) (up by 97 percent), 11.2 million (1990) (up by 69 percent), 18.6 million (2000) (up by 58 percent), 21.1 million (2004), 21.6 million (2005) (up by about 16 percent); 21.7 million (2006), and 22.2 million (2007), before it declined slightly to 21.3 million (2008). Throughout 1940 to 1980, naturalized citizens numbered 7 - 7.5 million and the 1990 number of 8 million was similar (7.9 million). The number of naturalized citizens increased to 12.5 million in 2000 and naturalized citizens have continued to increase in number--13.1 million (2004), 13.5 million (2005), 13.8 million (2006), 15.1 million (2007), and 15.5 million (2008). The recent increase number of naturalized citizens increased slightly (15.6 million or 42.1 percent) and reflects both changing status on citizenship and changing migration into the United States. Percentages of foreign-born residents naturalized in censuses and surveys underestimate the naturalization rates for lawful permanent residents. Current Population Survey estimates differed slightly from the ACS estimates for 2007 (38.1 million foreign-born, including 21.9 million noncitizens and 16.2 million citizens) and for 2008 (38.0 million foreign-born, including 21.6 million noncitizens and 16.3 million citizens) with significant differences in percent of naturalized citizens and percent of noncitizens over 2006-2008. Overall, the foreign-born population decreased in 2007-2008, but the foreign-born population was about the same over the 2006-2008 period.

With recognition of this gap in citizenship and the sizable share of unauthorized residents among noncitizens, proposals for providing a “pathway to citizenship” have appeared frequently in media discussions and immigration policy debates. Disparities in naturalization levels for the foreign-born population and origin groups and the question of differing propensities to naturalize are part of debates over which immigrant groups are assimilating in American society and which groups remain set apart within communities and the nation. The occurrence and timing of naturalization are as complex as
immigrants are heterogeneous. In the 1990 census, there were origin differentials for long-term residents with higher levels for Asians and Africans than Europeans, Canadians, or non-Mexican Latin Americans, and lower naturalization for Mexicans, and this divide between Mexicans and others was apparent in pre-1990 surveys (Woodrow 1992, 1995; Woodrow and Passel 1990). Certain immigrants, e.g., Asians, have appeared more likely to naturalize and to do so more quickly than others, especially Mexicans (Liang 1994; Jasso and Rosenzweig 1990), as annual naturalization statistics show Asians as having fewer years of residence than North Americans.

Preference categories and numerical limitations for family or labor migration were established in the Immigration and Nationality Act Amendments of 1965 and 1977. Individuals who wished to immigrate could be admitted for lawful permanent residence according to three major principles for setting priorities or preferences for lawful immigration: family unification, labor requirements, and humanitarian considerations. Later, the IA1990 introduced diversity criteria as a basis for admission, increased employment-sponsored visas, and slightly altered family-sponsored preference categories effective during fiscal year 1992.

The largest visa numbers were set aside for unification of families: second preference for spouses and unmarried sons and daughters of permanent resident aliens, fifth preference for siblings of U.S. citizens (at least 21 years of age) and their spouses and children, and first preference for unmarried sons and daughters of U.S. citizens and their children. The remaining amount (10 percent) was allocated equally among third preference for members of the professions of exceptional ability and their spouses and children, fourth preference for married sons and daughters of U.S. citizens and their spouses and children, and sixth preference for workers in skilled or unskilled occupations in which laborers are in short supply in the United States and their spouses and children. The nonpreference category of other qualified applicants was available according to residual visa numbers. The U.S. immigration system extends considerable capability to U.S. citizens, whether native-born or naturalized, to become sponsors of a family member as a legal immigrant, in that citizens may not only sponsor relatives under the limited preference categories but may also sponsor immediate relatives, that is, parents, spouses, and minor children, without limitations. The U.S. government is the sponsor for those immigrants admitted to meet labor needs and for humanitarian reasons. Demand levels are high for family reunification, so the right to sponsor family members is an incentive for naturalizing. Certain occupations and jobs are only available to citizens, as another incentive.

Understanding the determinants of naturalization is essential to explaining the integration of families and expansion of family networks across national boundaries. Given the role of family preferences in the U.S. legal immigration system, most legal immigration is either explicitly or implicitly family based as most legal immigrants are admitted with their nuclear family unit or joining nuclear family members, and they may also be reuniting with extended family members who previously migrated.

