THE MEASUREMENT OF INTERNATIONAL MIGRATION IN U.S. CENSUS SURVEYS: AN EXAMINATION OF CURRENT POPULATION SURVEY DATA MATCHED TO SOCIAL SECURITY ADMINISTRATION RECORDS

James D. Bachmeier[1]
Jennifer Van Hook[2]
Bert M. Kestenbaum[3]

ABSTRACT
This paper evaluates the measurement of international migration in U.S. Census surveys using data from the Current Population Survey linked to the Social Security Administration database. Results indicate that match rates vary as one would expect given existing knowledge about the unauthorized foreign-born population: match rates are significantly lower among Mexicans, recently arrived migrants, those of prime working age, and those lacking other survey indicators of legal residency status. The examination also finds variation in the consistency of survey responses with information in the administrative records across immigration related survey items. Specifically, among matched records, over 90 percent of survey responses for country or region of birth match the administrative data. However, consistent with previous research, there is much less consistency between the survey and administrative data with respect to year of arrival in the United States. The paper concludes with a discussion of the implications of the findings and avenues for future research on the U.S. foreign-born population using Census survey data matched to administrative records.

[1] Department of Sociology, Temple University
[2] Population Research Institute, Pennsylvania State University
[3] Social Security Administration

THE MEASUREMENT OF INTERNATIONAL MIGRATION IN U.S. CENSUS SURVEYS: AN EXAMINATION OF CURRENT POPULATION SURVEY DATA MATCHED TO SOCIAL SECURITY ADMINISTRATION RECORDS

INTRODUCTION

Large-scale immigration over the past five decades has had dramatic demographic impacts on the United States (Bean and Stevens 2003; Portes and Rumbaut 2006). With the movement of the nation's large "Baby Boom" cohort into retirement over the course of the next decade and a half, immigration's impact on the economic and social fabric of the country is likely to be even more significant (Myers 2007), and research on the foreign-born population and their children will carry ever more social and public policy significance.

One concern among researchers, given the growing importance of immigration research, is that adequate data for studying the U.S. foreign-born population remain extremely scarce (Bachmeier, Van Hook, and Bean 2014; Clark and King 2008; Massey and Bartley 2005; Van Hook et al. 2015). The most commonly used data sources for immigration research are large population surveys sponsored by the U.S. Bureau of the Census, including the American Community Survey (ACS), Current Population Survey (CPS), and the National Health Interview Survey (NHIS). These surveys are valuable sources of information on the foreign-born population due to the fact their very large sample sizes allow analysts to examine numerous sub-populations of immigrants and employ multivariate statistical techniques.

These Census surveys, however, are limited in that they include very few questions related to the immigration experience – typically, country of birth, year of arrival, citizenship, and English language proficiency[1]. Moreover, concerns have been raised about the quality of the immigration-

[1] The ACS includes country of birth, year of immigration, citizenship and English language proficiency. The CPS includes the same immigration related variables as the ACS except for English language proficiency, but unlike the ACS the CPS also includes questions about respondents' parents' place of birth. The NHIS includes country of birth, year of immigration, citizenship and English language proficiency. The level of detail in these variables provided in public-use data files varies across Census surveys. For example, the ACS and CPS

related data collected in Census surveys (Redstone and Massey 2004), especially in the post-9/11 immigration context, which has seen large increases in the enforcement of unauthorized migration at Federal, state, and local levels (Lofstrom, Bohn, and Raphael 2011; Meissner et al. 2013). Perhaps the most significant limitation in Census survey data on the foreign-born population is the lack of information on immigrants' legal status.

This paper addresses these interrelated concerns – about the accuracy of immigration-related information in Census surveys and the lack of legal status indicators – by analyzing data from the Current Population Survey matched with administrative records maintained by the Social Security Administration. Census survey data are regularly matched to government administrative database records, and these linked data are a valuable, but underutilized, source of information on the U.S. foreign-born population. Specifically, as detailed below, linked data are likely to facilitate crucial research on the unauthorized immigrant population that cannot be carried out using public-release samples, which lack indicators of immigrants' legal residency status. Also, linked data allow us to examine the accuracy of immigration-related survey questions.

BACKGROUND

CPS-SSA Match Rates

As discussed in greater detail below, Census survey data records are routinely matched to government administrative databases, such as the SSA. One important way we use these data is to examine the percentage of foreign-born population in Census surveys that is successfully matched to administrative records. This is helpful because it might facilitate future research on the unauthorized foreign-born population. To explain, large-scale Census surveys typically allow analysts to distinguish between immigrants who are naturalized citizens and those who are not. Among the latter, however, it is impossible to separate unauthorized immigrants from those in the country under legal auspices. While some analysts have used other indicators included in Census

---

public-use files provide detailed country of birth codes, while the NHIS only releases broad region of birth codes in their public-use samples.

surveys, such as military service, welfare receipt, and employment in government, Van Hook et al. (forthcoming) have shown that such indirect indicators fall far short of distinguishing between legally resident and unauthorized residents. Van Hook et al. (Van Hook et al. 2015) also demonstrate that the introduction of additional indicators of legal status can significantly improve legal status imputation methods. The presence or absence of a successful record linkage could serve as a potentially important such indicator, if there were sufficient evidence that non-citizen respondents in Census surveys that cannot be matched to administrative data are very likely to be unauthorized migrants.

One caveat to this line of reasoning, however, is that immigrants may not be matched to administrative data for reasons other than their residency status. Some legal migrants may never seek employment in the U.S., for example, and therefore never apply for a Social Security Number. Moreover, record linkages may fail because other information used in the matching algorithm, such as a person's name given on the survey, does not match the information in the administrative database.

