

METHODOLOGY AND DETAILED ESTIMATES OF FOREIGN-BORN EMIGRATION,  
RETURN IMMIGRATION, AND NET EMIGRATION, 1996-2005

Estimates prepared for the U.S. Census Bureau and Sabre Systems, Inc.  
in association with the project Estimates By Migration Status 2 (EMS2)

Deliverable 3.2

Jennifer Van Hook  
Bowling Green State University

Frank D. Bean  
University of California—Irvine

Jennifer Van Hook, Ph.D. is an Assistant Professor in the Department of Sociology at Bowling Green State University and Frank D. Bean is a Professor of Sociology at the University of California, Irvine. This work was developed under subcontracts with Sabre Systems, Inc. and utilized funds provided by the U.S. Census Bureau. The opinions expressed in this work are those of the authors and do not necessarily reflect those of either Sabre Systems, Inc. or the Census Bureau.

## METHODOLOGY AND DETAILED ESTIMATES OF FOREIGN-BORN EMIGRATION, RETURN IMMIGRATION, AND NET EMIGRATION, 1996-2005

### *Summary*

This report presents the methodology and detailed estimates of foreign-born emigration by year, age, sex, country or region of origin, and citizenship and migration status. The estimates are based on the newly-developed CPS Match method (Van Hook, Zhang, Bean, and Passel 2006). This method takes advantage of the unique sample design of the CPS. The CPS follows housing units—but not necessarily individuals—over a period of 16 months. Individuals in the March CPS not successfully followed up include those who died, internal migrants, and emigrants. The CPS Matching method uses statistical methods to estimate the proportion of emigrants among those not followed up.

The development of new estimates of foreign born emigration is important for the production of accurate population estimates. As part of the U.S. Census Bureau's "Measuring Migration Across U.S. Borders" program funded in fiscal year 2003, the Immigration Statistics Staff is working to improve the estimates of international migration at the national level. These estimates are critical to creating total population estimates, as international migration constitutes nearly half of the population change from year to year. Because the current estimation method produced unreasonable levels and trends of international migration in the late 1990s, there is an urgent need for a new method. The Immigration Statistics Staff is working to produce new estimates of immigration based on year of entry (one year ago) data from the American Community Survey, yet the methodology requires new estimates of emigration in order to derive overall estimates of net international migration.

These are not the first estimates of foreign-born emigration. The Census Bureau had been producing these estimates using data from the decennial census since the former Immigration and Naturalization Service (INS) stopped maintaining administrative records on emigrants in the 1950s. In using the estimates of international migration to evaluate the results of Census 2000, research indicated that the method being used to estimate foreign-born emigration failed to produce reasonable levels in the 1990s, which led to inaccurate estimates of the stock of foreign-born on April 1, 2000 (Mulder et al. 2002).

Tables 1-3 report the estimated coefficients from models of internal migration, matching across CPS years  $t$  to  $t+1$ , and mortality. The coefficients in the models are used to produce predicted probabilities of matching across years, internal migration, mortality, and non-follow-up, which in turn are used to calculate the emigration estimates. Tables 4-15 report the estimated emigration, return immigration, and net emigration rates. Return immigration is the rate at which emigrants return to the U.S., and net emigration is the difference between emigration and return immigration. Tables 4-15 also report the predicted probabilities of matching across years, internal migration, mortality, and non-follow-up.

The estimates presented here tend to be higher than those in prior work (Van Hook, et al. 2006; 2005). One reason is that the estimates in this report are based on additional CPS years (1996-2005 instead of 1996-2003) and emigration appears to have increased in recent years. Second, the estimates presented here use the Census Bureau's internal CPS files rather than the public-use CPS files. It is not possible to match large portions of the oversample across years using the public use files due to problems in the data associated with identification numbers. Usage of the internal Census files permits the entire Hispanic oversample to be used, which leads to higher emigration rates. Finally, modifications were made in how emigration is estimated for children, which results in higher emigration estimates. These modifications are described in full in the report.

The estimates here differ from prior estimates based on the CPS-Matching Method because we have refined the estimation program. One refinement is that the estimation routine now pulls in out-of-range values (i.e., emigration probabilities less than zero or greater than one) while maintaining the same average emigration probability within country, age, and sex groupings. This means that emigration estimates are never permitted to be negative, although net emigration estimates may still be negative if return immigration estimates turn out to be larger than emigration (this can result from measurement or sampling error). Another refinement is that the algorithm introduces an element of realism down to the level of individual cases by shifting all non-zero probabilities of emigration, internal migration, mortality, and non-response to cases that actually were not followed up in  $t+1$ . Again, this is done while still maintaining the same average probabilities within country, age, and sex groupings (accomplished by scaling up the probabilities among non-follow-ups). A third refinement is that the program now forces the components of non-follow-up (probabilities of emigration, internal migration, death, non-response among non-follow-ups) to sum to 100%. This is accomplished by rescaling the components proportionately. This means that the components for the final emigration estimates (shown in the tables) will be exactly equal to the percentage non-matched minus the sum of the percentage of internal migrants, deaths, and non-responses.

## LIST OF TABLES

- Table 1. Logistic Regression Models of Internal Migration by Mexican-origin, Sex, and Generational Status
- Table 2. Logistic Regression Models of Matching Across Adjacent Years (t to t+1) in March CPS, by Mexican-origin, Sex or Age Group, and Generational Status
- Table 3. Discrete-time Event History Models of Mortality by Nativity (Logistic Regression Coefficients)
- Table 4. Foreign-born Emigration (Estimates and Components), Based on CPS Matching Method
- Table 5. Foreign-born Emigration Based on CPS Matching Method, by Year (3-year moving averages) and Mexican Origin
- Table 6. Foreign-born Emigration Based on CPS Matching Method, by 5-Year Age Groups
- Table 7. Foreign-born Emigration Based on CPS Matching Method, by Sex and Mexican Origin
- Table 8. Foreign-born Emigration Based on CPS Matching Method Among Males, by 5-Year Age Groups
- Table 9. Foreign-born Emigration Based on CPS Matching Method Among Females, by 5-Year Age Groups
- Table 10. Foreign-born Emigration Based on CPS Matching Method Among Mexicans, by Sex and 10-Year Age Groups
- Table 11. Foreign-born Emigration Based on CPS Matching Method Among Non-Mexicans, by Sex and 10-Year Age Groups
- Table 12. Foreign-born Emigration Based on CPS Matching Method, By Years in the U.S. and Mexican Origin
- Table 13. Foreign-born Emigration Based on CPS Matching Method, By Country or Region of Birth
- Table 14. Foreign-born Emigration Based on CPS Matching Method, By Mexican Origin, Citizenship, and Migration Status
- Table 15. Foreign-born Emigration Based on CPS Matching Method, By Period, Mexican Origin, Citizenship, and Migration Status

## METHODOLOGY AND DETAILED ESTIMATES OF FOREIGN-BORN EMIGRATION, RETURN IMMIGRATION, AND NET EMIGRATION, 1996-2005

Of the components of population change, the numbers of births, deaths, and new legal immigrant visas issued each year are known with considerable accuracy, if not virtual certainty, because the U.S. vital registration system and the former INS (now “U.S. Citizenship and Immigration Services”), are required by law to count these events and collect other data on them. In contrast, official statistics on emigration from the United States are virtually non-existent. The Immigration and Naturalization Service (INS) kept track of departing foreign-born emigrants from 1908 to 1957 (Woodrow-Lafield 1998), but eventually discontinued this practice due to concerns about the quality of the resulting data (Kraly 1998). Other direct methods for measuring emigration, such as multiplicity surveys attempting to identify emigrants by interviewing their relatives in the United States (Woodrow-Lafield 1996) and the use of administrative records (Duleep 1994), have met with, at best, limited success.

The relative inadequacy of emigration statistics can pose a problem for the production of population estimates. For example, national- and sub-national-level postcensal population estimates, produced annually by the U.S. Census Bureau, depend on the accuracy with which the components of demographic change are measured; population estimates built up through the cohort component method will be too low if emigration is overestimated and too high if it is underestimated. The population estimates have far reaching consequences on state and local decision-making such as the distribution of public funds and city planning. Furthermore, the population estimates serve as the basis for the development of sampling weights for many major surveys, including the Current Population Survey. In the case of residual estimates of unauthorized migrants, the accuracy of emigration rates among the legal foreign-born is critical; the estimates of unauthorized migrants vary directly with inaccuracies in measurement of emigration (Bean et al. 2001; Van Hook and Bean 1998). The level of emigration (as well as selectivity of emigrants) is also important for assessing how immigrant populations change with time in the United States. Without information about emigration, it is difficult to discern whether changes over time in such characteristics as health status, welfare receipt, income, or employment are attributable to living longer in the United States, aging, emigration, or some combination of the three.

Of necessity then, emigration has been estimated with a variety of indirect demographic methods, the most prominent of which is the residual method. The residual method estimates emigration by comparing the size of foreign-born cohorts between two decennial censuses. Residual estimates, however, are sensitive to inconsistencies in enumeration and reporting error between the two censuses, and ill-suited for measuring the emigration of recent arrivals, many of whom were not living in the United States at the time of the first census. In an effort to improve the quality of emigration estimates, we developed an alternative method for estimating emigration that takes advantage of the longitudinal nature of the Current Population Survey (described in Van Hook et al., 2005, 2006). Since the publication of our initial work, we made several adjustments to the original methodology and produced more detailed estimates of emigration by year, age, sex, duration in the U.S., country of birth, and immigration status. We believe that the adjustments to the methodology are likely to have improved the quality of the estimates. In this report, we describe the new methodology and present the detailed estimates. We do not review prior literature on emigration estimates here because a review and critique of

prior emigration methodologies and estimates is included in our prior work (Van Hook, et al. 2005, 2006).

Before delving into the methodology, we provide a brief overview for those who wish to skip over the details in the next section. Our emigration estimates are based primarily on analyses of rates of attrition in Current Population Survey data. A key feature of the CPS sample design—one that is critical for our purposes—is that it follows housing units over time. The CPS interviews occupants of the housing units for 4 consecutive months; these units do not appear in the next 8 monthly CPSs. Then, the occupants of the same housing units are interviewed for 4 additional consecutive months. Each interview, numbered 1 through 8, is referenced by the “months-in-sample” variable. With this design, those in months-in-sample 1 through 4 in year  $t$  appear in months-in-sample 5 through 8 in year  $t+1$ . It is important that the sample is of *addresses*, not *individuals*. Thus, if a CPS respondent moves to a new address, he/she is not followed. Rather, the new occupants of the original housing unit are interviewed and the original respondent is dropped from the sample. This feature of the CPS sample design permits us to use follow-up rates—i.e., the proportion of persons in months-in-sample 1–4 in one year who are successfully interviewed as members of months-in-sample 5–8 in the following year—as a basis for estimating emigration. (See U.S. Census Bureau (2002a) for a detailed description of the CPS design.)

Individuals in the March CPS in one particular year (year  $t$ ) who do not appear in the following year’s March CPS (year  $t+1$ ) include those who died, internal migrants (who moved to other residences in the U.S.), emigrants who moved out of the country, and a residual group who cannot be matched for other reasons. Madrian and Lefgren (1999) estimate that 29 percent of those eligible for follow-up in the March CPS 1980-1998 surveys were not successfully followed up. Based on known rates of internal migration and mortality (derived from the CPS and NCHS statistics), Madrian and Lefgren (1999) also estimate that 16.3 percent moved to another address in the United States and 0.9 percent died, leaving 11.8 percent who were not followed up for other reasons. Of the residual 11.8 percent, some may have moved to another country while others may not have been followed due to non-response, coding error, or some other reason. Thus, 11.8 percent is the maximum percent of emigration, and this figure is almost certainly far too high because it does not take into account other reasons for non-follow-up. Our basic task is to subdivide the residual into emigration and residual-non-follow-up components. On the basis of prior knowledge and some assumptions about factors affecting the rates of internal migration, mortality, and non-follow-up, we use statistical methods to estimate the probability that non-matched individuals died, moved internally, emigrated, and were not followed for other reasons. We do not explicitly assign individuals categorically as an emigrant or not an emigrant. Rather, each individual is assigned a *probability* that they emigrated. We average the probability of emigration across all foreign-born who first appeared in the March CPS in year  $t$  to estimate the proportion of emigrants among the foreign-born.

One advantage of the CPS Matching Method is that it does not depend critically on the consistency of year-of-entry or coverage. Unlike the residual method, which compares the sizes of foreign-born cohorts between two data sources collected ten years apart, the CPS Matching Method follows *individuals* over time. There is no need to assume consistency in coverage or reporting between the two surveys because all social and demographic information (age, period-of-entry, place of birth, sampling weight) is obtained from a single data source: the CPS in year  $t$ . This feature of the CPS Matching Method is particularly valuable because it permits the estimation of emigration rates for groups defined on the basis of time-varying characteristics

such as health status, income, poverty, or welfare receipt. Another advantage is that the CPS Matching Method estimates emigration rates for recently-arrived foreign-born persons in the same manner as earlier arrivals. The new method is therefore more likely to produce comparable estimates across different period-of-entry groups than the residual method.

As elaborated below, the CPS Matching estimates depend critically on the accuracy of certain assumptions. Two of the most significant are that (1) emigration rates among second-generation native-born adults are negligible and (2) foreign-born and the second generation adults have similar patterns of non-follow-up due to unmeasured causes (while controlling for a number of socioeconomic factors).

The estimates we present here tend to be higher than those in prior work (Van Hook, et al. 2006; 2005). One reason is that the estimates presented in this report are based on additional CPS years (1996-2005 instead of 1996-2003) and emigration appears to have increased in recent years. Second, the estimates presented here use the Census Bureau's internal CPS files rather than the public-use CPS files. It is not possible to match large portions of the oversample across years using the public use files due to problems in the data associated with identification numbers (the origin of these problems are still not fully understood by the authors). Usage of the internal Census files permits the entire Hispanic oversample to be used, which leads to higher emigration rates. Finally, modifications were made in how emigration is estimated for children and in the level of country-of-origin detail used in the production of estimates, which results in higher emigration estimates. These modifications are described in full in the section below.

## METHODOLOGY

### Basic Approach

We begin by representing the proportion of persons in the CPS not followed up ( $u$ ) as the sum of the proportion who migrated within the United States ( $m$ ), the proportion who died in the United States ( $d$ ), the proportion who emigrated ( $e$ ), and the proportion who were not followed up for other reasons ( $r$ ). These components can be estimated for subgroups of the population. Thus, for the foreign born ( $f$ ), we represent the relationship as:

$$u^f = m^f + d^f + e^f + r^f \quad (1)$$

Most of these terms may be estimated from existing data. The non-follow-up probability ( $u^f$ ) may be estimated as the number of persons followed up in the March CPS in year  $t+1$  divided by the number eligible to be matched in the March CPS in year  $t$ . The proportion of internal migrants ( $m^f$ ) may be estimated, with certain adjustments, from the place-of-residence-one-year-ago question in the CPS. The probability of death ( $d^f$ ), a small component except in the older ages, may be estimated for the foreign born using the National Health Interview Survey or NHIS (Palloni and Aries 2004). We are left with the proportion of emigrants ( $e^f$ ) and residual non-follow-up probability ( $r^f$ ) for the foreign born. Once we estimate  $r^f$  we can solve for  $e^f$ .

To estimate  $r^f$ , we make two assumptions. The first is that foreign born and second generation adults age 15+ ( $s$ ) have the same non-follow-up probabilities after adjusting for compositional

differences in demographic characteristics. Second generation adults are the U.S.-born adult children of the foreign born. Thus:

$$\begin{aligned} r^f &= r^s \\ &= u^s - m^s - d^s - e^s. \end{aligned} \quad (2)$$

We choose second-generation adults rather than all native-born adults as a comparison group for reasons explained further below. Substituting equation (2) into equation (1) and solving for  $e^f$  yields:

$$e^f = u^f - m^f - d^f - (u^s - m^s - d^s - e^s). \quad (3)$$

The second assumption we make involves the value of  $e^s$ . Fernandez (1995) estimates that during the 1980s, roughly 48,000 U.S. born emigrated per year. This level of emigration amounts to an annual rate of about .02 percent among all U.S. natives. Even if all native-born emigrants were second generation (that is, U.S. born children of foreign-born parents) this level of emigration would amount to an annual rate of 0.2 percent for the second generation<sup>1</sup> and the rate is most likely even lower for second generation adults (because second generation children are more likely to emigrate with their foreign-born parents). Work with more recent data suggests that even this small level of emigration is too high, perhaps by a factor of 3 (Gibbs et al. 2003)<sup>2</sup>. Thus, we make the assumption that the emigration probability of second-generation adults is negligible or essentially zero, and equation 3 reduces to an expression that can be calculated with existing data:

$$e^f = u^f - m^f - d^f - (u^s - m^s - d^s). \quad (4)$$

*Native-born Comparison Group.* The selection of a native-born comparison group is an important issue. The underlying assumptions of the matching method are that (1) the native comparison group has very low rates of emigration, and (2) behaves similarly to the foreign born with respect to the factors other than emigration affecting residual non-follow-up. Satisfying both assumptions simultaneously may be difficult. On the one hand, the third-or-higher generation (i.e., U.S.-born children of U.S.-born parents) may serve as a good comparison group because they may be less likely to emigrate than the second-generation as they tend to have fewer family connections overseas. On the other hand, the second generation may serve as a good comparison group because they may behave more similarly to the foreign born vis-à-vis non-response than the third-or-higher generation (based on standard ideas about assimilation). If we obtained up-to-date estimates of emigration among the second generation that could be factored into the final estimates, we may be able to relax the first assumption. We believe it would be more difficult to relax the second assumption due to the difficulty in directly measuring generational differences in attrition in the CPS. Therefore, to increase the likelihood

---

<sup>1</sup> This figure is based on Passel and Edmonston's (1994) estimate that there were 24,006,000 and 24,354,000 second generation persons living in the U.S. in 1980 and 1990, respectively. Thus the average annual emigration rate over the 1980-decade is  $48,000 / (24,006,000 + 24,354,000)/2 = .002$ .