In addition to individuals and sponsors, the nation-state is relevant for studying migration. Capturing the influences of nations of origin and destination may not be as straightforward as identifying sets of social, economic, and political characteristics, but crucial elements lie within theoretical frameworks on international migration (Massey et al. 1998). As Jasso and Rosenzweig (1987) noted, the numbers and origins, histories, and demographic characteristics of immigrants are to some degree the consequence of the selection criteria, although unauthorized migrant flows may occur. Whereas initiating forces may help sustain a migration stream, both as to new migrants and minimizing return migrants, social networks play special roles in the perpetuation of international migration via cumulative causation and integration of families, households, and communities. Specialized institutions or special programs emerge for immigrant assistance, such as applying to naturalize or to sponsor family members. Understanding the determinants of naturalization is essential to explaining the multiplying effect of past immigrants for further immigration that integrates families and alters the dispersion of family networks across national boundaries.

Sociological studies using census data have long explored patterns of naturalization as a consequence of human capital, social capital, and origin country characteristics by distance, English-speaking, and political structures. Rare insights for naturalization experiences of lawfully admitted immigrants came from a sample drawn from the cohort admitted to permanent resident alien status during July 1970- June 1971 with naturalizations through February 1981 Jasso and Rosenzweig (1986, 1990). Immigrants of different visa categories were differentially likely to have naturalized in the first decade. Male immigrants as spouses of permanent residents or refugees had the highest probabilities of having naturalized among men, and women immigrants who were spouses of siblings of citizens or labor-certified immigrants had the highest among women. Women from the Western Hemisphere seemed to have lesser likelihood of naturalizing. Thus, immigrants with presumably fewer relatives in the United States, especially refugees, labor-certified immigrants, and women entering as spouses of principals of siblings, were more likely to have naturalized. Among men immigrating from Eastern Hemisphere
countries, either husbands of U.S. citizens and nonimmigrant adjustees or those men with employment-sponsorship visas, those men advancing occupationally over 1977-1990 were more likely to naturalize independently of any initial advantage from employment sponsorship (Jasso and Rosenzweig 1995).

Jasso and Rosenzweig conducted their research in the early 1980s, and improvements in official statistics about immigrants and immigration seemed imminent (Levine, Hill, and Warren 1985). Although the former INS Statistics Office, now the Office of Immigration Statistics (OIS) in the Department of Homeland Security (DHS), U.S. Census Bureau, and other agencies have made major advances as to federal statistics on immigration (Larson 2004), naturalization rates are not published on a regular basis. At initiation of this project, several editions of the Statistical Yearbook of the U.S. Immigration and Naturalization Service had shown naturalization rates by country of birth based on the 1977 and 1982 INS Immigrant-Naturalization Cohorts and the 1996 and 1997 Statistical Yearbooks showed details by class of admission and occupation. Compositional differences seemed to explain earlier and more rapid naturalization among the 1982 admittees as more than one-half were from Asia, including refugees from Vietnam, Laos, and Cambodia.

Immigrants are highly heterogeneous on individual characteristics, and contexts of sending countries vary temporally on internal and external dimensions. Differences among immigrants of various origins in the occurrence and timing of naturalization may be associated with demographic characteristics, admission criteria (family preference categories, employment preference categories, numerically unlimited immediate relatives provisions, or refugee or asylee categories) and other admission circumstances. There may also be differences among cohorts in naturalizing, net compositional effects, so it is the multi-cohort dataset permits causal modeling of intercohort change. Such studies that address questions of intercohort and intracohort variation in propensities to naturalize are as crucial as with other behaviors and especially timely given that naturalizing accelerated in the 1990s.

Data Description and Data Access

Data Description
For this project, the former INS Immigration Statistics Division provided microdata files in May 1999 of immigrant and naturalization administrative records with encrypted identifiers for alien numbers. The records did not include names, certainly the most sensitive information. These immigrant files excluded records for those immigrants who received lawful permanent residence after attaining legal status through the Immigration Reform and Control Act of 1986. In early 2001, it was evident that access to these data was not possible due to statutory confidentiality protections of the IRCA legislation. The immigrant microdata files were similar to the annual immigrant public use files for 1972-2000 available from the National Technical Information Service (NTIS) or the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan. Naturalization microdata files were considered public records for some time.

Including recent immigrants who represented nearly one-half of foreign-born residents in 1990, the provisional dataset was created in the summer of 2000. Over June 1999-February 2000, work was extensive to create consistent variables and analysis files. In dealing with admission categories and origin countries across years of immigrant data, the work of Greenwood and Ziel (1997) was helpful, as were data documentation and various editions of Statistical Yearbook.