This caveat notwithstanding, one could conclude with greater certainty that a successful match is an indication of legal residence status if expected patterns emerged in an examination of match rates. Specifically, we would expect match rates to be the lowest in segments of the population known to have large numbers of unauthorized immigrants, such as those born in Mexico and Central America, those arriving in the U.S. relatively recently, those of prime working age, and among those lacking other indirect indicators of legal status in survey data (Passel and Cohn 2010).

Consistency of Survey and Administrative Information

Another way we use the linked SSA data is to evaluate the consistency of survey with administrative information on immigration items. Information about immigrants' place of birth, year of arrival, and citizenship provide the scaffolding for many, if not most, analyses about immigrants and their impacts on American society. For example, accurate information on place of

birth makes it possible to assess the size and growth of national origin groups. This is important because national origin groups vary considerably with respect to their reasons for migrating, the resources they arrive with, legal rights, and social service needs. Year of arrival is important for assessing the degree to which immigrant groups change with duration of U.S. residence, and thus serves as a crucial indicator in analyses about the pace and direction of immigrant integration. Citizenship is important for policy evaluation because naturalization is both an important signal of social and political incorporation and a gateway to many U.S. jobs, rights, and responsibilities.

Given the importance of these basic characteristics for research on immigrants, it is important to assess the validity of place of birth, year of arrival, and citizenship reporting in surveys, as we do here. In the United States, data on these items largely come from Census Bureau surveys, such as the Current Population Survey (CPS), the long form of the decennial Census (2000 and earlier), and the American Community Survey (ACS). Prior research evaluating the quality of immigration survey items have generally used demographic methods that compare aggregated estimates of the number of immigrants by year of arrival or citizenship based on survey data with estimates based on administrative data. For example, Passel and Clark (1997) and Van Hook and Bachmeier (Van Hook and Bachmeier 2013), compared the number of naturalized citizens reported in Census or ACS data with estimates based on the number of naturalizations recorded by the Office of Immigration Statistics. Both studies found that the number of naturalized citizens is over-estimated in Census/ACS data, possibly because some non-citizens misreport as citizens.

This aggregate approach has been helpful for assessing the overall consistency between administrative and survey data, but aggregated comparisons can be misleading. For example, differences could be attributable to differences in the coverage between survey and administrative data rather than reporting differences. To better assess consistency in reporting, we use linked data to compare a person's responses to CPS survey items to information about them in the Social Security Administration's NUMIDENT file. This individual-level comparison allows us to better

assess the consistency in reporting on place of birth, year of arrival, and citizenship with similar information in administrative data.

It is important to note some important limitations of our approach. First, our analyses are based on the linked CPS-NUMIDENT file for 2007, 2008, and 2009. This file was created by linking March CPS respondents with their records in SSA's NUMIDENT file, which contains demographic and programmatic information about those registered in SSA's record system (we describe this file and linkage methodology further below). Thus, we are able to assess reporting consistency only among CPS respondents who were successfully linked to the NUMIDENT file. Those without legal U.S. residency were very probably not linked. Therefore, our assessments of reporting consistency are most likely limited to legally-resident foreign-born. Nevertheless, we still examine the characteristics of those without a NUMIDENT linkage to assess the plausibility of their responses on citizenship. Since unlinked respondents are unlikely to be legally resident and therefore ineligible for naturalization, very few (if any) should self-report as a naturalized citizen. If large numbers report as naturalized citizens, this may signify reporting error for this data item.

A second limitation is that the immigration data in the NUMIDENT file are not exactly comparable to the CPS data. The most comparable data item is place of birth; consistent reporters will give the same response in both the CPS and NUMIDENT. However, year of arrival responses could plausibly differ for respondents for whom year of arrival is ambiguous. They may have entered and left the United States multiple times, or spent time in the U.S. before attaining legal status. Additionally, the data on citizenship in the NUMIDENT file most likely reflects citizenship at the time the person first entered the Social Security system; if a person subsequently naturalized, the NUMIDENT record may or may not have been updated to reflect this change in status. Therefore, we employ a looser definition of consistency for citizenship. If the person is reported as a citizen in the NUMIDENT file, then we expect the same response in the CPS. However, if the

person is reported as a non-citizen in the NUMIDENT file, then the person could report as either a citizen or non-citizen in the CPS and still be considered consistent in their reporting.

DATA AND ANALYTICAL APPROACH

Data

Current Population Survey (CPS)

Our analysis relies on two sources of data. The first is the Current Population Survey (CPS), a monthly survey administered by Bureau of Labor Statistics. The CPS is a representative sample of the U.S. civilian population and includes roughly 60,000 households that are interviewed in-person. The CPS is the nation's leading indicator of employment dynamics, but also includes monthly modules including questions on topics such as fertility and educational attainment. For our analyses, we pool the 2009-2013 March samples, also know as the Annual Demographic Supplement, and limit our analysis to immigrant respondents who are born abroad to non-U.S. citizen parents (N=71,237).

The NUMIDENT

The NUMIDENT is the electronic file of applications and re-applications for a social security card, in numerical sequence. The NUMIDENT also houses death information of which the Social Security Administration (SSA) is aware. Unlike the death information, which can be shared with the public (with the exception of death information received from State vital statistics offices), the sharing of application and re-application information presents issues of confidentiality of personal information.

In-house research using the NUMIDENT can draw on the whole array of administrative record systems maintained by SSA. Thus, Duleep (Duleep and Dowhan 2002) has drawn on the earnings histories maintained by SSA to compare the earnings patterns of immigrants and non-

immigrants. Currently there is research underway to investigate the differential expectation of life at age 65 for immigrants by region of birth.

The ongoing matches of the Current Population Survey and the Survey of Income and Program Participation with a set of extracts from various SSA data systems provide an opportunity for researchers in certain settings to access NUMIDENT data – for persons interviewed in those surveys. Unlike the extracts from the earnings data base, the social security benefits database, and the SSI database, the NUMIDENT extract contains all the NUMIDENT information.