<sup>2</sup> Fernandez (1994) and Gibbs et al. (2003) present the only empirically-based estimates of U.S. born emigration. Although this work has its own limitations, we have no alternative at this point other than to use it. Further work on U.S. born emigration is necessary to provide more support for the assumption that U.S. born emigration is so low.

that the second assumption about non-follow-up holds, we opt to use the second generation rather than all natives or the third-or-higher generation as the native comparison group.

While the choice of the second generation as a comparison group offers some advantages for adults, this probably not true for children. Both foreign-born and second-generation children are the children of foreign-born parents; they often share the same households and are likely to have similar emigration rates. Therefore, in the case of children, the estimated emigration rates of the second generation are not likely to be negligible, as required in the assumptions used to derive emigration rates for the foreign born. Thus, if we were to use the methodology outlined above for estimating children's emigration rates, we would almost certainly underestimate emigration rates for foreign-born children. For this reason, we treat children ages 0 to 14 differently from adults. Rather than estimating children's non-response rate, we assign foreign-born children the non-response rate ( $r^f$ ) estimated for their parent (whichever parent or guardian is identified in the PARENT line number variable). This appears to be a reasonable assumption since children are not actually interviewed in the CPS; rather, parents report about their children in the CPS. We also assign foreign-born children the internal migration rate ( $m^f$ ) estimated for their parent. This is justified because the internal migration measure in the CPS is retrospective and therefore cannot be used to estimate the probability of future internal migration for very young children. We use the same methodology for estimating mortality, and non-follow-up for the children as we do for the adults. Thus for children,

$$e^f = u^f - pm^f - d^f - pr^f. \quad (5)$$

where  $pr^f$  is the non-response rate and  $pm^f$  is the internal migration rate estimated for the child's parent. In our previous work, we used a slightly different strategy for estimating emigration among children, whereby we simply assigned children the estimated emigration rate of the parent. However, this was problematic because it could not detect circumstances, such as in the case of a divorce, in which a child may stay with one parent (or other relative) while the other emigrates or vice versa.

*Internal Migration.* We base our estimates of internal migration on the question in the CPS that asks where the respondent lived one year before. CPS respondents who lived abroad a year before (some of whom are "return immigrants" who emigrated but then returned to the U.S.) are excluded from the analytical sample since this group was not at risk of moving internally. However, because the internal migration question in the CPS is retrospective, the population at risk—as it is measured in the CPS in year  $t+1$ —excludes some who were actually at risk of moving internally in year  $t$  such as those who died in the U.S. or emigrated in the previous year and are therefore no longer in the CPS universe. The "true" population at risk of moving internally between  $t$  and  $t+1$  ( $P_t^*$ ) is therefore equal to:

$$P_t^* = P_{t+1} / (1 - e - d),$$

where  $P_{t+1}$  is the population at risk as it is measured in the CPS,  $e$  is the proportion emigrating, and  $d$  is the proportion dying in the U.S. between  $t$  and  $t+1$ . Because  $P_{t+1}$  is less than  $P_t^*$ , the unadjusted CPS-based estimates of internal migration, which use  $P_{t+1}$  as a base, are too high. We therefore adjust the internal migration probability ( $m$ ) whereby the adjusted probability  $m^* = m(1 - e - d)$ . For second generation adults, among whom  $e$  is assumed to be zero, the adjusted internal migration probability reduces to  $m^* = m(1 - d)$ . This means that equation 4 expands to:

$$e^f = u^f - m^f(1 - e^f - d^f) - d^f - [u^s - m^s(1 - d^s) - d^s],$$

and rearranging terms:

$$e^f = [u^f - m^f + m^f d^f - d^f - u^s + m^s - m^s d^s + d^s] / (1 - m^f). \quad (6)$$

For children, the equation is:

$$e^f = [u^f - pm^f + pm^f d^f - d^f - pu^s + pm^s - pm^s pd^s + pd^s] / (1 - m^f). \quad (7)$$

where  $pu^s$ ,  $pm^s$ , and  $pd^s$  are the child's parent's values for  $u^s$ ,  $m^s$ , and  $d^s$ , respectively.

*Emigration and Return Immigration.* Emigration estimates from the CPS Matching Method are likely to be larger than those based on the residual method because the residual method does not count as “emigrants” those who leave the United States but later return within the decade (i.e., so-called “return immigrants”<sup>3</sup>). Specifically, the residual method does not measure the *annual* number of emigrants directly. Rather, it typically estimates *net* emigration over a decade and then divides this estimate by ten to obtain average annual emigration.<sup>4</sup> However, some foreign-born persons may have been living in the United States both at the beginning and end of the decade while having made several trips back and forth during the decade. This phenomenon seems particularly important in the case of Mexican migration to the United States. Massey and his colleagues estimate that the average duration of a Mexican labor migrant's first trip to the United States is only 21 months and that one-third of these migrants return to the United States in a second trip within ten years of the first trip (Massey et al. 2002).

In general, emigration rates increase as the sub intervals over which cohorts are followed become shorter. The number of net emigrants over a ten year period, say 1990–2000, is equal to the sum of the number of emigrants each year ( $E_y$ ) minus the number of return immigration trips to the U.S. in each year among those who emigrated during 1990–2000 ( $R_y$ ), or:

$$E = \sum_{y=1990}^{y=2000} E_y - \sum_{y=1990}^{y=2000} R_y$$

Taking the annual average,  $\frac{E}{10} = \bar{E}_y - \bar{R}_y$ . (8)

By this relationship, the difference between average annual estimates that use a ten-year interval ( $\frac{E}{10}$ ), as do most residual methods, and a one-year interval ( $\bar{E}_y$ ), as does the CPS

---

<sup>3</sup> We use the term “return immigration” to denote immigration to the United States by former immigrants who have left the United States to live abroad, but have returned to the United States. We use this term to distinguish the phenomenon from “return migration” which is usually used to mean emigration from the United States or return by immigrants to their home country.

<sup>4</sup> Average annual emigration rates are sometimes computed by dividing the average annual emigration by the mid-period foreign-born population. A better method is to derive the annual rate as one minus the 10<sup>th</sup> root of the 10-year probability that an immigrant will *not* emigrate. In neither case, however, is annual emigration measured directly.

Matching Method introduced here, is equivalent to the average annual number of returns to the U.S. by former immigrants ( $\bar{R}_y$ ).

We derive *net* emigration measures comparable to those produced by residual methods and to those required for population estimates. Dividing each side of equation (6) by the population at risk of emigrating in year  $t$  averaged across all years in the decade ( $P_t$ ), we express the relationship in terms of rates or probabilities:

$$\frac{E/10}{P_t} = \frac{\bar{E}_y}{P_t} - \frac{\bar{R}_y}{P_t}. \quad (9)$$

Thus we estimate the average annual net emigration rate (estimated by the residual method), shown on the left-hand side, as the difference between the annual gross emigration rate (estimated by the CPS matching method) and an estimate of return immigration that we refer to here as the “return immigration ratio.” Not a proper rate or probability, the return immigration ratio is the number of return immigrants appearing in the year  $t+1$  CPS relative to the number of persons who were living in the U.S. and at risk of emigrating in year  $t$ . The denominator thus excludes the return migrants in year  $t+1$  (they were living abroad in year  $t$ ) but includes those who died or emigrated between year  $t$  and  $t+1$ .

### Estimation Strategy

In this section, we describe the specific statistical methodology used to produce the emigration estimates for foreign-born adults and children. The predicted probability of non-follow-up for each person can be estimated with logistic regression as shown below, where  $X_i$  is a vector of independent variables observed at time  $t$ , the “ $i$ ” and “ $n$ ” subscripts denote the foreign-born and second-generation samples, and the “ $f$ ” and “ $s$ ” superscripts denote the foreign-born and second-generation coefficients, respectively:

$$u_i^f = \frac{\exp(X_i' \beta^f)}{1 + \exp(X_i' \beta^f)} = F(X_i, \beta^f) \quad (10a\text{—foreign born})$$

$$u_n^s = \frac{\exp(X_n' \beta^s)}{1 + \exp(X_n' \beta^s)} = F(X_n, \beta^s) \quad (10b\text{—second generation})$$

Each component of non-follow-up in equation 5 can be similarly expressed and estimated:

<u>Foreign born:</u>	<u>Second Generation:</u>
$m_i^f = F(X_i, \mu^f)$	$m_n^s = F(X_n, \mu^s)$
$d_i^f = F(X_i, \delta^f)$	$d_n^s = F(X_n, \delta^s)$

In addition, we estimate for foreign-born children their parent's probability of non-follow-up, internal migration, and death:

Foreign born children:

$$pu_i^f = F(P_i, \beta^f)$$

$$pm_i^f = F(P_i, \mu^f)$$

$$pd_i^f = F(P_i, \delta^f),$$

where  $\mathbf{P}_i$  is a vector of parental and child characteristics observed at time  $t$ . A key insight of our method is that foreign-born and second generation adults have equivalent residual non-follow-up probabilities after removing the influence of compositional differences. To remove the influence of compositional differences, the second generation components of equations 3 through 6 are estimated as predicted probabilities for the foreign born using second generation coefficients. For example, the non-follow-up probability of the second generation *assuming foreign-born adults' composition*, designated here with an subscript "i" but an "s" superscript, is obtained by replacing the coefficients in (10a) with second generation coefficients:

$$u_i^s = F(X_i, \beta^s),$$

Our first key assumption is that foreign-born adults have the same residual non-follow-up rates as their second-generation counterparts. In other words:

$$\begin{aligned} r_i^f &= r_i^s \\ &= u_i^s - m_i^s - d_i^s - e_i^s \end{aligned}$$

Our second assumption is that emigration among second generation adults is zero:  $e_i^s = 0$ . Equation 4 (the estimate for foreign-born adults) therefore is expressed as:

$$e_i^f = u_i^f - m_i^f - d_i^f - (u_i^s - m_i^s - d_i^s)$$

After making adjustments for internal migration being measured retrospectively and rearranging terms (following equation 6), the probability of emigrating is estimated as:

$$e_i^f = \frac{u_i^f - m_i^f + m_i^f d_i^f - d_i^f - u_i^s + m_i^s - m_i^s d_i^s + d_i^s}{1 - m_i^f} \quad (11)$$

Equation 7 (the estimate for foreign-born children) is expressed as:

$$e_i^f = \frac{u_i^f - pm_i^f + pm_i^f d_i^f - d_i^f - pu_i^s + pm_i^s - pm_i^s pd_i^s + pd_i^s}{1 - pm_i^f} \quad (12)$$

Equations 11 and 12 are particularly useful because all components can be estimated with CPS and NHIS data. Averaging (11) and (12) across all foreign born yields the gross emigration rate among the foreign born, and the gross emigration rate minus the "return immigration" ratio yields the net emigration rate.

## Data

To estimate foreign-born emigration rates for the late 1990s and early 2000s, we use the 1996-2005 Annual Demographic Supplements to the March CPS, designated as the Annual Social and Economic Supplements beginning with March 2003. The supplements from this month offer several advantages over other months. The March Supplements contain a substantial range of socioeconomic and demographic information not in other months. The information needed to identify nativity and generational status appear in every monthly CPS since 1994, but only the March supplement contains the question on residence one year ago that we use to identify internal migrants and return immigrants. A further advantage of the March supplements is that the samples are larger than in other months. Since the mid-1970s, the March supplement has contained an oversample of Hispanics, a sampling scheme that effectively doubles the number of Hispanic households in the March Supplement. Beginning with the March 2002 CPS, the supplement has been expanded further by adding additional households from non-overlapping rotation groups in adjacent months. Since emigration is a relatively rare event, the larger samples provide more precise estimates. During the 1996–2005 period, the basis for CPS weights changed from the 1990 Census to Census 2000. The official change-over occurred with the March 2002 CPS which was the first to use weights based on Census 2000. However, the March 2001 SCHIP file and a research version of the March 2000 Supplement also used Census 2000-based weights. Where possible, we use the 2000-based weights.

The data we use here to estimate foreign-born emigration differ in two respects from the data used in our previous work. First, we now include data from the 2004 and 2005 March CPSs (we used 1996-2003 in the prior work). Second, we now use the internal (restricted-use) CPS files rather than the public-use files. This change is likely to have improved the estimates because it permitted us to use the entire Hispanic oversample. We discovered in our prior work that it is not possible to match large portions of the oversample from year  $t$  to  $t+1$  when we used the public use files. This problem appears to have been associated with missing components of the identification numbers for the oversample in the public-use files. When we matched the oversample from year  $t$  to  $t+1$  using the internal Census files (using a series of identification numbers available in the internal files), a much higher proportion matched across years than in the public-use files.

The analytical sample used to estimate emigration includes all foreign-born persons and second generation adults in the 1996 through 2004 CPS March samples who were eligible to be followed up in the following year. This means that the sample is restricted to those in months-in-sample 1–4 (or 5 in the case of the oversample). The final sample included 77,491 foreign born adults and children (35,204 men, 37,082 women, and 5,205 children) and 48,694 second generation adults (23,100 men and 25,594 women).

We also use the National Health Interview Survey-National Death Index (NHIS-NDI) data to model the probability of dying in the U.S. for the foreign born and native adults. Conducted each year since 1957, the NHIS is an annual survey of individuals age 18 and older about health status, health care, and insurance coverage. Beginning with the 1986 sample, NHIS respondents were linked to the National Death Index (NDI) files (a data base of all deaths in the United States) in order to ascertain vital status and age at death. NHIS respondents are matched on a number of identifiers, including social security number, first and last name, father's surname, and month and year of birth. Details about the methodology and quality of

matches are discussed in the NHIS documentation (NCHS 2000). As of the time we conducted our analysis, NHIS respondents had been linked to the 1987 through 1997 NDI files. The NHIS did not include a question on place of birth until 1989, so we use the 1989 through 1994 NHIS files, which are linked to the 1989-1997 NDI files. We organize the NHIS-NDI data in person-year records, including a record for each year of life lived by NHIS respondents from the time of the survey and the time of their death or censorship in 1997, whichever comes first. The analytic data file includes 344,536 person-year records for the foreign born (2,480 deaths) and 2,767,340 person-year records for natives (27,652 deaths).

Although we use non-CPS data for modeling mortality, this does not present a serious problem. Our method only requires that we obtain a vector of coefficients that can be used to predict the probability of non-follow-up, internal migration, and mortality. Once we estimate coefficients from a given sample, we apply the coefficients to the foreign-born in the CPS to calculate predicted probabilities of non-follow-up, internal migration, and mortality. Because children are not included in the NHIS and child mortality is likely to be very low, we made the assumption that the probability of death for children ages 0-14 is equal to zero. In future work, it would certainly be possible to try to account for mortality for children using vital records, but it is unlikely that such estimates would be substantively different from the current estimates.

### Non-follow-up

To determine whether a respondent in the March Supplement to the CPS in year  $t$  is successfully followed up the following year  $t+1$ , we match those eligible for follow-up in the 1996–2004 March CPSs with respondents in the following years' CPSs, 1997–2005. Except for the oversample, households from rotation groups 1–4 in each year  $t$  are matched to rotation groups 5–8 in the following year  $t+1$ <sup>5</sup>. Then, matching individuals in these households are identified. The matched and unmatched individuals in year  $t$  are used to measure follow-up rates. We match households across years using the following identifiers: primary sampling unit (H\_PSU), segment (H\_SEG), serial number (H\_SER), serial suffix (H\_SERSUF), and household number (H\_HHNUM), and we match individuals using the person line number. Because matched cases may not represent the same individual due to coding errors on the person or household identification variables, we also require consistency in sex and age before considering a case a “true” match.<sup>6</sup> We do not require consistency on race or Hispanic origin because the race question changes in 2003 (allowing responses in multiple categories) and because of response inconsistency and variability.

### Internal Migration

The CPS asks respondents whether he/she lived in a different residence one year before. We define the internal migration probability for the time period from year  $t$  to  $t+1$  as the proportion of movers among those who reported having lived in the United States one year before. This figure is adjusted in the final estimation of emigration for biases associated with internal migration being measured with retrospective data.

---

<sup>5</sup> The oversample is coded as having a month-in-sample of 5 in both year  $t$  and  $t+1$ .

<sup>6</sup> For example, a person at year  $t$  can be no more than 2 years younger than the matched case in year  $t+1$ .

## Return Immigration

We define “return immigrants” as the foreign-born population who reported in year  $t+1$  living abroad one year earlier, but also reported having come to live in the United States more than one year before. We estimate “return immigration” ratios as the number of adult and children return immigrants in year  $t+1$  divided by those foreign-born children and adults in the  $t+1$  CPS *who lived in the U.S.* in year  $t$  (that is, excluding return immigrants). This figure is then adjusted to account for emigration and mortality occurring between years  $t$  and  $t+1$  using the same logic as with the internal migration estimates, multiplying the ratio by  $(1-e-d)$ .