The data have the typical limitations of administrative records. Some variables are missing for 1980-1983 because primary data were lost, and an alternative database, based on I-551 forms for alien registration receipt cards, included more limited detail. Until the mid-1980s, production of microdata files was poorly documented and counts from statistical systems in the 1970s were apparently reconciled with performance analysis systems by case duplication within preference category. The revised 1979 immigrants microdata set has fewer immigrant records than the 1979 Statistical Yearbook because 64,000 duplicate cases were deleted, apparently introducing little bias (Greenwood, McDowell and Trabka 1991; Greenwood and McDowell 1999). Naturalizations microdata files were sometimes incomplete, such as in 1994 when about 23,000 cases were omitted, primarily from San Francisco, Chicago, and Houston (Jasso and Rosenzweig 1990: 379) and in 1996 when, for approximately 100,000 naturalization cases, there was some missing data on marital status, occupation, naturalization provision, country of former allegiance, residence, and sex. Naturalization records are particularly incomplete for immigrants under 16 years at admission; they derive citizenship from parental naturalization, but the derivative citizenship applications may not be filed promptly or the data captured.

Data Access
The initial grant application was submitted in February 1996 with a letter affirming cooperation of the Immigration and Naturalization Service as data providing agency. Following scientific review, the project was approved to be awarded
funding, the principal investigator notified the INS Immigration Statistics Office, and the specific arrangements were
developed. Meeting security clearance requirements, the investigator was granted sworn status as an INS Expert, and two
other members of the research team were given permission for data management and analyses. A fourth member of the
research team was allowed to work only with data without identifiers. The principal investigator wrote guidelines for the
project data enclave and provided these to the Immigration Statistics Division. This arrangement for access to INS microdata
files was unique, but the arrangement resembled ones for other federal agencies, such as the Census Bureau, although the
arrangement was less structured by the INS. Principal investigators engaged with the New Immigrant Survey have also held
sworn status for access to confidential immigrant microdata, as have agency contractors.

By the summer of 1999, the data were provided to the principal investigator and housed in project offices in the Secure Data
Laboratory established at the Social Science Research Center at Mississippi State University. (There was a slight delay when
INS provided data on IBM cartridges that were incompatible with available drives, but INS then provided the data on Zip
disks.) Throughout the project, the principal investigator maintained personal communications with Michael D. Hoefer,
Director, Immigration Statistics Division, and suggested collaborative work. At a May 2000 meeting, the principal
investigator discussed data quality issues, availability of data for IRCA-amnestied immigrants, continuing sworn status as an
Expert, and data access for a competing continuation application in 2001. At a February 2001 meeting, the principal
investigator sought INS cooperation in developing public-use immigrant-naturalization cohort microdata files, pursuant to
OMB Circular A-110, discussed continuing sworn status as an Expert and obtaining further datasets, and discussed
possibilities for future grant applications, including the competing continuation application in 2001 or 2002. There were
additional personal inquiries (Spring 2004 and Spring 2006) about data access. INS staff had reduced the level of geographic
detail on immigrant records, amid agency concerns about highly detailed analyses by place or zipcode of residence and
country of origin.

This project does not include any immigrant microdata files after 1992 or any naturalization microdata files after 1996. As of
the 2001 fiscal year data, the INS ceased releasing public-use immigrant microdata files perhaps because some researchers
had analyzed immigrant statistics for small areas by demographic characteristics and country of origin. The Department of
Homeland Security was organized in 2003 and the 2001 Statistical Yearbook of the Immigration and Naturalization Service
published in February 2003 was the last in that series.

Unlinked Datasets
Of immigrants admitted 1978-1991, more than three-quarters (79.8 percent) initially intended to reside in one of the six key
states, especially California or New York, or in Virginia, Maryland, Massachusetts, or Pennsylvania, which was also found
for 1975-1996 (Woodrow-Lafield, Xu, Poch, and Kersen 1999). States with established foreign-born populations showed
higher percentages of immigrants under family-sponsored preferences, reflecting the presence of family networks and
strength of social ties, and examinations of percentages of immigrants under employment-sponsored preference categories
shows less pronounced differences among states (Woodrow-Lafield, Xu, Poch, and Kersen 1999a). These findings are

Another analysis of the immigrant microdata files focused on child immigrants by demographic characteristics, country of
origin, class of admission, and intended residence, with examination of class-of-admission and region-of-origin patterns
(Woodrow-Lafield, Xu, Poch, and Kersen 1999b). There were more males than females among children, although there were
more females in the 0-4 year category. Over time, the percent 11-17 years has increased, possibly as waiting times for
second preference visas have lengthened.