Several items on the application for a social security card are useful for studying immigration. These include city and state or country of birth, citizenship (U.S. citizen, legal alien allowed to work, legal alien not allowed to work, and other[2]), and date of original SSA application[3]. The date of the original application is useful because it sets an upper limit on the immigrant's date of entry into the United States. Often it will be very close to the date of entry – if the immigrant is in a hurry to get a number (and hasn't been here before without getting a number). In late 2002 SSA began the Enumeration at Entry initiative (similar to the better-known Enumeration at Birth initiative), in which a prospective immigrant applies for a social security number on the Application for Immigrant Visa (Form DS-230 or DS-260).

The NUMIDENT used to be a paper file. The electronic NUMIDENT contains over a billion records for almost 500 million number-holders, but does not contain *all* applications for a social security card. Until the mid-1970s, when a claim for a retirement, disability, or death benefit was adjudicated, the common practice among adjudicators was to remove the original application for a social security number from the paper file to associate with the documentation of the claim, with its place in the file taken by a "claim" record which did not contain all the items on the original

---

[2] In the past there also were these choices: (e) alien student, restricted work authorized and (f) conditionally legalized alien.

[3] There is also a particular "entry" code on the NUMIDENT which identifies emergency numbers issued to certain Indochinese refugees (Kestenbaum).

application, such as place of birth. Thus there is a "missing data" issue when using NUMIDENT data, of greater significance for very early birth cohorts.

Another limitation of the the NUMIDENT file is that it may not contain up-to-date information about citizenship. While persons who become naturalized citizens should request a replacement card that reflects their new status, there is no requirement to do so. If they don't, SSA will be unaware of the change. Additionally, until the 1970s, SSA assigned Social Security Numbers (SSNs) based on the applicant's allegations about identifying information without requiring corroborative evidence. As the SSN became increasingly used as a multi-purpose identifier by government, business, and other organizations, Congress legislated evidence requirements for SSN assignment because of concerns about illegal aliens working, as well as SSN fraud and abuse of public entitlement programs. It took some time for SSA to implement the Congressional mandate, but by May 15, 1978 all applicants were required to provide evidence of: (a) age, identity, and U.S. citizenship or lawful alien status for **original** SSNs; and (b) identity for **replacement** SSN cards. There remains no requirement to produce evidence for the alleged place of birth.

The NUMIDENT often has several records per number holder in cases in which replacement cards were requested when a card was lost or if information changed, such as surname upon marriage or citizenship upon naturalization. For ease of processing, we created a "best" record from the set of NUMIDENT records belonging to the number holder, where the "best" citizenship information was deemed to be that on the latest record and the application date that we saved was the first one.

<u>CPS-SSA match</u>

There is a very long history of matching Census Bureau surveys to Social Security Administration administrative data. A pilot match linked the March 1964 CPS to SSA earnings data. The 1973 Exact Match was a major undertaking that linked the March 1973 CPS to SSA earnings and benefit data and to individual tax data from the Internal Revenue Service. Probably more than

100 substantive and methodological papers used the 1973 Exact Match. More recently, linkages of the March CPSs to SSA administrative files has been carried out routinely for over a decade. Originally, the SSN was collected in the CPS from survey respondents willing to provide it and was the variable used to link on during the beginning of this period. As the number of refusals to provide the SSN mounted, the Census Bureau sought and found other linkage mechanisms. In particular, the Bureau used individual income tax files from the IRS with address and SSN information to match the CPS name and address to the IRS name and address and then the IRS SSN to the SSA SSN. Ultimately, the CPS stopped collecting the SSN altogether.

The percent matched (unweighted) are 87%, 86%, and 88% for the March 2009, March 2010, and March 2011 CPSs used in our analysis. Of course, the percentage matched is different for different groups. In particular, the percent matched is 89% for the (CPS) native-born, 85% for foreign-born naturalized citizens, and only 56% for foreign-born non-citizens.

Measures

We examine CPS-NUMIDENT match rates and consistency of information for three immigration related variables: place of birth, year of entry, and citizenship. Place of birth is measured in the CPS using a survey questions which simply asks respondents where they were born. In cases where respondents fail or refuse to provide their place of birth (just 0.8 % of the matched sample) country of birth is allocated by the Census Bureau. The corresponding information in the NUMIDENT is drawn from the SSN application form, which instructs applicants to indicate the city, state and country in which they were born. In presenting our results by place of birth, we distinguish between eight different countries or regions: (1) Mexico (2) Canada (3) Central America (4) Caribbean (5) South America (6) Europe (7) Asia, and (8) all other countries.

Year of entry in the CPS is based on the following question: "When did you come to the U.S. to stay?" Because the wording of this question is relatively vague, it is possible that respondents who have made multiple trips to the U.S. may provide their most recent arrival as opposed to their

first(Redstone and Massey 2004). Year of arrival information in the CPS is reported in intervals rather than in single-years and the width of the intervals varies both by survey year and year of entry. We recode these intervals into the following four decades: (1) entered prior to 1980; (2) entered between 1980 and 1989; (3) entered between 1990 and 1999; and (4) entered in 2000 or later.

Year of entry in the NUMIDENT file simply refers to the date on which the SSN applicant's information appears in the database. A relatively small number of persons are enumerated upon their admission into the United States as legal permanent residents, meaning that their entry into the NUMIDENT file coincides with their entry into the country. For most immigrants in the CPS, their date of entry into the United States will precede the date of entry into the NUMIDENT file, assuming they do in fact have a record in the file.

Citizenship is measured in the CPS by a question asking foreign-born respondents whether or not they are naturalized citizens. The response category to this question distinguishes between naturalized U.S. citizens and non-citizens, but does not distinguish between different legal status categories (e.g., LPRs, temporary workers, unauthorized migrants, etc.) among non-citizens.