### Predicted probabilities of non-follow-up, internal migration, and mortality

To obtain values for the components of equations 11 and 12, we start by estimating three sets of weighted logistic regression models: the first predicts non-follow-up among those eligible to be followed up; the second set predicts internal migration (i.e., living at a different address from the year before) among those who were living in the United States the year before; and the third set predicts a one-year probability of dying in the United States. Predicted one-year probabilities of deaths occurring in the U.S. for adults ages 18 and over are obtained from the National Health Interview Survey-National Death Index for 1989–97. Using a person-year file, we estimate separate logistic regression models for the foreign born and natives predicting whether a person died in the U.S. during the year, including as independent variables age, sex, race/ethnicity, and general health status. Because of unavailability of questions on parents’ place of birth in the NHIS, we estimate the native mortality models on all natives together rather than for solely the second generation. The independent variables in our models include sex, age, race/ethnicity, and general health. We use coefficients from the “foreign born” and “native” mortality models to generate, respectively, the “foreign born” and “second generation” predicted probabilities for immigrants in the CPS. The model estimates are presented in Table 3.

The models for non-follow-up and internal migration are estimated for all persons separately by age group (0-14, 15+), sex (among 15+ only), Mexican/non-Mexican ethnicity, and generational status (1<sup>st</sup> and 2<sup>nd</sup> generation). In models estimated for those age 15+, we include as independent variables whether the person was in the CPS oversample, homeownership status, age, year, school enrollment status, education, and country of birth (for models estimated on the foreign-born) or parents’ country of origin (for models estimated on the second generation). Because education was not significant in any of the internal migration models, education was included in the models of non-follow-up only<sup>7</sup>. For models estimated for children ages 0-14, the independent variables include whether the person was in the oversample, parental homeownership status, sex, age, year, parental education, parent’s age, whether the parent was successfully followed up, and country of birth. We estimated a total of 8 models of internal migration (Mexican/non-Mexican by generation for men and women separately) and 10 models of non-follow-up (Mexican/non-Mexican by generation for men, women, and children separately). We present the estimates for all models in Tables 1, 2, and 3.

---

<sup>7</sup> We found that the emigration estimates are remarkably stable across model specifications and do not change very much when additional variables such as health status, household composition, and detailed race/ethnicity are added to the models.

We generate predicted values of the likelihood of non-follow-up and of internal migration from the appropriate model. Two sets of predicted values of non-follow-up and internal migration are calculated for each foreign-born adult: first using foreign-born coefficients and second, using second-generation coefficients (a total of 4 predicted values). Predicted values for each sex and race/ethnic group are derived from each groups' corresponding models. For example, the predicted values for Mexican males come from the "Mexican male" models. Parental predicted values are merged onto their children's data records. The predicted values of non-follow-up, internal migration, and death are then used in equations 11 and 12 to estimate an individual-level predicted probability of emigration for foreign-born adults and children. Finally, we average the individual-level probabilities of emigration to obtain an estimate of the emigration rate for all foreign born and for foreign-born subgroups by age, sex, country-of-origin, and year-of-entry.

We calculated all standard errors using the methodology provided in the CPS documentation by applying the "b" factors associated with Hispanics (U.S. Census Bureau 2002b).

### Consistency Checks and Adjustments

After estimating emigration probabilities and the underlying components (non-follow-up, internal migration, mortality, and non-response probabilities), we ran the estimates through a series of consistency checks and made refinements to the estimates as necessary. We first checked the estimated emigration probabilities for out-of-range values (less than zero or greater than one), and then pulled in out-of-range values while maintaining the same average emigration probability within country, age, and sex groupings. In the equations below,  $e^u$  is the unadjusted estimated probability of emigration, and  $x$ ,  $m$ , and  $n$  are the mean, maximum, and minimum values of  $e$  within each country, age, and sex group, respectively. Then we estimate an adjusted estimate of emigration,  $e^{a1}$ , separately for each group as:

$$e^{a1} = x + (e^u - x) * S$$

$$\text{where } S = \min[(1-x)/(m-x), (x)/(x-n)].$$

Note that all the estimates are adjusted, not just the out-of-range estimates. Also, the minimum, maximum, and mean are evaluated within each country, age, and sex group and then the formula is applied in order to pull in the out of range values within each group. This procedure forces the emigration estimates to fall between zero and one but does not change the average of the estimate, although net emigration estimates may still be negative if return immigration estimates turn out to be larger than emigration (this can result from measurement or sampling error).

Another potential problem is that people who are followed up can be estimated as having non-zero probabilities of emigration, mortality, internal migration, and non-response. Therefore, after making the first adjustment to emigration (described above), we shifted all non-zero probabilities of emigration, internal migration, mortality, and non-response to cases that actually were not followed up in  $t+1$ . Again, this is done while still maintaining the same average probabilities within country, age, and sex groupings. This was accomplished by scaling up the probabilities among non-follow-ups. For each group, we use the algorithm below to estimate the adjusted probabilities of emigration, death, internal migration, and non-response ( $e^{a2}$ ,  $d^{a2}$ ,  $m^{a2}$ , and  $r^{a2}$ , respectively). Thus, in the case of internal migration:

$$m^{a2} = m^u (s_m^T / s_m^{nm}) \text{ if NF} = 1;$$

$$m^{a2} = 0 \text{ if NF} = 0,$$

where  $m^u$  is the unadjusted probability,  $s_m^T$  is the weighted sum of  $m^u$  for the entire group,  $s_m^{nm}$  is the weighted sum of  $m^u$  over the cases in the group that were not followed up, and NF is a dummy variable indicating whether the case was successfully followed up (1 = not followed up, 0 = followed up). We make this adjustment in a similar manner for the other components:

#### Death

$$d^{a2} = d^u (s_d^T / s_d^{nm}) \text{ if NF} = 1;$$

$$d^{a2} = 0 \text{ if NF} = 0;$$

#### Non-response

$$r^{a2} = r^u (s_r^T / s_r^{nm}) \text{ if NF} = 1;$$

$$r^{a2} = 0 \text{ if NF} = 0.$$

When we adjusted the emigration probabilities, the value of the unadjusted probability is taken from  $e^{a1}$  (referred to in the previous step):

$$e^{a2} = e^{a1} (s_{e(a1)}^T / s_{e(a1)}^{nm}) \text{ if NF} = 1;$$

$$e^{a2} = 0 \text{ if NF} = 0.$$

A third problem is that the components of non-response among those not followed up may not add up to 100%. Therefore, after making the above two adjustments, we forced the components of non-follow-up (probabilities of emigration, internal migration, death, non-response among non-follow-ups) to sum to 100%. This is accomplished by rescaling the internal migration (m), death (d), non-response (r) components proportionately, but not changing the estimate of emigration (e). Thus for cases that were not successfully followed up, we make the following adjustments:

$$d^{a3} = d^{a2} / (d^{a2} + m^{a2} + r^{a2}) * (1 - e^{a2})$$

$$m^{a3} = m^{a2} / (d^{a2} + m^{a2} + r^{a2}) * (1 - e^{a2})$$

$$r^{a3} = r^{a2} / (d^{a2} + m^{a2} + r^{a2}) * (1 - e^{a2})$$

where the a3 superscript indicates the adjusted probabilities and the a2 superscript indicates the probabilities from the second adjustment described above. This adjustment forces the components for the final emigration estimates to be exactly equal to the percentage non-matched minus the sum of the percentage of internal migrants, deaths, and non-responses.

## RESULTS

Among the foreign-born in the 1996-2005 CPSs, 34.2% were not successfully followed up. We estimate that 14.8% were not followed up because they moved to another residence in the U.S., 0.8% died, 14.4% were non-responders, and 4.3% emigrated (with a standard error of 0.1% and 95% confidence interval ranging from 4.1 to 4.4 (Table 4). For a population of 29,988 thousand foreign-born (as in the March 2000 CPS), this rate translates into roughly 1,276 thousand emigrants per year. At the same time, we estimate a return immigration rate of 1.0% ( $\pm 0.1\%$ ), translating into about 300 thousand return immigrants annually. Subtracting return immigration from total emigration yields annual net emigration of 3.3 percent or 976 thousand net emigrants per year.

The estimated emigration rates fluctuate from year to year (Table 5). For example, net emigration appears lowest in 2000 (2.1%; 1.6% among Mexicans and 2.3% among non-Mexicans) and highest in 2002 (4.5%; 5.2% among Mexicans and 4.2% among non-Mexicans).

Emigration and return immigration tend to increase with age until ages 20-24, after which emigration and return immigration generally decline (Table 6). Taking emigration and return immigration together, net emigration appears to be lowest for children ages 0-14 (1.9%) and highest for young adults, reaching 6.1% among those age 20-24, 4.4% among those age 25-29, and 4.0% among those age 30-34.

Males are more likely to emigrate than females—5.3 percent versus 3.2 percent—and are significantly more likely to be return immigrants (1.2 versus 0.7%), but not enough to offset their significantly higher emigration rates (Table 7). Net male emigration (4.0%) remains significantly higher than net female migration (2.5%). The gender difference in net emigration is greater among Mexicans (5.3% vs. 2.2%) than non-Mexicans (3.4% vs. 2.6%). Age patterns of net emigration also vary by gender. Among males, net emigration peaks at ages 20-24 (8.2%), but is also high among teenagers ages 15-19 (6.5%), and adults ages 25-34 (4.2%), 35-39 (5.0%), and 40-44 (4.1%) (Table 8). Net emigration is particularly high among working-aged Mexican men; the age pattern is much flatter among non-Mexican men (Tables 10 and 11). Emigration rates tend to be lower for females across most age groups except among girls ages 0-14 (2.2% compared with 1.6% among boys) (Table 9). Further inspection shows that net emigration is particularly low among Mexican girls (0.9% compared with 3.1% among non-Mexican girls) but high among Mexican women age 65+ (5.2%) (Table 10).

When we examine the emigration rates by duration of residence in the United States, we find that, in general, emigration rates are highest for recent arrivals and decline significantly with time in the United States (Table 12). Return immigration rates are higher for recent arrivals (0-4 years in the country) than earlier arrivals (5-9 and 10+ years in the country), suggesting that circular migration is more common among recent arrivals.

There is considerable variability in emigration by country or region of birth (Table 13). Foreign born from Africa are estimated as having very high emigration and net emigration rates (so high in fact that we are inclined to doubt the accuracy of the estimate). Other countries with emigration rates of 5% or higher include Mexico (5.3%), North America (primarily Canada) (5.8%), and India (9.5%). Countries with emigration rates of 4.0% to 4.9% include Dominican Republic (4.0%), and Europe (4.8%). Of these high-emigration groups, the relatively high return immigration rates for Mexicans and Dominicans stand out as they are about 50% or more higher

than any of the other countries or regions. This distinction reflects the circular migration patterns that are commonly observed for Mexican and Dominican migrants. On balance then, Mexicans and Dominicans do not have the highest net emigration rates. For example, net emigration appears to be higher than Mexicans among those from North America, Europe, India, and Africa.

Emigration estimates are highest among legal non-immigrants (10.2%), followed by unauthorized migrants (5.7%), legal permanent residents (4.5%), refugees (2.8%) and naturalized citizens (2.5%) (Table 14). The rank ordering by legal status is the same for Mexicans and non-Mexicans (although we do not produce estimates for Mexican refugees and legal non-immigrants since there were so few identified in the data). Return immigration rates follow the same ordering: legal non-immigrants (2.6%), followed by unauthorized migrants (1.5%), Legal Permanent Residents (1.2%), refugees (0.7%) and naturalized citizens (0.3%). Thus circular migration patterns emerge most prominently for those with temporary or irregular legal statuses, such as in the case of legal non-immigrants and unauthorized migrants.

Finally, emigration and net emigration rates appear to be higher for the 2000-2004 period than the 1996-1999 period among Mexican unauthorized migrants (among whom net emigration is estimated at 3.7% in the earlier period and 5.0% in the later period) and among non-Mexican legal non-immigrants (among whom net emigration is estimated at 5.3% in the earlier and 9.6% in the later period) (Table 15).

## CONCLUSIONS

We have developed and revised a new method for estimating foreign-born emigration using data hitherto untapped for that purpose. This involves mainly using year-to-year matched data files from the March Supplements to the Current Population Survey. The results from this method presented here demonstrate the feasibility of the approach for producing emigration estimates for a multi-year time period from several pooled CPS samples from adjacent years. They also include detailed estimates by year, age, sex, country of birth, duration in the U.S., and migration status using data from the 1996–2005 period.

The estimates tend to be higher than those produced by the residual method. For example, we estimate a net emigration rate of 3.3%, but residual-based estimates tend to close to 1.0% (Warren and Peck 1980: 1.2%; Ahmed and Robinson 1994: 1.2%; and Mulder 2003: 0.9%). This suggests a much higher level of movement to and from the United States among the foreign born than has been assumed in official Census Bureau estimates, but which is nevertheless consistent with other sources about the circularity of migration flows among Mexican, Dominican, and other Latin American immigrants, and the short duration of stays of legal temporary migrants including those with work or student visas. For example, Jasso and Rosenzweig (1982) estimated emigration rates for the 1971 immigrant admission cohort (based on follow-up rates in the Alien Address Report Program in January 1979). They found much higher rates of emigration than residual-based estimates at the time, which were about 1.0%. For example, the annual emigration rates were estimated to be higher than 2.1% and as high as 9.2% for Mexicans. For no single national origin group did the lower boundary of the estimate dip below 1.0%, and the middle estimates ranged from 1.0% (China) to 13.5% (South America).

If emigration really is as high as 3% (as we estimate here), this would mean that current population estimates, particularly of the foreign born, are probably inaccurate. First, the population estimates produced in the Census Bureau's Population Estimates Series would be too high, particularly for Hispanics. This would, in turn, mean that the sampling weights for the CPS and other Census surveys would be too high. Higher emigration rates would also have implications for estimates of the unauthorized foreign-born population, albeit in complex ways. Residual-based estimates of the unauthorized foreign-born population require independent estimates (that is, independent of the Census or CPS) of the legally-resident foreign-born population. This estimate is constructed through the demographic analysis of legal admission records and estimates of mortality and emigration. The estimate of the legally-resident population would be too high if emigration is underestimated, which in turn would mean that the residual (the unauthorized component) would be overestimated. But at the same time, if emigration is underestimated and the estimate of the total foreign-born population is based on the CPS, the sampling weights could be too high (particularly for Hispanics). This would mean that the total foreign-born population used in residual-based estimates would be too high, and the estimate of the unauthorized population would be too low. In other words, changes in the level of emigration would have counter-balancing effects, one working through the estimates of the legal population and another operating through the CPS sampling weights. At this moment, the net effect is unknown; more work would be required to assess what the net effect on estimates of the unauthorized population would be.

Further development and assessment of the underlying assumptions of the CPS-matching method is warranted. One key parameter that permits us to solve the multi-equation system of relationships is the assumption that emigration is negligible for second generation adults. It is important that future research explore the sensitivity of the foreign-born emigration estimates to various assumptions about the level of second generation adult emigration. However, to the degree that emigration for the second generation is non-negligible, or becomes so, then the foreign-born emigration estimates of the kind introduced here will actually be too low—thus indicating that the residual-based estimates may underestimate foreign-born emigration to an even greater extent than suggested here. Another key assumption is that the factors associated with non-response operate in the same manner for the second generation as for the first generation. This assumption may not hold for certain groups. For example, Mexican foreign-born may be more likely to avoid CPS interviewers than U.S.-born Mexican-Americans because large portions of Mexican immigrants are unauthorized. If we were to underestimate non-response among the foreign-born, this would cause emigration to be overestimated.

Nevertheless, the CPS-matching method has great potential. Emigration estimates developed with the CPS-matching method could be incorporated into other applications including national (and subnational) population estimates and residual-based methods for estimating unauthorized migration (e.g., Passel et al. 2004a). In particular, the CPS-matching method could be used to generate emigration statistics for recently-arrived foreign born, large numbers of whom are not legal permanent residents and thus likely to have much higher emigration rates than earlier arrivals. In addition, we believe that the CPS-matching method could be used to shed light on the type and degree of selectivity related to emigration because the method permits the estimation of emigration rates for large foreign-born subgroups, such as the unemployed, welfare recipients, and persons in relatively good or poor health. The development of estimates for such groups could only enhance the scientific and policy relevance of the procedure introduced here.

## REFERENCES

- Ahmed, B. and J.G. Robinson. 1994. "Estimates of Emigration of the Foreign-born Population: 1980-1990." Population Division Working Paper No. 9, U.S. Census Bureau. Washington, DC. (<http://www.census.gov/population/www/techpap.html>).
- Bean, Frank D., Rodolfo Corona, Rodolfo Tuiran, Karen Woodrow-Lafield, and Jennifer Van Hook. 2001. "Circular, Invisible, and Ambiguous Migrants: Components of Difference in Estimates of the Number of Unauthorized Mexican Migrants in the United States." *Demography* 38(3): 411-422.
- Duleep, Harriet 1994. "Social Security and the Emigration of Immigrants." *Social Security Bulletin* 57(1): 37-51.
- Fernandez, Edward W. 1995. "Estimates of the Annual Emigration of U.S. Born Persons by using Foreign Censuses and Selected Administrative Data: Circa 1980." Population Division Technical Working Paper No. 10. U.S. Census Bureau: Washington, DC. (<http://www.census.gov/population/www/techpap.html>).
- Gibbs James, Gregory Harper, Marc Rubin, and Hyon Shin. 2003. "Evaluating Components of International Migration: Native-Born Emigrants." Population Division Working Paper No. 63. (<http://www.census.gov/population/www/techpap.html>).
- Jasso, Guillermina and Mark R. Rosenzweig. 1982. "Estimating the Emigration Rates of Legal Immigrants Using Administrative and Survey Data: The 1971 Cohort of Immigrants to the United States." *Demography* 19(3): 279-290.
- Kraly, Ellen Percy. 1998. "Emigration: Implications for U.S. Immigration Policy Research." Pp. 587-618 in *Migration Between Mexico and the United States: Binational Study, Vol. 2 Research Reports and Background Materials*. Mexican Ministry of Foreign Affairs and the U.S. Commission on Immigration Reform: Mexico City and Washington, DC.
- Madrian, Brigitte C. and Lars John Lefgren. 1999. "An Approach to Longitudinally Matching the Current Population Survey." NBER Technical Working Paper No. 247. National Bureau of Economic Research. November. (<http://papers.nber.org/papers/T0247>)
- Massey, Douglas S., Jorge Durand, and Nolan J. Malone. 2002. *Beyond Smoke and Mirrors: Mexican Migration in an Era of Economic Integration*. Russell Sage Foundation: New York.
- Mulder, Tammany J., Betsy Guzmán, and Angela Brittingham. 2002. "Evaluating Components of International Migration: Foreign-Born Emigrants." Demographic Analysis-Population Estimates (DAPE) Research Project, Population Division Working Paper No. 62. U.S. Bureau of the Census. Washington, DC. April. (<http://www.census.gov/population/www/documentation/twps0062.html>)
- Mulder, Tammany. 2003. "Foreign-born Emigration from the United States: 1990 to 2000." Paper presented at the 2003 annual meetings of the Population Association of America, Minneapolis, MN.