Creation of Provisional Dataset

Record Linkage
A major portion of the project was creation of the linked records dataset. The linking strategy relied primarily on matching
immigrant records and naturalization records with encrypted alien numbers. More intricate strategies based on the alien
number (or A-number) with sequential ordering and full names as used by government statisticians could not be explored
because the same information was not released to the principal investigator. Those statistical reports indicated that alien
number was more essential than name, and the percentage of 1977 Cohort records linked with only the alien number that
were found mismatched by using name was small (.5 to 1.0) (INS, undated documents). Multi-person linkages (parent-child,
husband-wife, sibling-sibling) were not feasible without actual, unencrypted alien numbers, although further work may
explore modeling. The linking strategy involved data blocking, consistency checking, and deletion of erroneous or suspect
cases.
First, the record linkage was tested for the 1982 cohort (Woodrow-Lafield, Xu, Kersen, and Poch 1999). Then, record linkage procedures were implemented with remaining cohorts (1978-1981 and 1983-1991) (Woodrow-Lafield, Xu, Kersen, and Poch 2001b, 2004). The discovery was made that some immigrant and naturalization records were duplicated across years and within year as previously found for the INS Immigrant-Naturalization Cohorts for 1977 (2,960 duplicate cases) and 1982 (12 cases). Of 9.1 million immigrant records over 1978-1992, we focused on 8,926,246 records for 1978-1991, and we found 98.2 percent acceptable as “original” records after designating 159,420 records as duplicate records to be excluded. Among 5.5 million naturalization records over 1978-1996, we focused on 5,538,975 records, and we found 98.8 percent acceptable as “original” records after designating 62,987 records as duplicate records to be excluded. Duplicated cases appeared to stem from statistical processing as, pursuant to the Act of October 28, 1977, refugees became eligible for adjustment to lawful permanent residence under the preference system. The overcount rate of actual immigrants in publicly released data for 1978-1991 is about 1.8 percent. The number of duplicated identification numbers was only 139,612 as some numbers were found more than once in a pool of 299,032 records containing duplicated numbers. Error in publicly released naturalization data for 1978-1996 is thus an overcount rate of naturalizations by about 1.2 percent. The number of duplicated identification numbers was only 61,914 in a pool of 124,901 records with duplicated numbers, as the remaining records were third or higher-order incidences. To summarize in a different way, about 1.8 percent of immigrant counts and 1.2 percent of naturalization counts, as released, should not be counted.

After record linkage of 1978-1991 immigrant and 1978-1996 naturalization records, the linked dataset included 2,805,599 immigrant records (31.5 percent) matched with a naturalization record and 6,120,647 immigrant records (68.5 percent) remained as not yet corresponding with a naturalization record. The unmatched immigrant records included more than three million child immigrants possibly eligible to have derived or to yet derive citizenship upon parental naturalization or eligible to naturalize upon reaching 18 years. Most unmatched naturalization records were for immigrants entered before 1978.

Although the linked dataset was considered provisional pending official resolution of duplicates, those issues were never resolved by INS statisticians. The principal investigator discussed data quality issues and provided listing of records in question. However, the dataset is a substantial resource about naturalization of recent immigrants to the United States. Creation of microdata files for public use was deferred due to the provisional nature of the database given data quality issues on immigrants and naturalizations 1978-1996 and the necessity of cooperative arrangements with the Department of Homeland Security. As mentioned earlier, DHS has been evaluating various policies for publications, microdata release, and tabulations. The linked dataset raises particular issues also as to the appropriate schedule for releasing naturalization cohort data, i.e., after ten years, fifteen years, twenty years, or more than twenty years as resident.

Data Summary
Overall, of 8.9 million immigrants, 2.8 million (or 31.5 percent) were naturalized by 1997 (Woodrow-Lafield, Xu, Kersen, and Poch 2004, 2001b). After excluding children whose naturalization is poorly measured, 1.8 million adult immigrants (or 35.6 percent) were naturalized by 1997. Slightly less than one-half of the earliest cohort was naturalized after 19 years. More than one-half of some cohorts have probably naturalized by now.

Descriptive Studies
Descriptive analyses showed the beginning of the escalation in naturalizations in 1995-1996. These also illustrate the typical progression of certain immigrants toward naturalizing soon after reaching eligibility. For male immigrants, female immigrants, and total immigrants for each country of origin, charts were created to illustrate the observed timing of naturalization, showing cumulative naturalization by calendar year for each cohort (Woodrow-Lafield, Xu, Kersen, and Poch 2000a). Three-dimensional charts were also created for each region of origin showing cumulative naturalization rates by entry cohort for duration years (Woodrow-Lafield, Xu, Kersen, and Poch 2000b). Descriptive analyses showed the beginning of the escalation in naturalizations in 1995-1996, also noted in Statistical Yearbooks of the Immigration and Naturalization Service.