Citizenship information in the NUMIDENT file is based on an item from the SSN application asking applicants to check one of four categories: U.S. Citizen, Legal Alien Allowed to Work, Legal Alien not Allowed to Work, or Other. These categories are collapsed into two, citizens and non-citizens, to be comparable to the CPS measure of citizenship. It is important to note that citizenship information in the NUMIDENT file will most often reflect a person's citizenship status at the time he or she applied for a SSN. Changes to citizenship status, for example if an LPR subsequently naturalizes, will most likely *not* be reflected in the NUMIDENT file.

Federal surveys, including the CPS**,** almost universally do not include direct measures of the legal residency status of non-citizens. They do, however, include indirect indicators that allow users to conclude with a relatively high degree of certainty that persons with such traits are legal

residents. These characteristics have been used by analysts, the Pew Hispanic Center in particular, to identify legal non-citizens in legal status imputation methods (Passel and Cohn 2010). We use the indicators available in the CPS in an analogous way in order to identify what we refer to as the "probably legal" non-citizen population. Probably legal non-citizens are those persons in the CPS with any of the following characteristics: (1) entered the United States prior to 1980[4]; (2) reports receipt of SSI, Social Security, or Medicare; or (3) is a veteran or active-duty service member of the U.S. military.

Analytical Strategy

We examine match rates of the foreign-born population in the CPS to the SSA NUMIDENT file. The match rate is simply the weighted percentage of the CPS sample that is successfully matched to the NUMIDENT. Match rates are reported separately by country / region of origin, decade of arrival in the U.S., age, gender, and citizenship and probable legal residency status.

The second emphasis of the analysis is concerned with the consistency, among successfully matched cases, of place of birth, period of arrival, and citizenship information between the CPS and NUMIDENT. With respect to place of birth, agreement between the two data sources is defined as cases in which *country of birth* (as opposed to regions) provided in the CPS is the same as that which appears in the NUMIDENT. The *consistency rate* is the percentage of cases in which there is country of birth agreement between the two data sources. This rate is examined for country / region of birth, and citizenship and probable legal residency status, as reported in the CPS. We also examine the consistency rate for *reported* country of birth and *allocated* country of birth information separately.

Two factors render the examination of consistency rates with respect to year of arrival less precise compared to country of birth. The first, as discussed above, is the fact that the CPS data

[4] The logic behind this approach is that the 1986 Immigration Reform and Control Act (IRCA) provided the opportunity for unauthorized residents who could prove continuous residence in the U.S. up to January 1, 1982 to regularize their status. Following others, we assume that all unauthorized residents entering the U.S. prior to 1980 adjusted to legal status.

report year of arrival in intervals rather than as single years. Secondly, also mentioned above, very often the date of the creation of a record in the NUMIDENT file will not match, and will most likely be later than, an individual's date of entry into the United States. We thus examine the weighted percentage of matched cases for which the date of entry into NUMIDENT falls within, before, or later than the time interval of arrival reported in the CPS. For persons enumerated by the SSA at the time of the legal entry into the U.S., we expect a relatively high rate of consistency between the two data sources (i.e., the date of entry into NUMIDENT will fall within the reported interval of arrival in the CPS a high percentage of the time). Furthermore, we expect the consistency rate to be lower among the majority of immigrants not enumerated upon entry, but in cases where the NUMIDENT year does not fall within the reported CPS interval, we expect that in the overwhelming majority of cases, the NUMIDENT year will be *later than* the reported CPS interval of arrival. Finally, due to the relatively small sample of persons enumerated at entry, we examine period of arrival consistency rates only for reported CPS responses, and exclude cases for which year of arrival has been allocated. For persons not enumerated at entry, we report consistency rates for reported versus allocated information separately, and, as with place of birth, we expect to find substantially lower consistency rates for the allocated information, relative to information provided by respondents.

Finally, we also examine consistency rates for citizenship responses. We assume that these rates will be the least precisely estimated of the three for reasons discussed above, namely that there is no certainty that persons who apply for a SSN as legal aliens, and who subsequently become naturalized citizens will update their citizenship information in the NUMIDENT. Due to this phenomenon, the examination of consistency rates with respect to citizenship is more likely to reveal the accuracy, or lack thereof, of citizenship information contained in the NUMIDENT file as opposed to in the CPS. Citizenship consistency rates are examined here by country / region of birth and citizenship and probable legal residency status, as reported in the CPS. We expect that the

consistency rate will be *lowest* among persons reporting as naturalized citizens in the CPS, insofar as these persons applied for their SSNs as legal non-citizens and subsequently naturalized without updating their citizenship status with the SSA.

RESULTS

CPS-SSA Match Rates

We begin with an examination of match rates presented in Table 1 by country/region of birth and decade of immigration, both as reported in the CPS. Overall, match rates among all foreign-born are about 71 percent. Match rates are substantially higher, approximately 88 percent, among persons arriving in the U.S. prior to 1980, which constitutes a very established population with very few unauthorized residents. Match rates fall among more recent arrivals to a mere 55 percent among those arriving in the country after 2000.

Table 1 also reveals predictable patterns across and within countries/region of birth. Regions of the world from which relatively small numbers of unauthorized immigrants hail, such as Europe, Asia, and Canada, show the highest match rates and relatively little variation in the rates across decade of arrival categories. But match rates among Mexicans (51 percent), Central Americans (60 percent), and South American (71 percent) are notably lower, and these relatively lower match rates are driven by the especially low rates among more recent arrivals. This pattern is strongest among Mexican immigrants. Mexicans who report in the CPS that they immigrated prior to 1980 are matched a rate of 85 percent, comparable to immigrants from other parts of the world, but the match rate decreases to 70, 43 and 26 percent among those immigrating during the 1980s, 1990s, and 2000s, respectively.