- National Center for Health Statistics (NCHS). 2000. "National Health Interview Survey Multiple Cause of Death Files—2000 Linkage." Hyattsville, MD: National Center for Health Statistics.
- Palloni, Alberto and Elizabeth Arias. 2004. "Paradox Lost: Explaining the Hispanic Adult Mortality Advantage." *Demography* 41(3): 385-416.
- U.S. Census Bureau. 2002a. *Current Population Survey: Design and Methodology*. Technical Paper No. 63RV. Washington, DC. March. (<http://www.bls.census.gov/cps/tp/tp63.htm>)
- U.S. Census Bureau. 2002b. "Annual Demographic Survey March Supplement: 2002 Methodology and Documentation." Washington, DC: U.S. Census Bureau. (<http://www.bls.census.gov/cps/ads/adsmain.htm>).
- Van Hook, J., F.D. Bean, and J.S. Passel. 2005. "Foreign-born Emigration: A New Approach and Estimates Based on Matches CPS Files." SABRE Systems.
- Van Hook, J., W. Zhang, F.D. Bean, and J.S. Passel. 2006. "Foreign-born Emigration: A New Approach and Estimates Based on Matches CPS Files," *Demography* 43: 361-382.
- Van Hook, Jennifer and Frank D. Bean. 1998. "Estimating Unauthorized Migration to the United States: Issues and Results," pp. 511-550 in *Migration Between Mexico and the United States, Research Reports and Background Materials*, Mexico City and Washington, D.C.: Mexican Ministry of Foreign Affairs and U.S. Commission on Immigration Reform.
- Warren, Robert and Jennifer M. Peck. 1980. "Foreign-born Emigration from the United States: 1960 to 1970." *Demography* 17(1): 71-84.
- Woodrow-Lafield, Karen. 1996. "Emigration from the United States: Multiplicity Survey Evidence." *Population Research and Policy Review*. 15: 171-199.
- Woodrow-Lafield, Karen. 1998. "Viewing Emigration at Century's End." Pp. 683-694 in *Migration Between Mexico and the United States: Binational Study, Vol. 2 Research Reports and Background Materials*. Mexican Ministry of Foreign Affairs and the U.S. Commission on Immigration Reform: Mexico City and Washington, DC.

Table 1. Logistic Regression Models of Internal Migration by Mexican-origin, Sex, and Generational Status

	Mexican				Non-Mexican			
	Men		Women		Men		Women	
	Gen 1	Gen 2	Gen 1	Gen 2	Gen 1	Gen 2	Gen 1	Gen 2
In oversample	-0.915 ***	-1.000 ***	-0.988 ***	-1.119 ***	-0.779 ***	-0.681 ***	-0.740 ***	-0.710 ***
Homeowner	-0.810 ***	-1.457 ***	-0.867 ***	-1.650 ***	-0.935 ***	-1.482 ***	-0.953 ***	-1.476 ***
Cuba					0.204	0.363 *	0.231	0.065
China					0.042	-0.025	-0.019	-0.161
India					0.134	0.475 *	0.266 *	0.300
Other Asia					0.087	0.081	0.138 *	-0.030
Africa					0.076	0.093	0.068	-0.108
Europe					-0.193 *	0.179 *	-0.121	-0.022
N America					0.264	0.178	0.175	0.201
Age 15-24	1.012 ***	1.072 ***	1.585 ***	1.485 ***	1.697 ***	1.817 ***	1.675 ***	1.859 ***
Age 25-34	0.732 ***	1.010 ***	1.099 ***	1.268 ***	1.470 ***	1.734 ***	1.442 ***	1.837 ***
Age 35-44	0.386	1.005 **	0.764 ***	0.988 ***	1.031 ***	1.329 ***	1.048 ***	1.205 ***
Age 45+	0.081	0.109	0.398	0.239	0.580 ***	0.753 ***	0.553 ***	0.714 ***
1997	-0.239	-0.019	-0.144	-0.104	-0.039	0.061	0.189	-0.112
1998	-0.293 *	0.255	-0.216	-0.163	-0.057	-0.008	0.044	0.033
1999	-0.287 *	0.004	-0.368 *	-0.216	0.089	0.034	0.165	0.010
2000	-0.121	-0.248	-0.178	-0.180	0.036	0.051	0.106	0.094
2001	-0.731 ***	-0.729 **	-0.517 ***	-0.379	-0.235 *	-0.017	-0.015	-0.264 *
2002	-0.566 ***	0.005	-0.461 ***	-0.336	-0.220 *	0.157	0.069	-0.081
2003	-0.423 ***	-0.150	-0.455 ***	-0.340	-0.264 **	-0.170	-0.165	-0.133
2004	-0.666 ***	-0.192	-0.598 ***	-0.467 *	-0.184	-0.040	0.033	-0.017
Enrolled in HS	-0.516 **	-0.701 ***	-0.507 **	-1.036 ***	-0.927 ***	-0.831 ***	-0.565 ***	-0.933 ***
Enrolled in College	-0.831 **	-0.344	-0.711 **	-0.148	-0.442 ***	-0.408 **	-0.157	-0.488 ***
Intercept	-0.997 ***	-1.256 ***	-1.490 ***	-1.280 ***	-2.087 ***	-2.321 ***	-2.384 ***	-2.098 ***
N	12,198	4,754	10,827	5,212	21,915	18,286	25,315	20,301
Pseudo R-square	0.082	0.145	0.098	0.181	0.087	0.146	0.088	0.157

Table 2. Logistic Regression Models of Matching Across Adjacent Years (t to t+1) in March CPS, by Mexican-origin, Sex or Age Group, and Generational Status

	Mexican					Non-Mexican				
	Men		Women		Children	Men		Women		Children
	Gen 1	Gen 2	Gen 1	Gen 2	Gen 1	Gen 1	Gen 2	Gen 1	Gen 2	Gen 1
In oversample	-0.073	-0.073	-0.079	-0.136 *	-0.425	-0.396 ***	-0.447 ***	-0.362 ***	-0.488 ***	0.033
Homeowner	1.172 ***	1.101 ***	1.023 ***	1.014 ***	0.216	0.986 ***	1.020 ***	0.906 ***	1.007 ***	0.518
Cuba						0.112	0.094	0.104	0.154	-0.298
China						-0.092	0.068	-0.216 **	-0.141	0.918
India						-0.093	0.098	0.013	-0.145	1.374 *
Other Asia						0.058	-0.116	-0.029	-0.195 *	0.105
Africa						-0.260 **	0.577	-0.299 **	0.106	0.168
Europe						0.024	0.054	-0.049	-0.088	0.496
N America						-0.238 *	0.079	-0.113	-0.109	0.844
Male					0.172					-0.066
Age 0-4					-0.012					-0.573
Age 5-9					-0.155					-0.097
Age 15-24	-0.892 ***	-0.508 **	-0.820 ***	-0.967 ***		-0.755 ***	-0.490 ***	-0.837 ***	-0.967 ***	
Age 25-34	-0.334 **	-0.419 **	-0.273 *	-0.666 ***		-0.544 ***	-0.514 ***	-0.493 ***	-0.793 ***	
Age 35-44	0.003	-0.128	0.122	-0.268		-0.181 **	-0.114	-0.163 **	-0.388 ***	
Age 45-64	0.223	0.133	0.180	-0.150		0.058	0.265 ***	0.069	0.057	
1997	-0.147	-0.143	-0.088	-0.110	-0.069	-0.184 *	-0.158	0.012	-0.108	-1.238
1998	0.087	-0.290	0.020	-0.075	-0.246	-0.129	-0.089	0.018	-0.115	0.069
1999	-0.078	-0.226	0.124	-0.005	0.399	-0.083	0.024	0.042	-0.017	0.306
2000	-0.172	-0.287	-0.064	-0.237	-0.468	-0.027	-0.037	0.081	-0.150	-0.334
2001	-0.121	0.117	0.136	-0.032	-0.317	-0.146	-0.160	-0.030	-0.079	-0.019
2002	-0.191 *	-0.079	0.040	0.044	-0.209	-0.169 *	-0.161	-0.034	-0.132	-0.261
2003	-0.194 *	-0.305	-0.138	-0.150	-0.922 *	-0.503 ***	-0.724 ***	-0.323 ***	-0.565 ***	-0.457
2004	-0.642 ***	-0.748 ***	-0.539 ***	-0.541 ***	-0.968 *	-0.476 ***	-0.614 ***	-0.350 ***	-0.466 ***	-1.599 **
Enrolled in HS	0.660 ***	0.287 *	0.399 **	0.199		0.686 ***	0.483 ***	0.446 ***	0.619 ***	
Enrolled in College	1.090 ***	0.232	0.638 **	0.266		0.166	-0.028	0.217 *	-0.014	
No HS degree <sup>1</sup>	0.031	0.161	0.342 **	0.214	-0.069	-0.057	0.083	0.156 **	-0.130	0.081
HS diploma	0.027	-0.032	0.327 **	0.179	-0.190	0.071	-0.058	0.098 *	-0.136 *	0.704 *
Some college	-0.030	0.030	0.214	0.033	-0.119	0.013	-0.026	0.097	-0.038	0.174
Parent was followed up					7.033 ***					7.192 ***
Parent Age 15-29					1.522					-0.845
Parent Age 30-39					1.443					-0.372
Parent Age 40-49					1.609 *					-0.308
Parent Age 50-59					1.621					0.615
Intercept	0.248	0.450	0.166	0.720 **	-4.239 ***	0.672 ***	0.727 ***	0.689 ***	1.149 ***	-3.149 **
N	12,612	4,766	11,068	5,234	2,442	22,592	18,334	26,014	20,360	2,763
Pseudo R-square	0.093	0.072	0.077	0.073	0.807	0.074	0.072	0.062	0.081	0.794

<sup>1</sup>respondent's education for men's and women's models, parent's education for children's models.

Table 3. Discrete-time Event History Models of Mortality by Nativity (Logistic Regression Coefficients)

	Immigrants	Natives
Intercept	-2.017 ***	-1.711 ***
Male	0.517 ***	0.531 ***
<u>Race/Ethnicity</u>		
Mexican	0.336 ***	0.044
Other Hispanic	0.067	0.204 ***
NH White	0.230 ***	0.205 ***
Black (Other)	0.176	0.297 ***
<u>Age</u>		
18-24	-3.984 ***	-4.042 ***
25-34	-3.722 ***	-3.782 ***
35-44	-3.324 ***	-3.149 ***
45-54	-2.756 ***	-2.330 ***
55-64	-1.818 ***	-1.535 ***
65-74 (75+)	-0.994 ***	-0.819 ***
<u>General Health Status</u>		
Excellent	-1.181 ***	-1.711 ***
Very Good	-1.135 ***	-1.501 ***
Good	-0.937 ***	-1.127 ***
Fair (Poor)	-0.586 ***	-0.653 ***
Number of person-years	344,536	2,767,340

Source: 1989-1994 National Health Interview Surveys linked to the 1989-2004 National Death Index

\*\*\* p<.001 \*\* p<.01 \*p<.05

Reference categories are noted in parentheses.

Table 4. Foreign-born Emigration (Estimates and Components), Based on CPS Matching Method

Component	Percentage	SE	90% Confidence Intervals		
			Low	High	
(1) Non-match	34.2	0.3	33.8	34.6	
(2) Internal Migration	14.8	0.2	14.5	15.1	
(3) Mortality	0.8	0.0	0.7	0.9	
(4) Non-follow-up	14.4	0.2	14.0	14.7	
(5) =(1)-(2)-(3)-(4)	Emigration	4.3	0.1	4.1	4.4
(6) Return Immigration	1.0	0.1	0.9	1.1	
(7) = (5) - (6)	Net Emigration	3.3	0.1	3.1	3.5

Notes. Emigration rates are based on analysis of pairs of matched CPS files from 1996-97 through 2004-05. Net emigration is equal to emigration minus "return immigration." Non-match, internal migration, mortality, and non-follow-up are components of the formula used to estimate emigration. The standard errors of the components are estimated using the "b" factors for Hispanics in the Source and Accuracy statement of the CPS documentation. The standard error of the net immigration rate is estimated as the standard error of difference of emigration and return immigration.

Table 5. Foreign-born Emigration Based on CPS Matching Method, by Year and Mexican Origin

Year*	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
<b>All Foreign-born</b>							
1996	4.5 3.8, 5.1	1.2 0.9, 1.5	3.3 2.6, 4.0	32.3 30.9, 33.7	17.3 16.2, 18.5	0.7 0.5, 1.0	9.8 8.9, 10.6
1997	4.2 3.6, 4.8	1.2 0.9, 1.5	3.0 2.3, 3.7	33.2 31.9, 34.6	16.3 15.2, 17.4	0.7 0.4, 0.9	12.0 11.1, 13.0
1998	3.9 3.3, 4.4	1.1 0.8, 1.4	2.7 2.1, 3.4	32.1 30.7, 33.4	15.5 14.5, 16.6	0.7 0.5, 0.9	12.0 11.0, 12.9
1999	4.1 3.6, 4.7	1.0 0.7, 1.3	3.1 2.5, 3.8	31.4 30.0, 32.7	16.1 15.0, 17.2	0.7 0.4, 0.9	10.5 9.6, 11.3
2000	3.3 2.8, 3.8	1.3 1.0, 1.5	2.1 1.5, 2.6	32.0 30.8, 33.2	16.1 15.2, 17.1	0.7 0.5, 0.9	11.8 11.0, 12.7
2001	4.6 4.0, 5.2	1.0 0.7, 1.3	3.6 3.0, 4.2	33.2 31.9, 34.5	13.5 12.6, 14.4	0.7 0.5, 1.0	14.4 13.4, 15.3
2002	5.3 4.7, 5.9	0.8 0.5, 1.0	4.5 3.9, 5.2	33.1 31.8, 34.3	13.9 13.0, 14.8	0.7 0.5, 0.9	13.2 12.3, 14.1
2003	3.7 3.2, 4.2	0.7 0.5, 0.9	3.0 2.5, 3.6	37.7 36.4, 38.9	13.1 12.2, 14.0	1.3 1.0, 1.6	19.6 18.5, 20.6
2004	4.7 4.2, 5.3	0.7 0.5, 1.0	4.0 3.4, 4.6	40.8 39.6, 42.1	12.4 11.6, 13.3	0.8 0.6, 1.0	22.9 21.8, 24.0
<b>Mexican</b>							
1996	4.9 3.7, 6.2	1.9 1.1, 2.7	3.0 1.6, 4.5	38.5 35.7, 41.3	22.7 20.3, 25.1	0.5 0.1, 1.0	10.4 8.6, 12.1
1997	6.2 4.9, 7.6	1.7 0.9, 2.4	4.6 3.1, 6.1	40.7 38.0, 43.4	20.4 18.2, 22.7	0.4 0.1, 0.8	13.6 11.7, 15.5
1998	4.2 3.1, 5.3	1.3 0.6, 1.9	2.9 1.6, 4.2	35.6 32.9, 38.3	17.7 15.6, 19.9	0.5 0.1, 0.9	13.2 11.3, 15.1
1999	4.8 3.6, 5.9	1.4 0.7, 2.0	3.4 2.0, 4.7	36.0 33.4, 38.7	16.7 14.6, 18.7	0.4 0.1, 0.8	14.2 12.3, 16.1
2000	3.5 2.6, 4.4	1.9 1.2, 2.6	1.6 0.5, 2.7	40.6 38.2, 43.0	20.0 18.1, 22.0	0.5 0.1, 0.8	16.6 14.8, 18.4
2001	5.5 4.4, 6.6	1.0 0.6, 1.5	4.4 3.2, 5.6	38.1 35.8, 40.4	15.4 13.7, 17.1	0.4 0.1, 0.7	16.9 15.1, 18.7
2002	6.8 5.6, 8.0	1.6 1.0, 2.2	5.2 3.9, 6.6	37.7 35.4, 40.0	16.0 14.3, 17.8	0.5 0.1, 0.8	14.3 12.7, 16.0
2003	5.0 4.0, 5.9	0.8 0.4, 1.3	4.1 3.1, 5.2	39.6 37.4, 41.8	17.1 15.5, 18.8	0.7 0.3, 1.1	16.8 15.1, 18.5
2004	6.4 5.3, 7.5	1.0 0.5, 1.4	5.4 4.2, 6.6	49.6 47.3, 51.9	13.3 11.7, 14.8	0.7 0.3, 1.0	29.3 27.2, 31.3
<b>Non-Mexican</b>							
1996	4.3 3.6, 5.0	0.9 0.6, 1.3	3.4 2.6, 4.2	29.9 28.3, 31.5	15.3 14.1, 16.6	0.8 0.5, 1.1	9.5 8.5, 10.6
1997	3.4 2.8, 4.1	1.1 0.7, 1.4	2.4 1.7, 3.1	30.4 28.9, 32.0	14.8 13.6, 16.0	0.8 0.5, 1.1	11.5 10.4, 12.5
1998	3.7 3.1, 4.4	1.1 0.7, 1.4	2.6 1.9, 3.4	30.8 29.2, 32.4	14.8 13.6, 16.0	0.8 0.5, 1.1	11.5 10.5, 12.6
1999	3.9 3.3, 4.6	0.9 0.6, 1.2	3.0 2.3, 3.7	29.6 28.1, 31.1	15.9 14.6, 17.1	0.8 0.5, 1.1	9.0 8.1, 10.0
2000	3.2 2.7, 3.8	1.0 0.7, 1.3	2.3 1.6, 2.9	28.4 27.0, 29.9	14.5 13.4, 15.6	0.8 0.5, 1.1	9.9 9.0, 10.8
2001	4.2 3.5, 4.8	1.0 0.6, 1.3	3.2 2.5, 3.9	30.9 29.4, 32.4	12.6 11.5, 13.7	0.9 0.6, 1.2	13.2 12.1, 14.3
2002	4.6 3.9, 5.3	0.4 0.2, 0.6	4.2 3.5, 4.9	30.9 29.4, 32.4	12.9 11.8, 14.0	0.8 0.5, 1.1	12.6 11.5, 13.7
2003	3.1 2.5, 3.6	0.6 0.3, 0.8	2.5 1.9, 3.1	36.7 35.2, 38.2	11.1 10.1, 12.1	1.6 1.2, 2.0	21.0 19.7, 22.2
2004	3.9 3.3, 4.5	0.6 0.4, 0.9	3.3 2.6, 3.9	36.7 35.2, 38.2	12.1 11.0, 13.1	0.8 0.5, 1.1	19.9 18.6, 21.1