Modeling Strategies
Continuous-Time Proportional Hazards Models
With this comprehensive data on naturalization outcomes, statistical models may be applied for understanding the transition to naturalized citizenship and explaining how quickly or slowly immigrants become naturalized. Using continuous time Cox proportional hazards with a standard assumption on the hazard function, statistical models focused on the timing and occurrence of naturalization and utilized several covariates: demographic characteristics (age at admission, sex, country of
birth, marital status at admission), admission circumstances (class of admission, prior nonimmigrant experience), and year of admission. Visa category indirectly reflects human capital levels and potential social capital. The immigration-to-naturalization process is a single-state non-repeatable event process. Modeling duration to naturalization is advantageous in incorporating fixed characteristics at time of migration and the passage of time itself. To the extent that the functional form of the hazard of naturalizing differs among subgroups, covariate effects may be misestimated. Either estimating determinants of the cumulative probability of having naturalized as of ten years (Jasso and Rosenzweig 1986) or within broad duration intervals (Rytina 2008) would be a more restricted analysis. Although some immigrants may not have naturalized because they had left the United States, these models did not incorporate emigration as a competing outcome because emigration is poorly assessed (Woodrow-Lafield 1996). The numbers of immigrants from major sending countries were sufficient for examining detailed admission covariates with origin-gender-specific models on timing of naturalization, with controls for age, cohort, and other covariates.

Project Studies. These Cox regression hazards models have been calculated for region of birth groups and more than twenty country-of-birth groups (Woodrow-Lafield et al. 2000a, b, 2001a, b, 2004, 2005; Woodrow-Lafield et al. 2001, b, 2003; Woodrow-Lafield 2001). These models were estimated for exploring the role of gender and admission criteria in explaining propensities in naturalizing, specifically, (1) for immigrant cohorts 1978-1987 from the ten leading countries of birth (China, India, Korea, the Philippines, Vietnam, Jamaica, Mexico, Cuba, the Dominican Republic, and Colombia); (2) for immigrant cohorts 1978-1991 from the major regions of origin (Europe, Eastern Asia, South-Eastern Asia, South-Central and Western Asia, Africa, Oceania, Central and North America, Caribbean, and South America); (3) for immigrant cohorts 1978-1991 from Canada and European countries (Great Britain, Poland, Ireland, the former Soviet Union, France, Germany, Greece, Italy, Portugal, and Romania); and (4) for male immigrant cohorts 1983-1991 from Mexico, China, and India, adding occupational variables.

In initial Cox proportional hazards models for Mexicans admitted in 1982 and 1987, we examined covariates in relation to the timing and occurrence of naturalization (Woodrow-Lafield, Xu, Kersen, and Poch 1999). Key findings were: 1) Mexican women were more likely to naturalize than their male counterparts; 2) recent Mexican immigrants were more likely to naturalize and to naturalize faster than earlier immigrants; and 3) the propensity to naturalize depended considerably on labor force attachment and gender. These results enrich understanding of naturalization for the Mexican-born population in the United States.

Key Findings. Considering the Cox regression hazards models for immigrants from the ten leading countries (Woodrow-Lafield, Xu, Kersen, and Poch 2004), certainly, younger immigrants are more likely to naturalize, as are married immigrants. Past nonimmigrant experience is associated with higher propensity in naturalizing, although this effect is probably underestimated. Among immigrants from major origin countries and regions, some origin groups show lower propensities in naturalizing than other groups after controlling for demographic characteristics and admission circumstances.

The type of admission, employment-sponsored, family-sponsored, or immediate relatives provisions, is associated with propensities in naturalizing. Those immigrants most advantaged or most committed in naturalizing quickly are employment-sponsored immigrants, professionals and skilled workers and their spouses. They may have greater interest in enhanced employment opportunities and their educational levels may facilitate the process or they may need to sponsor other family members given they are not known to have relatives residing in the United States.

Family reunification may be an incentive for naturalizing for immigrants having the least number of relatives already in the U.S. Spouses of citizens, spouses of permanent residents, spouses of siblings of citizens, and spouses of sons and daughters of citizens naturalized faster than some other family-sponsored immigrants. Parents of citizens were less likely to naturalize even after controlling for demographic and origin characteristics, suggesting inability or disinterest. Those immigrants joining relatives with established U.S. residence may draw upon these social capital resources for English acquisition and for facilitating naturalization, and these relatives may also be helpful in understanding citizenship as conveying ability to bring parents.