Our examination of match rates continues in Table 2, which presents percentages by country/region of birth, birth cohort, and gender. Following the same logic used to interpret the results in Table 1, if the presence/absence of a CPS-SSA match approximates an immigrant's legal residency status, we would expect match rates to be lowest among Mexican and Central American

immigrant men in their prime working ages. This hypothesis is supported by the results in Table 2, which shows that working-age Mexicans and Central Americans have the lowest match rates, with the lowest rates observed among young men. For example, Mexican-born men between the ages of 18 and 29 are matched at a rate of about 31 percent, slightly lower than their female peers, 34 percent. Match rates among young Central American men are even lower at about 30 percent, 16 percentage points lower than their similarly aged Central American female peers. Overall, male match rates (68 percent) are somewhat lower than among women (73 percent), and match rates among both men and women improve with increases in age.

We turn in Table 3 to an examination of match rates by citizenship and the presence or absence of indicators in the CPS data of legal residency status. Match rates are relatively high among those reporting as naturalized citizens in the CPS (86 percent) followed by non-citizens with characteristics that makes it highly probable that they are legally resident in the country (80 percent). It follows that non-matches are concentrated overwhelmingly among foreign-born respondents without any of the "probably legal" characteristics. Notably, match rates among Mexican-born U.S. citizens and those in the probably legal category, 81 and 77 percent, respectively, are somewhat comparable to persons in these categories born elsewhere. However, and as expected, match rates are especially low, just 35 percent, among Mexican-born respondents lacking any of the probably legal characteristics.

Consistency of CPS and Administrative Records

How consistent is information provided in the CPS with administrative records in the NUMIDENT file? Table 4 addresses this question for country / region of birth, and shows the weighted percentage of cases in which country / region of birth reported in the CPS matches the country / region in the NUMIDENT file. The percentages are presented for each of the three citizenship and legal residency categories, and separately depending on whether the CPS information is reported or instead allocated by the Census Bureau. The results in Table 4 point to

several important findings. First, when place of birth is *reported* there is a high degree of agreement with the place of birth information in the NUMIDENT file. This holds regardless of the citizenship and probable legal residency status of the respondent. Across all national/regional origin groups, place of birth responses reported in the CPS match those in the NUMIDENT about 95 percent of the time. There is little variation in this percentage across countries and regions worth noting.

A second important result in Table 4 suggests that agreement between the CPS place of birth and that found in the NUMIDENT is infrequent when country of birth is allocated in the CPS. It should be noted that the allocation rate (the percentage of observations in the CPS sample that is imputed by the Census Bureau rather than being provided by the respondent) is relatively low. Because very few place of birth responses are allocated in the CPS, many of the percentages for allocated cases in Table 4 are based on a relatively small number of observations. Nevertheless, even when examining the percentages for the total foreign-born population (right-most column), Table 4 suggests less than 20 percent of the allocated place of birth information in the CPS matches the NUMIDENT.

In the NUMIDENT file, a subset of foreign-born persons are enumerated and their information is entered into the SSA database at the same time that they are legally admitted into the United States. The year of arrival information included in the NUMIDENT file for these "enumerated-at-entry" persons, which constitutes a small minority of all the foreign-born in the NUMIDENT is compared to their year of immigration information provided in the CPS in Table 5. Due to the relatively small sample size of such persons, we report only percentages among those with a valid year of arrival response in the CPS (i.e., persons with allocated CPS year of arrival information are excluded). Because the CPS information is recorded in intervals, rather than in individual years, we report in Table 5 the percentage of observations for which the year of entry

into the NUMIDENT falls within the CPS interval, and if it does not, whether the date in the administrative record is earlier or later than the CPS interval.

Again, due to the relatively small number of observations, it is more useful to focus on the percentages for the total foreign-born population rather than for individual citizenship / probable legal status categories. Consistent with previous research demonstrating the relative difficulty of measuring immigrants' year of immigration, the results in Table 5 suggest far less agreement between CPS year of arrival responses and the corresponding information in the NUMIDENT. Overall, the NUMIDENT year of arrival falls within the interval reported in the CPS 68 percent of the time. This percentage, however, varies considerably across countries and regions in ways that one might expect given regional patterns of migration to the United States. For example, Mexicans and other Latin Americans are (a) more likely to have spent time in the U.S. without documents, and thus their year of entry into the U.S. may not correspond with the year in which they adjust to legal residency status and (b) more likely to engage in circular or repeat migration, regardless of legal status. As a result, the agreement between CPS and NUMIDENT information on year of arrival is the lowest, typically below 50 percent. Also consistent with this interpretation of the results in Table 5 is that fact that when the NUMIDENT year of arrival information does not fall within the CPS interval, the NUMIDENT information is most often *later* than the year reported in the CPS.

We turn in Table 6 to an examination of the agreement between CPS and NUMIDENT year of arrival information among persons *not* enumerated by the SSA at the time of their admission into the United States. As would be expected, the degree of consistency is substantially less among these persons compared to their counterparts who were enumerated at entry. Level of agreement when year of entry in the CPS is *reported* is especially low among Mexicans (19 percent), Central Americans (25 percent), and South Americans (37 percent). And when year of arrival is allocated in the CPS, correspondence with year of arrival information in the NUMIDENT is very infrequent. Just

eight percent of the time among the total foreign-born population does the year of entry in the NUMIDENT fall within the interval allocated in the CPS.