Notes. Emigration rates are based on analysis of pairs of matched CPS files from 1996-97 through 2004-05. Net emigration is equal to emigration minus "return immigration." Non-match, internal migration, mortality, and non-follow-up are components of the formula used to estimate emigration. The smaller numbers shown beneath the estimates are the lower and upper end of the 90% CI. These are based on standard errors estimated using the "b" factors for Hispanics in the Source and Accuracy statement of the CPS documentation. The standard error of the net immigration rate is estimated as the standard error of difference of emigration and return immigration.

\*Estimates for 2005 are not shown because 2006 data (t+1) are necessary to produce estimates for 2005 (t).

Table 6. Foreign-born Emigration Based on CPS Matching Method, by Age Group

Age Group	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
0-14	3.5 2.9, 4.2	1.7 1.2, 2.1	1.9 1.0, 2.7	35.6 33.9, 37.4	19.5 17.7, 21.2	0.0 0.0, 0.0	12.6 11.4, 13.9
15-19	4.7 3.9, 5.5	1.1 0.7, 1.5	3.6 2.7, 4.6	40.7 38.8, 42.6	18.5 17.0, 20.0	0.1 0.0, 0.3	17.3 15.8, 18.8
20-24	8.0 7.1, 8.8	1.9 1.5, 2.3	6.1 5.1, 7.0	53.3 51.7, 54.8	25.6 24.3, 27.0	0.2 0.0, 0.3	19.5 18.3, 20.7
25-29	5.9 5.3, 6.6	1.5 1.1, 1.8	4.4 3.7, 5.2	47.5 46.1, 48.9	24.0 22.8, 25.2	0.2 0.1, 0.3	17.4 16.4, 18.5
30-34	5.2 4.6, 5.7	1.1 0.9, 1.4	4.0 3.4, 4.7	38.6 37.3, 39.9	18.9 17.9, 19.9	0.2 0.1, 0.3	14.3 13.4, 15.3
35-39	4.2 3.6, 4.7	1.1 0.8, 1.3	3.1 2.5, 3.7	32.7 31.5, 34.0	14.3 13.4, 15.2	0.2 0.1, 0.3	14.0 13.1, 15.0
40-44	3.3 2.8, 3.8	0.7 0.4, 0.9	2.6 2.0, 3.2	30.1 28.7, 31.4	12.9 12.0, 13.9	0.3 0.1, 0.5	13.5 12.6, 14.5
45-49	2.9 2.4, 3.4	0.6 0.3, 0.8	2.3 1.7, 2.9	25.8 24.4, 27.2	8.8 7.9, 9.7	0.3 0.1, 0.5	13.8 12.7, 14.9
50-54	2.9 2.3, 3.5	0.5 0.3, 0.8	2.4 1.8, 3.0	25.2 23.7, 26.7	8.6 7.6, 9.6	0.6 0.3, 0.9	13.1 11.9, 14.3
55-59	2.7 2.0, 3.3	0.4 0.2, 0.7	2.3 1.6, 3.0	21.7 20.0, 23.3	7.3 6.3, 8.4	0.5 0.2, 0.8	11.1 9.9, 12.4
60-64	3.2 2.4, 3.9	0.5 0.2, 0.9	2.6 1.8, 3.5	22.5 20.6, 24.3	7.6 6.4, 8.8	1.5 1.0, 2.1	10.1 8.8, 11.5
65-69	2.5 1.8, 3.3	0.6 0.2, 0.9	2.0 1.1, 2.8	19.8 17.8, 21.8	3.8 2.9, 4.8	1.2 0.6, 1.7	12.3 10.6, 13.9
70-74	2.7 1.8, 3.6	0.6 0.2, 1.1	2.1 1.1, 3.1	22.1 19.8, 24.4	4.2 3.1, 5.3	3.0 2.1, 4.0	12.1 10.3, 13.9
75+	3.1 2.4, 3.8	0.2 0.0, 0.3	2.9 2.2, 3.7	29.1 27.1, 31.0	5.7 4.7, 6.7	8.1 7.0, 9.3	12.2 10.8, 13.6

Notes: See Table 5.

Table 7. Foreign-born Emigration Based on CPS Matching Method, by Sex and Mexican Origin

Sex and Mexican Origin	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
<b>All</b>							
Female	3.2 3.0, 3.5	0.7 0.6, 0.8	2.5 2.3, 2.8	31.2 30.6, 31.8	13.7 13.1, 14.3	0.7 0.6, 0.8	13.6 13.1, 14.0
Male	5.3 5.0, 5.6	1.2 1.1, 1.4	4.0 3.7, 4.3	37.1 36.5, 37.8	15.9 15.4, 16.4	0.8 0.7, 1.0	15.1 14.7, 15.6
<b>Mexican</b>							
Female	3.1 2.7, 3.6	0.9 0.7, 1.1	2.2 1.7, 2.7	35.7 34.5, 36.9	16.4 15.5, 17.3	0.5 0.3, 0.6	15.7 14.8, 16.7
Male	7.1 6.5, 7.6	1.7 1.4, 2.0	5.3 4.7, 6.0	43.4 42.3, 44.5	18.2 17.3, 19.0	0.6 0.4, 0.7	17.6 16.7, 18.4
<b>Non-Mexican</b>							
Female	3.3 3.0, 3.5	0.6 0.5, 0.8	2.6 2.3, 2.9	29.6 28.9, 30.2	12.7 12.2, 13.2	0.8 0.7, 1.0	12.8 12.3, 13.3
Male	4.4 4.1, 4.7	1.0 0.9, 1.2	3.4 3.0, 3.7	34.1 33.3, 34.8	14.8 14.2, 15.3	1.0 0.8, 1.1	13.9 13.4, 14.5

Notes: See Table 5.

Table 8. Foreign-born Emigration Based on CPS Matching Method Among Males, by Age Group

Age Group	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
0-14	3.5 2.6, 4.5	1.9 1.2, 2.6	1.6 0.4, 2.7	35.7 33.2, 38.1	19.5 17.1, 22.0	0.0 0.0, 0.0	12.6 10.9, 14.3
15-19	7.8 6.4, 9.3	1.3 0.7, 1.9	6.5 5.0, 8.1	41.4 38.8, 44.1	17.1 15.0, 19.1	0.2 0.0, 0.4	16.4 14.4, 18.4
20-24	10.6 9.3, 11.9	2.4 1.7, 3.0	8.2 6.8, 9.6	56.5 54.4, 58.5	25.6 23.8, 27.4	0.2 0.0, 0.4	20.1 18.4, 21.7
25-29	6.1 5.1, 7.0	1.9 1.4, 2.4	4.2 3.1, 5.2	51.2 49.3, 53.1	25.4 23.7, 27.0	0.2 0.0, 0.4	19.6 18.0, 21.1
30-34	5.4 4.6, 6.3	1.3 0.9, 1.7	4.2 3.2, 5.1	41.6 39.8, 43.4	20.2 18.8, 21.7	0.3 0.1, 0.4	15.6 14.3, 17.0
35-39	6.2 5.3, 7.1	1.3 0.8, 1.7	5.0 4.0, 6.0	36.1 34.3, 37.9	15.2 13.8, 16.5	0.3 0.1, 0.5	14.4 13.1, 15.7
40-44	5.1 4.2, 6.0	1.1 0.7, 1.5	4.1 3.1, 5.0	32.3 30.4, 34.2	13.4 12.0, 14.8	0.4 0.1, 0.6	13.4 12.0, 14.8
45-49	3.1 2.3, 3.9	0.7 0.3, 1.1	2.4 1.5, 3.3	28.2 26.2, 30.3	9.9 8.5, 11.3	0.4 0.1, 0.7	14.8 13.2, 16.4
50-54	3.2 2.3, 4.1	0.5 0.2, 0.9	2.6 1.7, 3.6	27.0 24.8, 29.2	9.4 7.9, 10.8	0.8 0.3, 1.2	13.7 11.9, 15.4
55-59	2.9 1.9, 3.9	0.5 0.1, 0.9	2.5 1.4, 3.5	22.8 20.4, 25.3	7.9 6.3, 9.5	0.6 0.2, 1.1	11.4 9.5, 13.3
60-64	3.8 2.5, 5.0	0.8 0.2, 1.3	3.0 1.6, 4.4	23.4 20.6, 26.2	8.0 6.2, 9.8	1.9 1.0, 2.8	9.8 7.8, 11.8
65-69	1.9 0.9, 2.9	0.6 0.0, 1.1	1.3 0.2, 2.5	20.4 17.4, 23.4	4.1 2.7, 5.6	1.4 0.5, 2.3	13.0 10.5, 15.5
70-74	2.1 0.8, 3.3	0.9 0.1, 1.7	1.2 -0.3, 2.6	23.7 20.0, 27.3	4.7 2.9, 6.5	3.9 2.2, 5.5	13.0 10.2, 15.9
75+	2.5 1.5, 3.6	0.2 -0.1, 0.5	2.3 1.2, 3.4	31.9 28.8, 35.0	6.6 4.9, 8.2	9.8 7.8, 11.8	13.0 10.8, 15.3

Notes: See Table 5.

Table 9. Foreign-born Emigration Based on CPS Matching Method Among Females, by Age Group

Age Group	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
0-14	3.6 2.6, 4.6	1.3 0.7, 2.0	2.2 1.0, 3.4	35.6 33.0, 38.2	19.4 16.8, 22.0	0.0 0.0, 0.0	12.7 10.9, 14.5
15-19	1.3 0.6, 1.9	0.9 0.3, 1.4	0.4 -0.4, 1.3	39.9 37.1, 42.7	20.1 17.8, 22.4	0.1 -0.1, 0.3	18.4 16.2, 20.6
20-24	4.7 3.7, 5.7	1.3 0.8, 1.8	3.4 2.3, 4.5	49.3 46.9, 51.6	25.7 23.6, 27.7	0.1 0.0, 0.3	18.8 17.0, 20.6
25-29	5.8 4.8, 6.7	1.0 0.6, 1.4	4.8 3.7, 5.8	43.4 41.3, 45.4	22.5 20.8, 24.2	0.1 0.0, 0.3	15.0 13.5, 16.4
30-34	4.9 4.0, 5.7	1.0 0.6, 1.4	3.9 3.0, 4.8	35.3 33.5, 37.2	17.4 16.0, 18.9	0.1 0.0, 0.3	12.9 11.6, 14.2
35-39	2.0 1.5, 2.5	0.9 0.5, 1.2	1.1 0.5, 1.8	29.1 27.4, 30.9	13.4 12.0, 14.7	0.1 0.0, 0.3	13.6 12.3, 14.9
40-44	1.4 0.9, 1.9	0.3 0.1, 0.5	1.1 0.6, 1.6	27.8 26.0, 29.6	12.5 11.1, 13.8	0.2 0.0, 0.4	13.7 12.3, 15.1
45-49	2.7 2.0, 3.4	0.4 0.1, 0.7	2.3 1.5, 3.0	23.5 21.5, 25.4	7.8 6.6, 9.0	0.2 0.0, 0.5	12.7 11.2, 14.2
50-54	2.6 1.8, 3.4	0.5 0.1, 0.8	2.2 1.3, 3.0	23.5 21.4, 25.6	7.8 6.5, 9.2	0.4 0.1, 0.8	12.6 11.0, 14.2
55-59	2.5 1.6, 3.3	0.4 0.0, 0.7	2.1 1.2, 3.0	20.7 18.5, 22.9	6.8 5.4, 8.2	0.5 0.1, 0.8	10.9 9.2, 12.6
60-64	2.7 1.7, 3.6	0.4 0.0, 0.7	2.3 1.2, 3.3	21.7 19.2, 24.2	7.3 5.8, 8.9	1.3 0.6, 2.0	10.4 8.6, 12.3
65-69	3.0 1.9, 4.2	0.6 0.1, 1.1	2.5 1.2, 3.7	19.3 16.7, 21.9	3.6 2.3, 4.8	1.0 0.3, 1.7	11.7 9.5, 13.8
70-74	3.2 1.9, 4.5	0.5 0.0, 1.0	2.7 1.3, 4.1	20.9 18.0, 23.9	3.8 2.4, 5.2	2.5 1.3, 3.6	11.5 9.2, 13.8
75+	3.5 2.5, 4.5	0.1 -0.1, 0.3	3.3 2.3, 4.3	27.1 24.7, 29.6	5.1 3.9, 6.3	7.0 5.6, 8.4	11.6 9.8, 13.3

Notes: See Table 5.

Table 10. Foreign-born Emigration Based on CPS Matching Method Among Mexicans, by Sex and Age Group

Sex and Age Group	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
<b>Mexican Males</b>							
0-14	2.5 1.3, 3.7	2.4 1.2, 3.6	0.1 -1.6, 1.8	36.7 33.0, 40.5	20.8 17.1, 24.6	0.0 0.0, 0.0	13.4 10.8, 16.1
15-24	14.3 12.6, 16.1	2.4 1.6, 3.2	11.9 10.0, 13.9	59.0 56.5, 61.4	23.7 21.6, 25.8	0.2 0.0, 0.5	20.7 18.7, 22.7
25-34	6.5 5.5, 7.5	1.9 1.3, 2.4	4.6 3.5, 5.8	48.3 46.3, 50.3	22.1 20.4, 23.7	0.2 0.0, 0.4	19.5 17.9, 21.1
35-44	8.7 7.3, 10.1	1.5 0.9, 2.1	7.2 5.7, 8.7	37.1 34.8, 39.5	13.7 12.0, 15.4	0.4 0.1, 0.7	14.4 12.7, 16.1
45-54	0.3 -0.1, 0.7	0.5 0.0, 0.9	-0.1 -0.7, 0.5	31.4 28.2, 34.6	11.1 8.9, 13.2	0.7 0.1, 1.3	19.3 16.6, 22.0
55-64	0.0 -0.1, 0.1	1.4 0.2, 2.6	-1.4 -2.6, -0.2	22.5 18.3, 26.8	8.7 5.9, 11.6	1.1 0.1, 2.2	12.7 9.3, 16.1
65+	2.0 0.4, 3.6	0.5 -0.3, 1.4	1.5 -0.4, 3.3	31.3 25.9, 36.7	8.7 5.4, 12.0	6.4 3.6, 9.3	14.2 10.2, 18.3
<b>Mexican Females</b>							
0-14	2.6 1.3, 4.0	1.7 0.6, 2.8	0.9 -0.8, 2.6	37.9 33.8, 42.0	21.4 18.0, 24.9	0.0 0.0, 0.0	13.8 10.9, 16.8
15-24	4.9 3.6, 6.1	1.4 0.7, 2.1	3.4 2.0, 4.9	50.5 47.6, 53.5	25.9 23.3, 28.4	0.1 -0.1, 0.3	19.7 17.4, 22.0
25-34	4.0 3.1, 4.9	1.0 0.5, 1.4	3.0 2.0, 4.0	39.4 37.2, 41.7	18.7 16.9, 20.5	0.1 0.0, 0.3	16.6 14.9, 18.3
35-44	2.6 1.7, 3.5	0.6 0.2, 1.0	2.0 1.1, 3.0	28.5 26.0, 30.9	12.6 10.8, 14.5	0.2 0.0, 0.5	13.0 11.1, 14.8
45-54	0.0 0.0, 0.0	0.4 -0.1, 0.8	-0.4 -0.8, 0.1	26.0 22.7, 29.3	8.8 6.7, 10.9	0.4 -0.1, 0.8	16.8 14.1, 19.6
55-64	0.0 0.0, 0.0	0.3 -0.2, 0.9	-0.3 -0.9, 0.2	22.1 17.9, 26.2	7.5 4.9, 10.2	0.8 -0.1, 1.7	13.7 10.3, 17.2
65+	5.5 3.1, 8.0	0.3 -0.3, 1.0	5.2 2.6, 7.8	26.9 22.1, 31.7	5.1 2.7, 7.5	4.9 2.5, 7.2	11.4 7.9, 14.8

Notes: See Table 5.