Filipinos, Chinese, and Vietnamese were naturalizing sooner than Indians, Koreans, Cubans, Colombians, Jamaicans, Dominicans, or Mexicans (Woodrow-Lafield, Xu, Kersen, and Poch 2000a, 2004). Other studies have shown Mexican immigrants as slower to naturalize, consistent with border or temporary migration explanations.

Differences in the timing and occurrence of naturalization were also found by gender for certain origin groups (Woodrow-Lafield, Xu, Kersen, and Poch 2000a). These differentials by gender and origin suggest cultural contexts, demographic
diversity, and social capital are crucial for explaining naturalization. Women from Cuba, the Dominican Republic, Jamaica, Colombia, and Mexico were likely to naturalize more quickly than men, but the reverse was the case for the Philippines, Vietnam, India, and Korea. Chinese men and women seemed at equal chances. When considering visa classes of admission, certain gender differentials are significant. Some results suggest marital homogamy on skills and education and positive selection for immigration and naturalization. Also, some women might be naturalizing not only because they have more human capital and can fulfill the requirements but also because they seek self-protection to insure their children’s current and future wellbeing in a diffuse manner, if not in the specific sense of public benefits access.

More recent cohorts seem to be naturalizing sooner than earlier cohorts implying they might reach higher naturalization levels (Woodrow-Lafield et al. 2004). These data precede passage of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA), also known as the 1996 Welfare Reform Act, and the 1996 Illegal Immigration Reform and Immigrant Responsibility Act (IIRIRA) that may have affected naturalization in the later 1990s and after 2000. Whether this is merely reflecting period influences, diminishing barriers or deterrents, greater incentives, or more facilitating networks or organizations has not been investigated.

Information on immigrant skill levels, or occupational backgrounds, are limited in these data, but initial evaluation indicates it helps explain differences in naturalization. For selected origin-cohort-specific immigrants, reported occupation at admission is useful in modeling duration to naturalization and more economically successful immigrants naturalize more quickly (Woodrow-Lafield, Xu, Poch, and Kersen 2001a).

Continuous-Time Parametric Regression Models
A number of alternative continuous-time hazards models were calculated based on varying distributions for the underlying hazard or survival function over duration to naturalization (the exponential, Weibull, Gompertz, log-normal, log-logistic, and generalized gamma distribution). This work is promising as to the value of selecting the best model according to the form of the underlying hazard or survival function over duration and incorporating statistical corrections for unobserved heterogeneity.

Project Studies. These models were developed for selected cohorts (1978, 1985, and 1990) of six origins, specifically, Cuba, El Salvador, Mexico, China, India, and the Philippines (Woodrow-Lafield, Xu, Poch, and Kersen 2001b, 2003). These analyses also compared models with and without correction for unobserved heterogeneity. Because this is extensive model testing, these studies used a more restricted scheme for visa class of admission: family-sponsored, employment-sponsored, parents or spouses as immediate relatives, and refugees or asylees. For each origin-cohort specific population, twelve models were estimated, so that 216 models were examined with three cohorts of six countries of origin. Similar analyses were done for all remaining Mexican and Chinese cohorts (totaling 168 models each) (Woodrow-Lafield, Xu, Poch, and Kersen 2010), for all Dominican cohorts (Woodrow-Lafield and Poch 2003), and for all Chinese cohorts (Woodrow-Lafield and Poch 2006) settling in New York.

Key Findings. These confirmed the hazard of naturalizing varies markedly over duration for origin groups. Asian and Latino immigrants progress to naturalization in contrasting temporal patterns. Asian immigrants include many who are early deciders to naturalize and their probabilities of naturalizing become lower, and the underlying hazard function form seems to be the log-logistic form, an inverted U. The Latino immigrants begin naturalizing more slowly and gradually do so more extensively in the second decade and perhaps longer, and the underlying hazard function form seems to be the Gompertz form, concave. This is relevant for managing programs for English language instruction and civics education which has been difficult in California with increasing numbers of immigrants seeking to enroll (Gonzalez 2007). For some groups, observed characteristics are adequate for measuring propensity in naturalizing, suggesting that admission categories sort some immigrant groups as to assimilation prospects and naturalization. Models controlling for unobserved heterogeneity were preferable for Asians and may capture the effect of initial and post-immigration characteristics promoting early naturalization, including socioeconomic assimilation.