Our final CPS-NUMIDENT comparison is done with respect to citizenship, the results of which are reported in Table 7. Overall, citizenship status reported in the CPS is consistent with the NUMIDENT file 62 percent of the time. This result is more likely an assessment of the citizenship information contained in the NUMIDENT than it is an indication of the validity of the citizenship measure in the CPS. It is likely that many persons were non-citizens when submitting their initial application for a Social Security Number, but may have subsequently naturalized and not updated their citizenship status with the SSA. This interpretation is supported by the fact that consistency rates are much lower among immigrants reporting as naturalized citizens in the CPS, 55 percent of whom are not identified as U.S. citizens in the NUMIDENT file. As might be expected, among non-citizens in the CPS that are successfully matched to the NUMIDENT file, the consistency rates are higher than among their naturalized counterparts, although, curiously, this is especially true among non-citizens with none of the characteristics in the data that would place then in the probably legal category. Among those with no indication of legal status, 85 percent have citizenship information that is consistent with the NUMIDENT, while only 64 percent in the probably legal group are identified as non-citizen aliens in the NUMIDENT.

DISCUSSION AND CONCLUSIONS

Owing to the ever-increasing impact of immigration on U.S. society and its economy, it is crucial that Federal data collection systems collect information about the foreign-born population that is both adequate and accurate in order to facilitate policy relevant social science research aimed at understanding trends in immigration and immigrant incorporation. In this paper we examined the rate at which foreign-born respondents in the 2009-2013 Current Population Survey (CPS) are successfully matched to administrative data at the Social Security Administration (SSA),

and in cases where matches are made, the degree of consistency in the immigration-related information in the two data sources.

We argue that the results on match rates presented above carry important implications pertaining to the *adequacy* of information collected about immigrants in federally administered surveys most often used in research on the nation's foreign-born population. The relatively low match rates among certain segments of the foreign-born population reinforce what analysts have long known about Census surveys: that they include large numbers of unauthorized immigrants (Massey and Bartley 2005). To the extent that social and economic outcomes among immigrants and their families hinge on the degree to which they are excluded from participation in key aspects of U.S. society by virtue of their unauthorized status, existing survey data falls short of the task of providing analysts the necessary information about the foreign-born population needed to adequately understand the dynamics of immigrant adaptation and incorporation (Bachmeier et al. 2014; Clark and King 2008; Van Hook et al. 2015; Yoshikawa 2011) (Bachmeier, Van Hook and Bean 2014; Clark and King 2009; Van Hook et al., forthcoming; Yoshikawa 2011).

However, our results also imply that in the absence of legal status information collected in Census surveys, existing legal status imputation methods used increasingly by data analysts stand to be improved dramatically through the use of survey data linked with administrative data because our results suggest that success or failure of an attempted match is a reliable indicator of legal residency status. Van Hook et al. (forthcoming) recently demonstrated that the indirect indicators of legal status provided in Census survey data (e.g., welfare receipt, veteran status) are insufficient to distinguish legal from unauthorized immigrants, because only a relatively small percentage of non-citizens report these characteristics. Van Hook and her colleagues simulated the use of administrative data to substantially increase the share of non-citizens that can be coded as probably legal, which in turn dramatically improves legal status imputation strategies. Thus, in the absence of survey information on immigrants' legal status, survey data linked to administrative

records represents the most promising and immediate avenue through which research with crucial public policy relevance can be advanced.

The results presented above also carry implications for the *accuracy* of immigration-related information collected in Census surveys. In particular, we find that place of birth information reported by immigrant respondents in the CPS is highly consistent with information from NUMIDENT records. Implications with respect to the consistency of year of arrival information and, especially, citizenship reporting, are less certain owing to differences in the timing of the collection of information in the NUMIDENT relative to the CPS. Finally, our examination of consistency rates suggests that CPS-NUMIDENT agreement is very infrequent when the information in the CPS is allocated, as opposed to being reported by the respondent. Thus, survey information that is allocated at a relatively high rate, such as year of arrival, will have greater measurement error.

REFERENCES

Bachmeier, J. D., J. Van Hook, and F. D. Bean. 2014. "Can We Measure Immigrants' Legal Status? Lessons From Two U.S. Surveys." *International Migration Review* 48(2):538–66.

Bean, F. D., and G. Stevens. 2003. *America's Newcomers and the Dynamics of Diversity*. Russell Sage Foundation.

Clark, R. L., and R. B. King. 2008. "Social and Economic Aspects of Immigration." *Annals of the New York Academy of Sciences* 1136(1):289–97.

Duleep, H. O., and D. J. Dowhan. 2002. "Insights From Longitudinal Data on the Earnings Growth of U.S. Foreign-Born Men." *Demography* 39(3):485–506.

Lofstrom, M., S. Bohn, and S. Raphael. 2011. *Lessons From the 2007 Legal Arizona Workers Act*. San Francisco, CA: Public Policy Institute of California.

Massey, D. S., and K. Bartley. 2005. "The Changing Legal Status Distribution of Immigrants: a Caution." *International Migration Review* 39(2):469–84.

Meissner, D., D. M. Kerwin, M. Chishti, and C. Bergeron. 2013. *Immigration Enforcement in the United States.* Washington, DC: Migration Policy Institute..

Myers, D. 2007. *Immigrants and Boomers*. Russell Sage Foundation.

Passel, J. S., and D. Cohn. 2010. *US Unauthorized Immigration Flows Are Down Sharply Since Mid-Decade*. Washington, DC: Pew HIspanic Center.

Portes, A., and R. G. Rumbaut. 2006. *Immigrant America*. Univ of California Press.

Redstone, I., and D. S. Massey. 2004. "Coming to Stay: an Analysis of the U.S. Census Question on Immigrants' Year of Arrival." *Demography* 41(4):721–38.

Van Hook, J., and J. D. Bachmeier. 2013. "Citizenship Reporting in the American Community Survey." *Demographic Research* 29(1):1–32.

Van Hook, J., J. D. Bachmeier, D. Coffman, and O. Harel. 2015. "Can We Spin Straw Into Gold? an Evaluation of Legal Status Imputation Approaches." *Demography*.

Yoshikawa, H. 2011. *Immigrants Raising Citizens: Undocumented Parents and Their Young Children*. New York: Russell Sage.