Table 11. Foreign-born Emigration Based on CPS Matching Method Among Non-Mexicans, by Sex and Age Group

Sex and Age Group	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
<b>Non-Mexican Males</b>							
0-14	4.3 2.9, 5.7	1.6 0.8, 2.5	2.7 1.1, 4.3	34.9 31.7, 38.1	18.5 15.3, 21.7	0.0 0.0, 0.0	12.0 9.8, 14.2
15-24	5.9 4.8, 6.9	1.6 1.1, 2.2	4.2 3.1, 5.4	44.7 42.5, 46.9	21.5 19.7, 23.3	0.2 0.0, 0.4	17.2 15.5, 18.8
25-34	5.2 4.4, 5.9	1.3 0.9, 1.7	3.8 3.0, 4.7	44.7 43.0, 46.4	23.2 21.7, 24.6	0.2 0.1, 0.4	16.1 14.8, 17.4
35-44	4.3 3.6, 5.0	1.0 0.7, 1.3	3.3 2.5, 4.1	33.0 31.5, 34.6	14.7 13.5, 15.9	0.3 0.1, 0.5	13.7 12.6, 14.9
45-54	4.0 3.3, 4.8	0.7 0.4, 1.0	3.4 2.5, 4.2	26.5 24.7, 28.2	9.2 8.1, 10.3	0.5 0.3, 0.8	12.7 11.4, 14.0
55-64	4.0 3.1, 5.0	0.4 0.1, 0.7	3.6 2.6, 4.7	23.2 21.1, 25.3	7.8 6.5, 9.1	1.2 0.6, 1.7	10.2 8.7, 11.7
65+	2.2 1.6, 2.9	0.5 0.2, 0.8	1.7 1.0, 2.5	25.2 23.2, 27.2	4.8 3.8, 5.7	5.3 4.3, 6.4	12.8 11.3, 14.4
<b>Non-Mexican Females</b>							
0-14	4.2 2.8, 5.6	1.1 0.4, 1.8	3.1 1.5, 4.7	34.0 30.7, 37.4	18.0 15.3, 20.7	0.0 0.0, 0.0	11.8 9.6, 14.1
15-24	2.4 1.7, 3.1	0.9 0.5, 1.4	1.5 0.6, 2.3	42.4 40.1, 44.7	22.0 20.1, 23.9	0.1 0.0, 0.2	18.0 16.2, 19.7
25-34	6.0 5.2, 6.8	1.0 0.6, 1.3	5.0 4.1, 5.9	38.9 37.2, 40.6	20.5 19.1, 21.9	0.1 0.0, 0.2	12.3 11.2, 13.5
35-44	1.4 1.0, 1.8	0.6 0.3, 0.8	0.8 0.4, 1.3	28.5 27.0, 30.0	13.0 11.9, 14.1	0.2 0.0, 0.3	13.9 12.7, 15.0
45-54	3.3 2.7, 4.0	0.5 0.2, 0.7	2.8 2.1, 3.6	22.8 21.3, 24.4	7.6 6.6, 8.5	0.3 0.1, 0.5	11.6 10.5, 12.8
55-64	3.1 2.3, 3.8	0.4 0.1, 0.7	2.7 1.9, 3.5	21.0 19.2, 22.8	7.0 5.8, 8.1	0.8 0.4, 1.2	10.1 8.8, 11.4
65+	3.0 2.3, 3.6	0.4 0.1, 0.6	2.6 1.9, 3.3	22.7 21.1, 24.3	4.2 3.4, 5.0	3.9 3.2, 4.7	11.6 10.4, 12.9

Notes: See Table 5.

Table 12. Foreign-born Emigration Based on CPS Matching Method, By Years in the U.S. and Mexican Origin

Mexican-origin and Years in U.S.	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
<b>All Foreign-born</b>							
0-4 yrs	7.8 7.2, 8.3	2.1 1.8, 2.4	5.7 5.0, 6.3	51.3 50.3, 52.4	25.1 24.0, 26.1	0.4 0.2, 0.5	18.1 17.3, 19.0
5-9 yrs	4.9 4.4, 5.4	1.4 1.2, 1.7	3.5 3.0, 4.0	39.9 38.9, 41.0	18.9 18.1, 19.8	0.4 0.3, 0.5	15.7 15.0, 16.5
10+ yrs	3.0 2.8, 3.2	0.5 0.5, 0.6	2.5 2.3, 2.7	27.5 27.0, 28.0	10.6 10.2, 10.9	1.0 0.9, 1.1	12.9 12.5, 13.2
<b>Mexican</b>							
0-4 yrs	9.3 8.3, 10.4	2.7 2.1, 3.3	6.6 5.4, 7.9	59.4 57.5, 61.2	27.7 26.1, 29.4	0.3 0.1, 0.5	22.0 20.5, 23.5
5-9 yrs	6.2 5.3, 7.1	1.9 1.4, 2.4	4.3 3.3, 5.3	45.7 43.9, 47.5	21.2 19.8, 22.7	0.3 0.1, 0.5	18.0 16.6, 19.4
10+ yrs	3.5 3.1, 3.9	0.7 0.5, 0.8	2.9 2.4, 3.3	30.9 29.8, 31.9	12.2 11.5, 13.0	0.7 0.5, 0.9	14.4 13.6, 15.2
<b>Non-Mexican</b>							
0-4 yrs	7.0 6.3, 7.6	1.7 1.4, 2.1	5.2 4.5, 6.0	47.3 45.9, 48.6	23.7 22.6, 24.8	0.4 0.2, 0.6	16.2 15.2, 17.2
5-9 yrs	4.2 3.7, 4.7	1.2 0.9, 1.4	3.0 2.5, 3.6	36.8 35.6, 38.1	17.7 16.7, 18.7	0.4 0.3, 0.6	14.5 13.6, 15.4
10+ yrs	2.9 2.6, 3.1	0.5 0.4, 0.6	2.4 2.1, 2.6	26.2 25.6, 26.8	10.0 9.5, 10.4	1.2 1.0, 1.3	12.3 11.8, 12.7

Notes: See Table 5.

Table 13. Foreign-born Emigration Based on CPS Matching Method, By Country or Region of Birth

Country or Region	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
Mexico	5.3 4.9, 5.7	1.4 1.2, 1.5	3.9 3.5, 4.4	40.0 39.1, 40.8	17.4 16.6, 18.2	0.5 0.4, 0.6	16.8 16.1, 17.4
Central America	3.8 3.1, 4.4	0.9 0.6, 1.2	2.9 2.1, 3.6	41.9 40.2, 43.5	16.5 15.2, 17.7	0.4 0.2, 0.7	21.2 19.8, 22.6
North America	5.8 4.4, 7.2	0.5 0.1, 0.8	5.4 3.9, 6.8	29.5 26.8, 32.2	12.5 10.6, 14.5	1.2 0.6, 1.9	10.0 8.2, 11.7
Caribbean	3.5 3.1, 3.9	0.9 0.6, 1.1	2.7 2.2, 3.1	33.1 32.0, 34.1	13.1 12.4, 13.9	0.7 0.5, 0.9	15.7 14.8, 16.5
Cuba	2.8 1.9, 3.6	0.6 0.2, 1.0	2.1 1.2, 3.0	27.6 25.4, 29.9	10.5 9.0, 12.1	1.4 0.8, 1.9	13.0 11.3, 14.7
Dominican Republic	4.0 2.8, 5.2	1.5 0.8, 2.3	2.5 1.0, 3.9	33.6 30.7, 36.6	12.7 10.6, 14.7	0.5 0.1, 0.9	16.5 14.2, 18.8
Other Caribbean	3.7 3.1, 4.2	0.8 0.5, 1.0	2.9 2.3, 3.5	34.6 33.3, 36.0	14.1 13.1, 15.0	0.6 0.4, 0.8	16.3 15.3, 17.4
Europe	4.8 4.2, 5.4	0.6 0.4, 0.7	4.3 3.7, 4.9	25.6 24.4, 26.7	8.8 8.1, 9.6	1.7 1.4, 2.1	10.2 9.4, 11.0
Asia	2.3 2.0, 2.5	1.0 0.8, 1.2	1.3 1.0, 1.6	30.7 29.8, 31.5	15.2 14.6, 15.9	0.7 0.5, 0.8	12.5 11.9, 13.1
China	4.1 3.2, 4.9	0.7 0.3, 1.0	3.4 2.5, 4.3	31.5 29.5, 33.4	12.4 11.0, 13.7	0.9 0.5, 1.3	14.1 12.7, 15.6
India	9.5 8.0, 10.9	0.8 0.3, 1.2	8.7 7.1, 10.2	33.4 31.0, 35.7	16.7 14.8, 18.6	0.6 0.2, 1.0	6.6 5.4, 7.9
Philippines	0.6 0.2, 0.9	2.2 1.6, 2.9	-1.7 -2.4, -0.9	26.2 24.3, 28.2	12.9 11.4, 14.4	0.7 0.3, 1.0	12.1 10.6, 13.5
Korea	1.0 0.4, 1.7	1.0 0.4, 1.6	0.0 -0.9, 0.9	32.5 29.6, 35.5	16.9 14.6, 19.3	0.7 0.2, 1.2	13.9 11.7, 16.1
Other Asia	0.1 0.0, 0.2	0.9 0.7, 1.2	-0.9 -1.1, -0.6	30.9 29.6, 32.3	16.7 15.6, 17.8	0.6 0.4, 0.8	13.6 12.6, 14.6
Africa	18.5 16.1, 20.9	0.9 0.3, 1.5	17.6 15.1, 20.0	40.9 37.9, 43.9	15.4 13.2, 17.6	0.7 0.2, 1.2	6.3 4.8, 7.8
Other	2.6 2.0, 3.3	0.6 0.3, 0.9	2.0 1.3, 2.7	29.8 28.0, 31.6	15.1 13.7, 16.6	1.2 0.8, 1.6	10.8 9.6, 12.1

Notes: See Table 5.

Table 14. Foreign-born Emigration Based on CPS Matching Method, By Mexican Origin, Citizenship, and Migration Status

Country or Region	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
<b>All Foreign-born</b>							
Naturalized	2.5 2.2, 2.8	0.3 0.2, 0.5	2.1 1.8, 2.4	23.2 22.4, 24.0	8.6 8.1, 9.1	1.4 1.2, 1.6	10.7 10.1, 11.3
LPR	4.5 4.2, 4.9	1.2 1.0, 1.4	3.4 3.0, 3.8	34.9 34.1, 35.8	16.0 15.3, 16.6	0.7 0.6, 0.9	13.7 13.1, 14.3
Unauthorized	5.7 5.2, 6.2	1.5 1.3, 1.8	4.2 3.6, 4.7	42.8 41.7, 43.8	20.3 19.4, 21.2	0.2 0.1, 0.3	16.5 15.8, 17.3
Refugee	2.8 2.1, 3.4	0.7 0.4, 1.1	2.0 1.3, 2.7	30.3 28.5, 32.0	15.6 14.3, 17.0	0.7 0.4, 1.0	11.2 10.0, 12.4
Non-Immigrant	10.2 8.4, 12.1	2.6 1.6, 3.5	7.7 5.6, 9.8	54.1 51.0, 57.1	28.6 25.9, 31.4	0.2 -0.1, 0.4	15.0 12.8, 17.2
<b>Mexican</b>							
Naturalized	2.0 1.3, 2.7	0.2 0.0, 0.4	1.8 1.1, 2.5	22.4 20.4, 24.4	8.7 7.3, 10.0	1.0 0.5, 1.5	10.8 9.3, 12.3
LPR	4.7 4.1, 5.3	1.4 1.1, 1.8	3.3 2.6, 4.0	37.4 35.9, 38.8	17.4 16.3, 18.5	0.6 0.4, 0.9	14.6 13.6, 15.7
Unauthorized	6.3 5.6, 7.0	1.7 1.3, 2.1	4.6 3.8, 5.4	44.3 42.9, 45.8	21.2 20.0, 22.4	0.2 0.1, 0.4	16.6 15.5, 17.7
<b>Non-Mexican</b>							
Naturalized	2.6 2.3, 2.9	0.4 0.2, 0.5	2.2 1.9, 2.5	23.3 22.5, 24.1	8.6 8.0, 9.2	1.4 1.2, 1.7	10.7 10.1, 11.3
LPR	4.4 4.0, 4.9	1.0 0.8, 1.2	3.4 2.9, 4.0	33.5 32.4, 34.6	15.1 14.3, 15.9	0.8 0.6, 1.0	13.1 12.4, 13.9
Unauthorized	4.9 4.2, 5.6	1.3 0.9, 1.6	3.7 2.9, 4.4	40.9 39.3, 42.4	19.3 17.7, 20.8	0.2 0.1, 0.4	16.5 15.3, 17.6
Refugee	2.8 2.1, 3.4	0.7 0.4, 1.1	2.0 1.3, 2.7	30.3 28.5, 32.0	15.6 14.3, 17.0	0.7 0.4, 1.0	11.2 10.0, 12.4
Non-Immigrant	10.4 8.5, 12.4	2.4 1.5, 3.4	8.0 5.9, 10.2	54.2 51.1, 57.4	28.8 26.0, 31.7	0.2 -0.1, 0.4	14.8 12.6, 17.0

Notes: See Table 5.

\*Virtually no Mexicans were classified as refugees or non-immigrants.

Table 15. Foreign-born Emigration Based on CPS Matching Method, By Period, Mexican Origin, Citizenship, and Migration Status

Country or Region	Emigration	Return Immigration	Net Emigration	Non-match	Internal migration	Mortality	Non-follow-up
<b>Mexicans, 1996-1999</b>							
Naturalized	1.8 0.6, 2.9	0.0 0.0, 0.0	1.8 0.6, 2.9	20.5 17.0, 24.0	9.4 6.9, 11.9	0.8 0.0, 1.5	8.6 6.2, 11.0
LPR	4.6 3.6, 5.5	1.5 0.9, 2.0	3.1 2.0, 4.2	36.8 34.6, 39.1	18.8 17.0, 20.6	0.6 0.2, 1.0	12.9 11.4, 14.5
Unauthorized	5.8 4.5, 7.0	2.1 1.3, 2.8	3.7 2.3, 5.1	42.5 39.9, 45.1	22.7 20.5, 25.0	0.2 0.0, 0.5	13.8 11.9, 15.6
<b>Mexicans, 2000-2004</b>							
Naturalized	2.1 1.2, 2.9	0.3 0.0, 0.6	1.8 0.8, 2.7	23.3 20.8, 25.8	8.3 6.7, 10.0	1.1 0.5, 1.7	11.8 9.9, 13.7
LPR	4.8 4.0, 5.6	1.4 0.9, 1.8	3.4 2.5, 4.4	37.7 35.8, 39.6	16.4 15.0, 17.9	0.6 0.3, 0.9	15.9 14.4, 17.3
Unauthorized	6.5 5.7, 7.4	1.6 1.2, 2.0	5.0 4.0, 5.9	45.1 43.4, 46.9	20.5 18.8, 22.3	0.2 0.1, 0.4	17.8 16.5, 19.2
<b>Non-Mexicans, 1996-1999</b>							
Naturalized	2.6 2.1, 3.1	0.4 0.2, 0.6	2.2 1.6, 2.7	21.4 20.1, 22.7	9.5 8.6, 10.4	1.1 0.8, 1.5	8.2 7.3, 9.1
LPR	5.0 4.2, 5.7	1.2 0.8, 1.6	3.7 2.9, 4.6	33.6 32.0, 35.2	17.2 15.9, 18.5	0.7 0.4, 0.9	10.8 9.7, 11.8
Unauthorized	5.4 4.2, 6.7	1.4 0.8, 2.0	4.0 2.6, 5.4	39.3 36.7, 42.0	20.4 18.2, 22.6	0.2 0.0, 0.5	13.3 11.5, 15.2
Refugee	2.2 1.3, 3.0	0.7 0.2, 1.2	1.5 0.5, 2.5	31.4 28.6, 34.2	18.5 16.2, 20.9	0.8 0.2, 1.3	10.0 8.2, 11.8
Non-Immigrant	9.0 6.0, 12.0	3.7 1.7, 5.6	5.3 1.8, 8.9	52.1 46.9, 57.2	31.5 26.7, 36.3	0.2 -0.3, 0.6	11.5 8.2, 14.8
<b>Non-Mexicans, 2000-2004</b>							
Naturalized	2.5 2.1, 2.9	0.3 0.2, 0.5	2.2 1.8, 2.6	24.5 23.4, 25.6	8.0 7.4, 8.7	1.6 1.3, 1.9	12.3 11.5, 13.1
LPR	4.0 3.4, 4.6	0.8 0.6, 1.1	3.2 2.5, 3.9	33.4 32.0, 34.9	13.5 12.4, 14.5	0.9 0.6, 1.2	15.0 13.9, 16.1
Unauthorized	4.7 3.9, 5.5	1.2 0.8, 1.6	3.5 2.5, 4.4	41.7 39.8, 43.6	18.7 16.8, 20.6	0.3 0.1, 0.5	18.0 16.5, 19.5
Refugee	3.2 2.3, 4.0	0.8 0.3, 1.2	2.4 1.4, 3.3	29.5 27.3, 31.8	13.7 12.0, 15.4	0.7 0.3, 1.1	12.0 10.4, 13.6
Non-Immigrant	11.3 8.8, 13.8	1.7 0.7, 2.7	9.6 6.9, 12.3	55.5 51.6, 59.5	27.2 23.7, 30.8	0.2 -0.2, 0.5	16.8 13.8, 19.7

Notes: See Table 5.