Future Research
These studies indicate the pace of naturalizing may be quickening, which is consistent with other evidence. More than ten million foreign-born residents had become naturalized during 1991-2008, about the same as in official naturalization statistics, without allowing for those who emigrated, died, or were not covered in the survey. Naturalization was least likely for the most recent post-2000 arrivals as only 10.2 percent were naturalized. About two in every five foreign-born persons who entered in 1990 to 1999 had naturalized by 2008. Nearly two-thirds (59.6 percent) of 1980 to 1989 entrants had become
citizens after two or more decades of residence. Among long-time residents since before 1980, three quarters (78.6 percent) are citizens.

Multiple projects may follow for further statistical modeling. With continuous time modeling, the emphasis is on cohort time and whether certain cohorts experience the transition more quickly than others. Future work is likely to include discrete time hazards modeling to more fully assess duration, period, and cohort effects. Due to provisions in the IIRIRA and the PRWORA, immigrants admitted after 1996 have more incentives to naturalize, for access to public benefits, to relieve their sponsoring family members of obligatory financial support, and for self-protection against deportation. Discrete analysis would also allow treatment of origin country shifts on dual citizenship or nationality.

This project was initiated when the former INS had made little progress in linking immigrant and naturalization records. The DHS Office of Immigration Statistics has undertaken more statistical analyses on naturalizations internally, beginning with a comparison of naturalization outcomes for the 1989, 1990, and 1991 cohorts with IRCA-amnestied immigrants (Rytina 2002). Releases include annual reports on naturalizations with demographic characteristics and country of origin and place of residence (Rytina and Saeger 2005; Simanski and Rytina 2006; Simanski 2007; Lee and Rytina 2009), geographic distributions of naturalizations (Simanski 2006), and trends by provision of law (Simanski 2007). One fact sheet addressed the difference between naturalization rates based on the American Community Survey and the 1975 and 1995 immigrant cohorts (Cornwell 2006). Two reports presented trends in cumulative naturalization rates after ten years for 1973-1995 and 1973-1999 immigrant cohorts (Baker 2007, 2009). One study is a multinomial logit analysis focused on naturalization for the 1977 and 1982 cohorts (Rytina 2008). Federal statistics on immigrant naturalization have improved. Further advances are needed and statistical modeling might be more likely to be undertaken by academic researchers as was the research project described here. Statistical policy on immigrant naturalization as an aspect of immigrant incorporation should be shaped.

The previous agreement for data access through special sworn status is unlikely to be possible with the Department of Homeland Security. The DHS Information Office concluded “that the Privacy Act prohibits us from disclosing individual records (even when stripped of personal identifiers) because of the possibility of identification” (personal communication, Michael D. Hoefer, 2004). Thus, in the current environment external researchers have extremely limited access to post-2000 immigrant and naturalization microdata files and such access may only be possible through a federally funded research data center or possibly through another type of restricted use agreement. Augmenting and updating the project’s current linked records dataset with post-1992 immigrant and post-1996 naturalization files has not been possible. Additionally, the initial aim of data sharing was deferred until a later project phase, partially due to the altered statistical policy regarding public use data releases. Certainly, a comprehensive immigration and naturalization data archive updated with current records would be very useful for immigration policy research.

The privacy office of the Department of Homeland Security recently faced criticisms from the Privacy Coalition for enabling government surveillance and intelligence programs. DHS, as was the INS previously, is a federal agency with missions largely oriented to enforcement rather than statistics-gathering. Within that context, statistics are for purposes of performance evaluation rather than as social indicators, but the Office of Immigration Statistics has a more elevated position in the DHS organizational chart than in the INS. Looking at the broader matrix of federal statistical infrastructure, developing a comprehensive program on naturalization research might involve advisory groups from professional organizations, interagency statistical expertise, and identifying similar data organization and statistical modeling from such agencies as Education, the Census Bureau, and Justice, specifically, the Bureau of Justice Statistics.

Summary

Beyond analyses involving either administrative records or surveys, more research is needed about immigrant decisionmaking to reside in the United States, to become naturalized citizens, and in navigating the application process as well as taking English classes and learning U.S. civics and history.

This project has demonstrated that existing large-scale data may usefully contribute longitudinal analyses that more accurately evaluate the transition to citizenship for the new immigration of Asian, Caribbean, African, and Central and South American origins. Assimilation processes for these new immigrants are set in different contexts from the historical ones for European immigrants—linguistic concentration, geographic clustering, and temporal continuity (Massey 1995). The sheer continuity of the "new regime" of immigration means that immigrant and ethnic groups are "augmented continuously with new arrivals from abroad" and as Massey (1995:644-645) further posits, ethnicity will become fragmented "along the lines of class, generation, and ancestry."
Gender patterns of immigration and immigrant incorporation for the United States are studied relatively little. Gender effects for naturalizing differ for origin groups and categories of admission in the immigration preference system. The multi-cohort approach allows examination of how the gender effect within country varies among cohorts. Further research is needed to fully investigate origin-specific variation by gender and admission criteria and to more fully understand the consequences of immigration admission policies for naturalization patterns and immigrant assimilation in America.