TABLES

Table 1. Percentage of Foreign-Born Individuals in the 2009-2011 March CPS Matched to SSA Records by Country of Birth and Decade of Entry into the U.S.

	*Decade of Entry**				
Place of Birth*	pre-1980	1980-1989	1990-1999	2000+	*Total*
Mexico	85.4	70.0	43.2	26.2	*50.5*
Canada	92.7	95.4	80.0	79.9	*87.6*
Central America	85.0	85.0	65.1	28.1	*59.7*
Caribbean	90.9	84.7	84.5	78.0	*84.6*
South America	88.6	83.1	75.6	53.0	*71.0*
Asia	87.4	84.3	83.6	75.5	*81.8*
Europe	91.0	85.7	85.5	75.8	*85.7*
All Other	85.4	80.4	82.7	77.0	*80.0*
Total	88.4	80.0	67.9	55.2	*70.7*

*based on CPS measure

Table 2. Percentage of Foreign-Born Individuals in the 2009-2011 March CPS Matched to SSA Records by Country of Birth, Sex, and Age Cohort

	*Men, by Age Group**						*Women, by Age Group**					
	18-29	30-39	40-49	50-64	65+	*Total*	18-29	30-39	40-49	50-64	65+	*Total*
Place of Birth*												
Mexico	30.7	35.8	58.8	73.7	79.8	*48.2*	33.7	41.9	60.9	76.7	75.9	*53.3*
Canada	70.6	97.2	88.0	92.6	82.8	*87.4*	76.9	79.2	92.6	93.0	88.9	*87.7*
Central America	29.6	47.3	67.4	81.3	81.3	*53.3*	46.2	59.1	73.3	84.0	82.8	*67.0*
Caribbean	76.7	81.2	84.2	89.0	89.6	*84.5*	82.2	81.6	83.1	85.7	89.1	*84.6*
South America	52.6	56.5	70.5	77.6	88.3	*67.0*	59.4	72.5	76.1	81.8	83.9	*74.9*
Asia	76.3	83.5	84.2	85.2	84.1	*82.9*	74.1	81.8	81.9	83.4	80.5	*80.9*
Europe	79.9	81.3	85.0	88.6	90.4	*86.1*	79.1	80.7	86.0	88.3	89.1	*86.0*
All Other	74.8	79.8	79.8	85.1	82.2	*80.1*	77.7	81.6	80.6	79.2	81.9	*79.9*
Total	*50.2*	*59.2*	*73.2*	*82.4*	*85.9*	*68.3*	*58.1*	*65.6*	*75.4*	*83.0*	*84.1*	*73.2*

*based on CPS measure

Table 3. Percentage of Foreign-Born Individuals in the 2009-2011 March CPS Matched to SSA Records by Citizenship and Legal Status Indicator

		Non-Citizens*		
Place of Birth*	Naturalized Citizens*	No Indication of Legal Status*	Probably Legal*	Total
Mexico	80.9	35.1	77.0	*50.5*
Canada	93.2	81.8	84.1	*87.6*
Central America	84.5	46.7	76.0	*59.7*
Caribbean	87.9	78.9	85.6	*84.6*
South America	87.2	55.0	82.3	*71.0*
Asia	85.5	77.0	76.2	*81.8*
Europe	88.4	79.4	86.2	*85.7*
All Other	83.9	76.1	85.4	*80.0*
Total	*85.6*	*55.4*	*80.1*	*70.7*

*based on CPS measure

Table 4. Percentage of Matched CPS-NUMIDENT cases in Which Place of Birth (POB) in the CPS is Consistent with the NUMIDENT File, by CPS Place of Birth, CPS Citizenship and Legal Indicator Status, and Whether Place of Birth is Reported or Allocated in the CPS

			Non-Citizens*					
	Naturalized Citizens*		No Indication of Legal Status*		Probably Legal*		*Total*	
	POB Reported*	POB Allocated*	POB Reported*	POB Allocated*	POB Reported*	POB Allocated*	POB Reported*	POB Allocated*
Place of Birth*								
Mexico	93.3	32.3	89.2	26.4	92.6	35.0 [a]	91.4	29.6
Canada	94.1	0 [a]	92.9	0 [a]	95.7	0 [a]	94.0	0 [a]
Central America	91.6	13.1 [a]	92.7	28.0 [a]	95.5	0.0 [a]	92.4	17.6
Caribbean	96.1	8.9	96.1	31.7 [a]	95.8	44.6 [a]	96.1	22.3
South America	93.8	41.8 [a]	93.8	8.3 [a]	94.4	0.0 [a]	93.9	26.5
Asia	96.3	0 [a]	97.6	0 [a]	96.9	0 [a]	96.8	0
Europe	95.1	6.8	97.4	15.8 [a]	97.6	23.6 [a]	96.0	10.6
All Other	86.7	24.2	92.5	0.0 [a]	84.5	0.0 [a]	89.4	14.4
Total	94.7	16.9	94.1	21.6	94.7	20.1	94.5	19.0

*based on CPS measure

[a] Percentages are based on an unweighted cell with fewer than 20 observations

Table 5. Percentage of Matched CPS-Numident Cases in Which Year of Entry (YOE) into NUMIDENT is Earlier than, Within, or Later than the YOE Interval Recorded in the 2009-2011 CPS, by Place of Birth, Cititizenship and Legal Indicator Status, and Whether YOE is Reported or Allocated: Persons Enumerated at the Time of Legal Entry into the United States