**Appendix A. SAS code for constructing data files and matching cases across years**

```

options error mprint;

** location on /ISS/ of CPS files **;
** location of core files **;
libname cps '/iss2/JVH/EMS2dat/';
** location of HHES March CPS files **;
libname cps96 '/iss/CPS4EMS2/march1996';
libname cps97 '/iss/CPS4EMS2/march1997';
libname cps98 '/iss/CPS4EMS2/march1998';
libname cps99 '/iss/CPS4EMS2/march1999';
libname cps00 '/iss/CPS4EMS2/march2000_2k';
libname cps00g '/iss/CPS4EMS2/march2000good';
libname cps01 '/iss/CPS4EMS2/march2001_schip_2k';
libname cps01mg '/iss/CPS4EMS2/march2001_mig';
libname cps02 '/iss/CPS4EMS2/march2002';
libname cps02mg '/iss/CPS4EMS2/march2002_mig';
libname cps03 '/iss/CPS4EMS2/march2003';
libname cps03mg '/iss/CPS4EMS2/march2003_mig';
libname cps03nw '/iss/CPS4EMS2/march2003_nwgt';
libname cps04 '/iss/CPS4EMS2/march2004';
libname cps04pr '/iss/CPS4EMS2/march2004_prelim';
libname cps05 '/iss/CPS4EMS2/march2005';

/*-----*/
/*-----*/
/* MACRO READING IN THE RAW MARCH CPS DATA AND RUNNING RECODES. THE 'RECODES' AND
'KEEPVARS' MACROS
ARE INVOKED WITHIN THIS MACRO. */
/*-----*/

/* notes: a_race is $1.; prdtrace is $2.;
a_hscol changes from $1. to $2.

*/

/*-----*/

%macro recodes ;
    yr=&yr;

if livqrt =1 then livqrt=1;/*living in a house*/
else if livqrt =11 then livqrt=2;/*living in a dormitory*/

```

```

else if livqrt in(5,6) then livqrt=3;/*living in a mobile home*/
else livqrt=4;/*other living quarters*/

if hscol =0 then enroll=3;/*not enrolled in school*/
else if hscol =1 then enroll=1;/*enrolled in high school*/
else if hscol =2 then enroll=2;/*enrolled in universities*/

if age ge 0 and age le 4 then agegrp=0;
else if age ge 5 and age le 9 then agegrp=1;
else if age ge 10 and age le 14 then agegrp=2;
else if age ge 15 and age le 19 then agegrp=3;
else if age ge 20 and age le 24 then agegrp=4;
else if age ge 25 and age le 29 then agegrp=5;
else if age ge 30 and age le 34 then agegrp=6;
else if age ge 35 and age le 39 then agegrp=7;
else if age ge 40 and age le 44 then agegrp=8;
else if age ge 45 and age le 49 then agegrp=9;
else if age ge 50 and age le 54 then agegrp=10;
else if age ge 55 and age le 59 then agegrp=11;
else if age ge 60 and age le 64 then agegrp=12;
else if age ge 65 and age le 69 then agegrp=13;
else if age ge 70 and age le 74 then agegrp=14;
else if age ge 75 then agegrp=15;

/*raceth: 1=white 2=black 3=asian 4=mexican 5=hispanic 6=other*/
if yr=2003 or yr=2004 or yr=2005 then do;
  if reorgn =1 then raceth=4;
  else if reorgn in(2,3,4,5) then raceth=5;
  else do;
    if race =4 then raceth=3;
    else if race =1 then raceth=1;
    else if race =2 then raceth=2;
    else raceth=6;
  end;
end;

else do;
  if reorgn in(1,2,3) then raceth=4;
  else if reorgn in(4,5,6,7) then raceth=5;
  else do;
    if race =4 then raceth=3;
    else if race =1 then raceth=1;
    else if race =2 then raceth=2;
    else raceth=6;
  end;
end;

if rprcitshp =4 or rprcitshp=5 then fb=1;
else fb=0;

if fb=0 then rpob=0;
else if rpenatvty = 315 then rpob=1; /* mexican */
else if rpenatvty = 337 then rpob=3; /* cuba */
else if rpenatvty = 339 then rpob=4; /* Dom Rep */
else if rpenatvty in(207,238,209) then rpob=5; /*China */

```

```

else if rpenatvty =231 then rpob=6; /* Philippines */
else if rpenatvty =217 or rpenatvty =218 then rpob=7; /* ??? */
else if rpenatvty =210 then rpob=8; /*India */
else if rpenatvty in(310,311,312,313,314,315,316,317,318) then rpob=2; /*
Central Am */
else if rpenatvty in (415,436,468,252) then rpob=10; /* North Africa */
else if rpenatvty ge 200 and rpenatvty le 253 then rpob=9; /* Asia */
else if rpenatvty ge 417 and rpenatvty le 462 then rpob=11; /* Africa */
else if rpenatvty ge 103 and rpenatvty le 148 then rpob=12; /* Europe */
else if rpenatvty ge 300 and rpenatvty le 304 then rpob=13; /* North Am */
else if rpenatvty ge 333 and rpenatvty le 389 then rpob=14; /* Caribbean */
else if rpenatvty=555 and reorgn in (1,2,3) then rpob=1; /* mexican */
else rpob=15; /* other */

/* THIS IS THE DEFAULT GENERATIONAL STATUS OF THE PERSON.
IF THE UI-LEGAL STATUS IMPUTATIONS ARE USED, THIS VARIABLE
WILL CHANGE SOMEWHAT AS SOME IDENTIFIED AS US-BORN ARE RECODED TO FB */

if FB=1 then generation=1;
else if rpenmtvty>78 or rpefntvty>78 then generation=2;
else generation=3;

if rpob=1 then mexi=1;
else if rpob ne 0 then mexi=0;

/* peinusyr has 2 or 4 digits, 2 for years less than 2000, and 4 for years 2000 and
later. */

if &yr <1999 and rpeinusyr >0 then avgpoe=&yr-(1900+rpeinusyr);
else if &yr >= 1999 and rpeinusyr >0 then avgpoe=&yr-(rpeinusyr);
if avgpoe >age then avgpoe=age;
if avgpoe ge 0 and avgpoe lt 5 then yrarriv=1;
else if avgpoe ge 5 and avgpoe lt 10 then yrarriv=2;
else if avgpoe ge 10 then yrarriv=3;

/*coding dummy variables*/
male=0;
if sex =1 then male=1;

white=0; black=0; mexican=0; hisp=0; other=0;
if raceth=1 then white=1;
else if raceth=2 then black=1;
else if raceth=4 then mexican=1;
else if raceth=5 then hisp=1;
else other=1;

if age ge 18 and age le 29 then ager=1;
else if age ge 18 and age le 39 then ager=2;
else if age ge 18 and age le 49 then ager=3;
else if age ge 18 and age le 59 then ager=4;
else if age ge 18 and age le 69 then ager=5;
else if age ge 18 and age le 79 then ager=6;
else if age ge 18 then ager = 7;

array ag[7] a1-a7;
do i = 1 to 7; ag[i]=0;

```

```

if ager=i then do;
  ag[i]=1;
end;end;

array hel[5] h1-h5;
do i = 1 to 5;
  hel[i]=0;
  if rhea=i then hel[i]=1;
end;

ownhome=0;
if tenure = 1 then ownhome=1;

if rhea=0 then delete;

/* check this how to identify Hispanic oversample */
  yr=&yr;
oversamp=0;
os = .; os = h_ser;
if os ge 50 then oversamp=1;

if rmigsame =2 then intmig =1;
else if rmigsame =1 then intmig=0;
else intmig=.;

prcitshp = rprcitshp;
pob = rpenatvty ;
mpob = rpemntvty ;
fpob = rpefntvty ;
hea = rhea;
migsame = rmigsame;
peinusyr = rpeinusyr;

%mend recodes;
/*-----*/
%macro perhh(yr,file,hh,p);

/*-----*/
/* read in person file */
/*-----*/
%if &yr <2000 %then %do;
  %let pid = a_lineno;
  %let race = a_race;
  %let hisp = a_reorgn;
%end;
%else %if &yr >=2000 and &yr <2003 %then %do;
  %let pid = a_lineno;
  %let race = a_race;
  %let hisp = a_reorgn;
%end;
%else %if &yr >=2003 %then %do;
  %let pid = a_lineno;
  %let race = prdtrace;
  %let hisp = prdthsp;
%end;
%if &yr = 2005 %then %do; %let st = gestcen ;%end;

```

```

%else %if &yr <2005 %then %do; %let st = hg_st60 ; %end;

DATA cps.p_file (keep = /* newpvars */ seq lineno parent rrp age maritl sex grdatn
race reorgn fnlwgt marsupwt hscol rmigsame dis pearval
rpenatvty rpeintvty rpefntvty rpeinusyr rprcitshp rhea );

set &file..&p (keep= /* oldpvars */ PH_SEQ &pid a_parent a_exrrp a_age a_maritl a_sex
a_hga
&race &hisp a_fnlwgt marsupwt a_hscol migsame dis_hp pearval
penatvty pemntvty pefntvty peinusyr prcitshp hea );

array oldvars [*] &pid a_parent a_exrrp a_age a_maritl a_sex a_hga
&race &hisp a_hscol migsame dis_hp
penatvty pemntvty pefntvty prcitshp hea;

array newvars [*] lineno parent rrp age maritl sex grdatn
race reorgn hscol rmigsame dis
rpenatvty rpeintvty rpefntvty rprcitshp rhea ;

do i = 1 to 17;
newvars[i] = .;
newvars[i] = oldvars[i];
end;
if &yr ne 1999 then do;
rpeinusyr = . ; rpeinusyr = peinusyr;
end;
else do;
rpeinusyr = peinusyr;
end;

seq = .;
seq = ph_seq;
fnlwgt = .;
fnlwgt = a_fnlwgt;

proc sort data = cps.p_file; by seq;
run;
/*-----*/

/*-----*/
/* read in HH file */
/*-----*/
%if &yr <=2000 %then %do;

DATA cps.h_file (drop =h_seq h_livqrt hg_st60 h_tenure h_mis ) ;
length h_hhnum $2. h_psu h_seg 4 ;
set &file..&hh (keep= H_SEG H_TENURE H_TYPE /*HRLONGID*/ h_seq h_numper h_month h_mis
h_hhnum h_livqrt h_tenure &st h_psu h_ser h_hhnum h_sersuf );
%end;
%else %if &yr > 2000 %then %do;

DATA cps.h_file (drop =h_seq h_livqrt hg_st60 h_tenure h_mis hrschips) ;
length h_hhnum $2. h_psu h_seg 4 ;
set &file..&hh (keep= H_SEG H_TENURE H_TYPE /*HRLONGID*/ h_seq h_numper h_month h_mis
h_hhnum h_livqrt h_tenure &st h_psu h_ser h_hhnum h_sersuf hrschips);
if hrschips = 3 or hrschips = 4 then delete; /* deleting schip oversample in 9th MIS */

```

```

%end;

array oldvars [*] h_livqrt &st h_tenure h_mis ;
array newvars [*] livqrt state tenure mis ;
do i=1 to 4;
    newvars[i] = .;
    newvars[i] = oldvars[i];
end;
seq=h_seq ;
proc sort data = cps.h_file; by seq;
run;
/*-----*/

/*-----*/
/* merge person and HH files and do recodes */
/*-----*/
data cps.perhh&yr (keep= ownhome tenure h1-h5 a1-a7 white black mexican hisp other male
sex
                                generation newhrlongid newhrlongidb yr agegrp raceth prcitshp
migsame enroll
                                hea intmig fb pob rpob mpob fpob peinusyr livqrt oversamp avgpoe
mexi yrarriv
                                state lineno age reorgn race parent fnlwgt grdatn seq hhnum mis rrp
                                maritl dis pearval wgt);
merge cps.p_file (in=p) cps.h_file (in=h);
    by seq;
    if p and h;
wgt = marsupwt;
/* CREATE NEWHRLONGID FOR FILES OLDER THAN 2000, USING THE INSTRUCTIONS PROVIDED
BY CYNTHIA DAVIS. */

/* convert h_psu, h_seg, h_ser, and h_hhnum to character */
psu = put(h_psu, 5.);
seg = put(h_seg, 4.);
hhnum = put(h_hhnum, 2.);
psu = translate(psu,'0',' ');

if h_sersuf eq 'A' or h_sersuf eq 'B' or h_sersuf eq 'C' or h_sersuf eq 'D'
or h_sersuf eq 'E' or h_sersuf eq 'F' or h_sersuf eq 'X' or h_sersuf eq 'Y'
or h_sersuf eq 'Z' or h_sersuf eq 'G' or h_sersuf eq 'V' or h_sersuf = 'W'
then ss = h_sersuf ; else ss = '2';

seg = right(seg);
hhnum = right(hhnum);
h_ser = right(h_ser);
h_sersuf = right(h_sersuf);

seg = translate(seg,'0',' ');
hhnum = translate(hhnum,'0',' ');
h_ser = translate(h_ser,'0',' ');

newhrlongid = psu||seg||h_ser||ss||hhnum;
newhrlongidb = psu||seg||h_ser||ss ;

```

```

%recodes;
/*-----*/
/* check for duplicate cases in person file */
/*-----*/

proc sort data = cps.perhh&yr;
by mis newhrlongid lineno;
data cps.doublep (keep = mis newhrlongid lineno doublep)
  cps.doubleh (keep = mis newhrlongid doubleh ) ;
set cps.perhh&yr;
by mis newhrlongid lineno;
retain doublep doubleh ;
if first.lineno then doublep=0;
doublep+1;
if last.lineno and doublep >1 then output cps.doublep;

if first.newhrlongid then doubleh=0;
if first.newhrlongid and lineno = 1 then doubleh+1;
if last.newhrlongid and doubleh >1 then output cps.doubleh;

data cps.perhh&yr ;
merge cps.perhh&yr cps.doubleh;
by mis newhrlongid ;

data cps.perhh&yr;
merge cps.perhh&yr cps.doublep;
by mis newhrlongid lineno ;
if doublep ge 2 and doublep ne . then delete;
if doubleh ge 2 and doubleh ne . then delete;
run;
proc means n mean min max;run;

%mend perhh;

/*-----*/
/* END OF MACRO READING IN THE RAW MARCH CPS DATA AND RUNNING RECODES.*/
/*-----*/
/*-----*/
/*-----*/
/* macro for matching files for year t = 1996-1999 */
/*-----*/

%macro matchcpsa(yra,yrb) ;

/*-----*/
/* select which match code to use */
/*-----*/
%if &yra <2000 %then %do;
  %let hid = newhrlongid ;
%end;
%else %if &yra >=2000 %then %do;
  %let hid = newhrlongid ;
%end;
/*-----*/
/* create person file for t+1 with adjusted MIS codes */
/*-----*/
data cps.yrb (keep=&hid lineno sex2 age2 mis ); /* based on person t+1 */

```

```

set cps.perhh&yrb;
sex2 = sex;
age2=age;
  if oversamp = 1 then mis = 9;
  if mis in(1,2,3,4) then delete;
  if oversamp = 0 then mis = mis-4;

data cps.yra;
set cps.perhh&yra;
if oversamp= 1 then mis=9;

/*-----*/
/* merge person files from t to t+1 */
/*-----*/
PROC SORT DATA=cps.yrb ; /* person t+1 */
by mis &hid lineno ;

PROC SORT DATA=cps.yra ; /* person t */
by mis &hid lineno ;

data cps.mrged&yra ; /* merged person t to t+1 */
merge cps.yra (in=a) cps.yrb (in=b) ;
by mis &hid lineno ;

if a and b then matched=1;
else if a and not b then matched=0;
else if b and not a then matched=.;
if matched=. then delete;

if sex = sex2 then sexdiff=0; else sexdiff=1;
agedif = age2-age;
if agedif in(0,1,2) then ragedif = 0; else ragedif=1;

s_a=0;
if (sexdiff=1 | ragedif=1) then s_a=1;

/* redo this code */
  elig=0;
  if ( oversamp=0 and mis in(1,2,3,4) ) or ((oversamp= 1 and mis=9) ) then elig=1;

if matched=1 and s_a=0 and elig=1 then match=1; /* good match */
if matched=1 and s_a=1 and elig=1 then match=2; /* ID matched, but not characteristics */
if matched=0 and elig=1 then match=3; /* no id matched but eligible to be matched */
if matched=. or elig = 0 then match=4; /* not eligible to be matched */

goodmatch=0;
if match=1 then goodmatch=1;

if matched=1 and s_a=0 then anymatch=1;
else anymatch=0;
run;

proc means data = cps.mrged&yra mean n;
var goodmatch anymatch matched;
class elig mis oversamp;