Given the volume of immigration and heterogeneity of immigration and the generalized pattern of unanticipated consequences of changes to immigration laws and policies, this work advances immigration studies for interpreting the significance of naturalization in America. The project is just beginning to disentangle influences of gender, origin, and admission criteria for timing of naturalization in the 1980s and 1990s. With multiple cohorts spanning multiple periods, causal modeling is possible of period influences, in addition to individual level and inter-cohort, for change in citizenship status and timing of naturalization. This immigration-to-naturalization multi-cohort dataset offers the promise of being a key data source with which to reach greater understanding of the macro-linkages and micro-linkages leading immigrants to become permanent settlers in the United States.

The surge in immigrants applying to naturalize since 1992 was expected because IRCA-legalized immigrants were becoming eligible, certain agency procedures and priorities encouraged applications (Green Card Replacement Program and Citizenship USA 1995), and immigrants may have sought to protect themselves and public benefits access after passage of anti-immigrant legislation in California in 1994 (Proposition 187) and the 1996 Welfare Reform Act (Espenshade, Baraka, and Huber 1997). The PRWORA included language to bar noncitizen immigrants from several federal and state benefits. A less immediate effect stemmed from the Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (IIRIRA) requiring that persons sponsoring family members for immigrant visas must meet requirements for the affidavit of support: their income must be at least 125 percent of the official poverty threshold applicable for their family plus the sponsored family members and that this financial responsibility is effective for ten years or until the sponsored family members became citizens.

Peak numbers of immigrants became eligible to naturalize, unprecedented numbers applied to naturalize, and naturalization approvals were at historic highs over 1995-1997. Asian countries had accounted for the greatest number of naturalizations for two decades, but, in 1996, North America became the leading region of birth among naturalizing immigrants with Mexico as the most prominent origin country. By 2000, the majority of foreign-born persons here for two decades were naturalized citizens, with 86 percent of Asian origin persons naturalized, and similar levels for Europeans (81 percent), Africans (80 percent), Caribbeans (80 percent), and South Americans (77 percent). Among long-resident Mexicans in 2000, 52 percent were naturalized. Mexicans, the single largest origin group, may now be at the threshold of embracing citizenship with the Mexican Nationality Act of 1998 that extended non-loss of Mexican nationality, including all rights of Mexican citizenship, to Mexicans who opted to obtain U.S. citizenship. With a history of naturalization as a “long gray welcome” (North 1987), the bipartisan Commission on Immigration Reform (1997) recommended reforms to increase the efficiency and integrity of the naturalization process. More than two million applications were delayed in processing over 1996-97. The USCIS implemented a new naturalization test in October 2008 with goals of standardization, fairness, and to encourage civic learning and patriotism. The new exam includes more civics-based vocabulary, and master list of questions and answers is available online along with study guides, as well as reading and writing vocabulary lists. Naturalization applications have increased as immigrants anticipated voting in presidential elections in 1996, 2000, 2004, and 2008, with subsequent delays in processing. A fee increase announced for the next year may also prompt more aliens to submit applications before it is effective.

The naturalization gap by origin is one of several prompting discussion about assimilability of contemporary immigrants and immigration policy, part of a backlash against globalization that seeks to halt social and economic progress of the late 20th century (Massey 2004). Going beyond cross-sectional analyses which give limited insights as to immigrant socioeconomic success and naturalization, immigrants are reshaping America so that assimilation is a two-way street (Massey 2004). Immigrants may be linguistically concentrated, but their children acquire English ability, as do adults over time, and original languages are unlikely to be retained over time (Rumbaut, Massey, and Bean 2006), and citizenship in original and destination nation may become more commonly held for U.S. citizens.

An immigration-to-naturalization multi-cohort dataset augmented with more recent naturalizations and immigrant cohorts would constitute a useful longitudinal data archive for multidisciplinary researchers in future decades. Although this immigration-to-naturalization multi-cohort dataset lacks socioeconomic characteristics, it is rich with details on admission
characteristics, country of birth, and timing of naturalization.

References


Gonzalez, Arturo. 2007. *California’s Commitment to English Language Earners: Caught Between Funding and Need*. Public Policy Institute of California.


Naturalization: Mexico, China, and India,” Presented at the 2001 annual meeting of the Population Association of America, Washington, D.C.