				*Non-Citizens**								
	Naturalized Citizens*			No Indication of Legal Status*			Probably Legal*			*Total*		
CPS Place of Birth*	Earlier than CPS Interval	Later than CPS Interval	Within CPS Interval	Earlier than CPS Interval	Later than CPS Interval	Within CPS Interval	Earlier than CPS Interval	Later than CPS Interval	Within CPS Interval	Earlier than CPS Interval	Later than CPS Interval	Within CPS Interval
Mexico	0.0	74.7	25.3 [a]	4.5	42.6	52.9	9.7	85.1	5.2 [a]	4.4	50.4	45.2
Canada	0.0	0.0	100.0 [a]	100.0	0.0	0.0 [a]	--	--	-- [a]	88.5	0.0	11.5 [a]
Central America	0.0	100.0	0.0	6.6	39.3	54.1 [a]	0.0	0.0	100.0 [a]	6.1	39.0	54.9
Caribbean	17.5	55.3	27.2 [a]	0.0	19.4	80.6	33.7	12.3	54.0 [a]	5.1	21.2	73.8
South America	0.0	100.0	0.0 [a]	15.5	24.8	59.7	0.0	0.0	100.0 [a]	11.6	40.6	47.8
Asia	6.2	13.8	80.0	3.2	18.6	78.1	0.0	17.6	82.4 [a]	3.4	18.1	78.5
Europe	0.0	0.0	100.0 [a]	0.0	35.6	64.4	0.0	13.5	86.5 [a]	0.0	30.6	69.4
All Other	--	--	-- [a]	0.0	48.2	51.8 [a]	0.0	0.0	100.0 [a]	0.0	30.6	69.4 [a]
Total	5.0	39.1	55.9	3.2	26.5	70.3	13.4	28.2	58.4	4.1	27.9	68.0

*based on CPS measure

[a] Percentages are based on an unweighted cell with fewer than 20 observations

Table 6. Percentage of Matched CPS-Numident Cases in Which Year of Entry (YOE) into NUMIDENT is Earlier than, Within, or Later than the YOE Interval Recorded in the 2009-2011 CPS, by Place of Birth, Cititizenship and Legal Indicator Status, and Whether YOE is Reported or Allocated: Persons Not Enumerated at the Time of Legal Entry into the United States

				*Non-Citizens**								
	Naturalized Citizens*			No Indication of Legal Status*			Probably Legal*			*Total*		
CPS Place of Birth*	Earlier than CPS Interval	Later than CPS Interval	Within CPS Interval	Earlier than CPS Interval	Later than CPS Interval	Within CPS Interval	Earlier than CPS Interval	Later than CPS Interval	Within CPS Interval	Earlier than CPS Interval	Later than CPS Interval	Within CPS Interval
Mexico												
YOE Reported*	22.6	57.1	20.3	30.9	53.8	15.3	13.2	60.0	26.9	25.1	56.0	18.9
YOE Allocated*	44.3	48.3	7.4	58.5	35.3	6.2	43.4	50.5	6.1	50.5	42.8	6.7
Canada												
YOE Reported*	17.8	41.4	40.7	16.7	28.4	54.9	13.3	42.6	44.0	16.7	37.2	46.1
YOE Allocated*	31.5	45.8	22.7	42.7	51.0	6.3 [a]	9.5	81.0	9.5 [a]	30.4	52.5	17.1
Central America												
YOE Reported*	18.8	53.3	27.9	12.7	64.7	22.6	9.4	61.9	28.7	15.1	59.6	25.3
YOE Allocated*	31.4	61.1	7.5	45.9	41.7	12.4	16.5	75.2	8.2 [a]	37.7	52.4	9.9
Caribbean												
YOE Reported*	18.6	30.8	50.5	10.6	31.2	58.1	14.2	34.4	51.4	15.7	31.3	53.0
YOE Allocated*	45.2	42.8	12.0	55.6	29.2	15.2	39.8	49.5	10.8	48.4	38.6	13.0
South America												
YOE Reported*	19.1	41.4	39.5	13.6	53.5	32.9	17.9	47.7	34.5	17.0	46.3	36.7
YOE Allocated*	48.8	45.1	6.0	52.5	44.1	3.4	37.2	56.1	6.7	49.3	45.8	4.9
Asia												
YOE Reported*	16.0	28.0	56.0	11.8	28.9	59.3	16.8	30.2	53.0	14.5	28.4	57.0
YOE Allocated*	49.2	45.8	5.0	58.1	34.9	7.0	62.6	31.4	6.0	52.7	41.6	5.7
Europe												
YOE Reported*	15.7	29.8	54.5	14.8	29.1	56.1	12.9	28.2	58.9	15.2	29.5	55.3
YOE Allocated*	48.2	41.9	9.9	50.6	42.7	6.6	51.0	42.2	6.8	49.0	42.1	9.0
All Other												
YOE Reported*	18.0	25.2	56.7	12.5	31.5	56.0	20.0	25.8	54.2	15.4	28.3	56.3
YOE Allocated*	40.1	53.6	6.2	54.3	36.0	9.7	51.2	48.8	0.0	46.5	46.5	7.0
Total												
*YOE Reported**	17.8	35.5	46.7	17.2	40.3	42.5	14.1	43.5	42.3	17.3	38.0	44.8
*YOE Allocated**	45.2	46.8	8.0	55.5	36.5	8.1	45.2	48.2	6.5	48.9	43.2	7.9

*based on CPS measure

Table 7. Percentage of Matched CPS-NUMIDENT cases in Which Citizenship (CIT) in the CPS is Consistent with the NUMIDENT File, by CPS Place of Birth, CPS Citizenship and Legal Indicator Status,

		*Non-Citizens**		
	Naturalized Citizens*	No Indication of Legal Status*	Probably Legal*	*Total*
CPS Place of Birth*				
Mexico	52.4	83.9	76.8	67.9
Canada	37.1	89.1	51.6	56.9
Central America	48.5	89.4	70.4	70.5
Caribbean	49.7	84.1	65.3	62.4
South America	47.4	84.1	55.9	61.9
Asia	46.9	86.0	60.9	61.4
Europe	39.8	86.8	46.1	52.4
All Other	46.4	81.4	59.8	63.9
Total	54.7	85.0	64.2	62.2