```

```

run;
%mend;
/* end of macro for matching files */

/*-----*/
/* END OF MACRO READING IN THE RAW MARCH CPS DATA AND RUNNING RECODES.*/
/*-----*/
/*-----*/
/*-----*/
/* macro for matching files for year t = 1996-1999 */
/*-----*/

/*-----*/

/* MERGING LEGAL STATUS (PASSEL ET AL). DO NOT RUN THIS SECTION IF YOU DO NOT INTEND
TO USE THE LEGAL STATUS IMPUTATIONS. Instead, run the alternative macro legalstatus
below (this
one is intentionally blank). */

proc format;
value citstat
1 - 3 = 'Native Born'
4 = 'Naturalized'
6 = 'Legal Immigrant'
7 = 'Illegal Immigrant'
8 - 9 = 'Refugee (Nat + Alien)'
11 - 33 = 'Non-Immigrants';
value uileg
1 - 3 = 'Native Born'
4 = 'Naturalized'
6 = 'Legal Immigrant'
7 = 'Illegal Immigrant'
8 - 9 = 'Refugee (Nat + Alien)'
11 - 33 = 'Non-Immigrants';
value immstat
1 - 3 = 'Native Born'
4 = 'Naturalized'
6 = 'Legal Immigrant'
7 = 'Illegal Immigrant'
8 - 9 = 'Refugee (Nat + Alien)'
11 - 33 = 'Non-Immigrants';

/*-----*/
%macro legalstatus(yr,yrb) ;

data lgstat (rename= (h_seq=seq a_lineno=lineno));
set cps.ext&yr ;

proc sort data=lgstat; by lineno seq;
proc sort data=cps.mrged&yrb; by lineno seq;

data cps.mrged&yrb; merge cps.mrged&yrb lgstat ; by lineno seq;
migstat=0;
if citstat in(1,2,3) then migstat=1;
else if citstat = 4 then migstat=2;

```

```

else if citstat = 6 then migstat=3;
else if citstat=7 then migstat=4;
else if citstat in(8,9) then migstat=5;
else if citstat ge 11 and citstat le 33 then migstat=6;
/*migstat: 1 = 'Native Born'
           2 = 'Naturalized'
           3 = 'Legal Immigrant'
           4 = 'Illegal Immigrant'
           5 = 'Refugee (Nat + Alien)'
           6 = 'Non-Immigrants';*/
generation=0;
if migstat ge 2 then generation=1;
else if mpob>78 or fpob>78 then generation=2;
else generation=3;
if migstat =4 then lgstat=0;/*illegal*/
else if migstat in (2,3,5,6) then lgstat=1;
if ui_wgt ne . then wgt=ui_wgt;

%mend legalstatus;
/*-----*/

%perhh(1996,cps96,hhld,person); run;
%perhh(1997,cps97,hhld,person); run;
%perhh(1998,cps98,hhld,person); run;
%perhh(1999,cps99,hhld,person); run;

/* %perhh(2000,cps00g,march00h,march00p); run; */
%perhh(2000,cps00,hhld,person); run;

%perhh(2001,cps01,hhld,person); run;
%perhh(2002,cps02,hhld,person); run;
%perhh(2003,cps03,hhld,person); run;

%perhh(2004,cps04,hhld,person); run;
%perhh(2005,cps05,hhld,person); run;

/*-----*/
/* matching */
/*-----*/
%matchcpsa(1996,1997); run;
%matchcpsa(1997,1998); run;
%matchcpsa(1998,1999); run;
%matchcpsa(1999,2000); run;
%matchcpsa(2000,2001); run;
%matchcpsa(2001,2002); run;
%matchcpsa(2002,2003); run;
%matchcpsa(2003,2004); run;
%matchcpsa(2004,2005); run;

/*-----*/
/* Fix 2000 migration variable--take from 2000good file */
/*-----*/
data m00 (keep = seq lineno intmig rmigsame ) ;
set cps00g.march00p (keep = ph_seq a_lineno migsame);
seq =.;seq = ph_seq;
lineno=.; lineno = a_lineno;

```

```

rmigsame = . ; rmigsame = migsame;
if rmigsame =2 then intmig =1;
else if rmigsame =1 then intmig=0;
else intmig=.;

proc sort data = m00; by seq lineno;
proc sort data = cps.mrged2000; by seq lineno;
data cps.mrged2000 (drop = rmigsame);
merge cps.mrged2000 (drop = intmig migsame) m00;
by seq lineno;
migsame = rmigsame;
proc means; run;

/*-----*/
/* now fix the weights for 2003 and 2004 */
/*-----*/

data w03 (keep = seq lineno wgt ) ;
set cps03nw.ASEC03PP;
seq = .;
seq = ph_seq;
lineno=.; lineno = a_lineno;
wgt = marsupwt;
proc sort data = w03; by seq lineno;
proc sort data = cps.mrged2003; by seq lineno;
data cps.mrged2003;
merge cps.mrged2003 (drop = wgt) w03;
by seq lineno;
proc means; run;
data w04 (keep = seq lineno wgt ) ;
set cps04pr.nwgt04pp;
seq = .;
seq = ph_seq;
lineno=.; lineno = a_lineno;
wgt = marsupwt;
proc sort data = w04; by seq lineno;
proc sort data = cps.mrged2004; by seq lineno;
data cps.mrged2004;
merge cps.mrged2004 (drop = wgt) w04;
by seq lineno;
proc means; run;
/*-----*/
/* Combine data files */
/*-----*/

data cps.combined (drop = doubleh doublep i marsupwt);
set cps.mrged1996 cps.mrged1997 cps.mrged1998 cps.mrged1999 cps.mrged2000 cps.mrged2001
cps.mrged2002 cps.mrged2003 cps.mrged2004 ;
s = substr(newhrlongid,6,1);
if oversamp = 1 then do;
  if yr in(1996, 1998, 2000,2002, 2004) and s in(5,6,7,8) then elig = 1;
  else if yr in (1997, 1999, 2001, 2003, 2005) and s in(1,2,3,4) then elig = 1;
  else elig = 0 ;
end;

if elig=1 and generation in(1,2);

```

```

proc means data=cps.combined; run;
proc means data = cps.combined mean n ;
var goodmatch;
class mis oversamp; run;
proc contents data = cps.combined ; run;
/* UI legal status and weights */

%legalstatus(96,1996); run;
%legalstatus(98,1998); run;
%legalstatus(99,1999); run;
%legalstatus(00,2000); run;
%legalstatus(01,2001); run;
%legalstatus(02,2002); run;
%legalstatus(03,2003); run;

/* combining files (with ui legal status) */
data cps.combinedui;
set cps.mrged1996 cps.mrged1998 cps.mrged1999 cps.mrged2000 cps.mrged2001 cps.mrged2002
cps.mrged2003 ;
s = substr(newhrlongid,6,1);
if oversamp = 1 then do;
  if yr in(1996, 1998, 2000,2002, 2004) and s in(5,6,7,8) then elig = 1;
  else if yr in (1997, 1999, 2001, 2003, 2005) and s in(1,2,3,4) then elig = 1;
  else elig = 0 ;
end;
if elig=1 and generation in(1,2);
proc means; run;
*/
/*-----*/
/*-----*/

run;

```

**Appendix B. STATA code, main program (Remigration706.5.do)**

```

capture log close
** Revised version of J. Van Hook s prog Emigration3d.do
** To match Census environment. Mainly directory location changes
** double asterisks denote changes from JVHs program
clear
set mem 500m
set matsize 400
set more off

log using REmigration706.2.log, replace

/* set the default directory */
*cd "C:\Documents and Settings\vanhook\My Documents\Jen\Sabre\"
cd "/iss2/JVH/EMS2stdat/"

use h* yr state /* f* l* w* */ *seq lineno mis a? agegrp male generation avgpoe mpob fpob
migsame grdatn enroll agegrp elig intmig age goodmatch raceth white h? mexican oversamp
ownhome black hisp rpob prcitshp hea parent gen wgt yrarriv if generation == 1 |
generation==2 using combined.dta,clear

/* origin for 2nd generation */
gen c = 0
replace c = mpob if generation==2
replace c = fpob if generation==2 & mpob <=72 & fpob >72

gen origin = 0
replace origin=1 if c == 315 /* mexican */
replace origin=3 if origin==0 & c == 337 /* cuba */
replace origin=4 if origin==0 & c == 339 /* Dom Rep */
replace origin=5 if origin==0 & (c ==207 | c==238 | c==209) /*China */
replace origin=6 if origin==0 & c == 231 /* Philippines */
replace origin=7 if origin==0 & (c == 217 | c == 218) /* ??? */
replace origin=8 if origin==0 & c == 210 /*India */
replace origin=2 if origin==0 & (c ==310 | c==311 | c==312 | c==313 | c==314 | c==315 |
c==316 | c==317 | c==318) /* Central Am */
replace origin=10 if origin==0 & (c==417 | c==436 | c==468 | c==252) /* North Africa */
replace origin=9 if origin==0 & c >= 200 & c <= 253 /* Asia */
replace origin=11 if origin==0 & c >= 417 & c <= 462 /* Africa */
replace origin=12 if origin==0 & c >= 103 & c <= 148 /* Europe */
replace origin=13 if origin==0 & c >= 300 & c <= 304 /* North Am */
replace origin=14 if origin==0 & c >= 333 & c <= 389 /* Caribbean */
replace origin=1 if origin==0 & c==555 & mexican==1 /* mexican */
replace origin = 15 if origin==0 & generation==2
replace origin = rpob if generation==1

/* combine some categories */
replace origin = 9 if origin==7 | origin==6
replace origin = 11 if origin==10
replace origin = 14 if origin==4

tab origin, gen(or)

capture gen hhseq = seq
sort yr state hhseq lineno

```

```

gen lths=(grdatn<39)
gen hs=(grdatn==39)
gen somecol=(grdatn==40|grdatn==41|grdatn==42)
gen yr97=0
replace yr97=1 if yr==1997
gen yr98=0
replace yr98=1 if yr==1998
gen yr99=0
replace yr99=1 if yr==1999
gen yr00=0
replace yr00=1 if yr==2000
gen yr01=0
replace yr01=1 if yr==2001
gen yr02=0
replace yr02=1 if yr==2002
gen yr03=0
replace yr03=1 if yr==2003
gen yr04=0
replace yr04=1 if yr==2004
gen en1=0
replace en1=1 if enroll==1
gen en2=0
replace en2=1 if enroll==2

gen amale = 1 if male==1 & agegrp>=3
replace amale = 0 if male == 0 & agegrp >=3
replace amale = 2 if agegrp <3

gen matchsamp=0
replace matchsamp=1 if elig==1
gen intsamp=0
replace intsamp=1 if intmig !=.
replace intmig = . if intsamp==0 | age==0
replace goodmatch=. if matchsamp==0

gen asian=0
replace asian=1 if raceth==3 | raceth==6

gen h = 0
replace h=1 if origin==1
gen nmex=0
replace nmex=1 if origin !=1

gen age1=0
gen age2=0
gen age3=0
gen age4=0
gen age5=0

replace age1=1 if agegrp <3
replace age2=1 if agegrp >=3 & agegrp <=4
replace age3=1 if agegrp >=5 & agegrp <=6
replace age4=1 if agegrp >=7 & agegrp <=8

```

```

replace age5=1 if agegrp >=9 & agegrp <=12
gen agela = (agegrp==0)
gen agelb = (agegrp==1)
gen agelc = (agegrp==2)

```

```

gen agrp =0
replace agrp=1 if agegrp >=0 & agegrp <3
replace agrp=2 if agegrp >=3 & agegrp <5
replace agrp=3 if agegrp >=5 & agegrp <7
replace agrp=4 if agegrp >=7 & agegrp <9
replace agrp=5 if agegrp >=9 & agegrp <13
replace agrp=6 if agegrp >=13

```

```

drop mexican
gen mexican = (nmex==0)
gen rhisp=hisp
drop hisp
gen hisp = (rhisp == 1 & mexican==0)
drop rhisp

```

```
set more off
```

```
quietly do parentvars706.do
```

```

global variables "oversamp ownhome or3-or11 male age2-age5 yr97 yr98 yr99 yr00 yr01 yr02
yr03 yr04 en1 en2 "
global kvariables "oversamp ownhome male or3-or11 agela agelb pagel-page5 yr97 yr98 yr99
yr00 yr01 yr02 yr03 yr04 plths phs psomecol "
**This program makes the paramter Probability of INTERNAL MIGRATION, TO ANY NEW HOUSE;
quietly do intmodels706v2.do

```

```

**GOOD match- this program predicts whether or not MATCHED across years;
global variables "oversamp ownhome or3-or11 male age2-age5 yr97 yr98 yr99 yr00 yr01 yr02
yr03 yr04 en1 en2 lths hs somecol "
global kvariables "pgoodmatch oversamp ownhome male or3-or11 agela agelb pagel-page5
yr97 yr98 yr99 yr00 yr01 yr02 yr03 yr04 plths phs psomecol "
quietly do gmmmodels706v2.do

```

```

**This program takes all parameters to form predicted values;
** then use to ESTIMATE EMIGRATION RATES;
** Predicted values are by generation, sex, major sending group (Mexican, not Mexican);
** NATIVE emigration rates from Census-Fernandez 1995 and Gibbs et al. 2003;
** Mortality data from Natl Health Intv Svy, pr of dying IN THE US, ?avgd across 1989-
1993;
** Combined set of annual surveys and compare to health data and see if show up;
** Then JVH with parameters and survival analysis;
quietly do rc706v2.do

```

```

**This program takes kids and matches to a parent, then makes kids' prob that of parents'
;
** This is also b/c 1st and 2nd generation kids are similar- if a CHILD, follow parents ;
quietly do parents706v2.do

```

```
tab avgpoe
```

```

capture drop r ar
capture gen r = 0
replace r=0
replace r = 1 if migsame==3 & avgpoe >1.5
replace r = . if migsame==0

capture gen ar=r*(1-(dmolnm+aaeerbnm)) if gen==1
replace ar=r*(1-(dmolnm+aaeerbnm)) if gen==1

/* cob has less detailed country categories */
gen cob = rpob
**combine smaller groups
replace cob = 10 if rpob==11 /* all africa */
replace cob = 14 if rpob==4 | rpob==3 /* caribbean */
replace cob=9 if rpob==5 | rpob==6 | rpob == 7|rpob==8 | rpob==9
**put all Asian pobs together

**adjusted for emigration and mortality
** ar from rc6.do
gen anet = aaeerbnm-ar

/* citizenship */
gen cit = 1 if prcitshp == 4
replace cit = 0 if prcitshp==5

/* health */
gen healthy = (hea == 1 | hea==2)

drop if age == 0

/* combine children */
replace agegrp = 1 if agegrp <3

/* merge on migration status from combinedui.dta */
sort yr state hhseq lineno
capture drop _merge
save all, replace

use yr state *seq lineno iteratn migstat ui* using combinedui, clear
capture gen hhseq = seq
sort yr state hhseq lineno
capture drop _merge
merge yr state hhseq lineno using all
drop if _merge==1

gen regsamp = (iteratn==1 | iteratn==0 | iteratn==.) & generation==1
gen uisamp = (migstat>1 & generation==1)
gen regsamp2 = (iteratn==1 | iteratn==0 | iteratn==.) & generation==1 & _merge==3

drop if regsamp==0 & uisamp==0

```

```

/* divide weights by 1,000 for display in tables */
replace wgt = wgt/1000
replace ui_wgt = ui_wgt/1000
gen combo = raceth+100*generation+1000*male

gen nonmatch = (goodmatch==0)

keep migsame aa* d* mis wgt generation regsamp uisamp eerb nml nonmatch iml mol anf r
ar anet ui_wgt mexican origin agegrp yr mexican male yrarriv rpob cob cit migstat healthy
combo intmig goodmatch

gen agegrpb = 1 if agegrp == 1
replace agegrpb = 2 if agegrp == 3 | agegrp==4
replace agegrpb = 3 if agegrp == 5 | agegrp==6
replace agegrpb = 4 if agegrp== 7 | agegrp==8
replace agegrpb = 5 if agegrp == 9 | agegrp==10
replace agegrpb = 6 if agegrp == 11 | agegrp==12
replace agegrpb = 7 if agegrp >=13

/* Table 4 */
table generation [aw=wgt] if generation==1 & regsamp==1, m c(m aaerbnm m ar m r m anet m
nonmatch )
table generation [aw=wgt] if generation==1 & regsamp==1, m c(m daimlnm m dmolnm m danfnm
rawsum wgt)

/* Table 5 */
table yr [aw=wgt] if generation==1 & regsamp==1 , m c(m aaerbnm m ar m anet m nonmatch)
table yr [aw=wgt] if generation==1 & regsamp==1 , m c(m daimlnm m dmolnm m danfnm rawsum
wgt)

table yr [aw=wgt] if generation==1 & regsamp==1 & mexican==1 , m c(m aaerbnm m ar m anet
m nonmatch)
table yr [aw=wgt] if generation==1 & regsamp==1 & mexican==1 , m c(m daimlnm m dmolnm m
danfnm rawsum wgt)

table yr [aw=wgt] if generation==1 & regsamp==1 & mexican==0 , m c(m aaerbnm m ar m anet
m nonmatch)
table yr [aw=wgt] if generation==1 & regsamp==1 & mexican==0 , m c(m daimlnm m dmolnm m
danfnm rawsum wgt)

/* Table 6 */
table agegrp [aw=wgt] if generation==1 & regsamp==1 , m c(m aaerbnm m ar m anet m
nonmatch)
table agegrp [aw=wgt] if generation==1 & regsamp==1 , m c(m daimlnm m dmolnm m danfnm
rawsum wgt)

/* Table 7 */
table male [aw=wgt] if generation==1 & regsamp==1 , m c(m aaerbnm m ar m anet m
nonmatch)
table male [aw=wgt] if generation==1 & regsamp==1 , m c(m daimlnm m dmolnm m danfnm
rawsum wgt)

table male [aw=wgt] if generation==1 & regsamp==1 & mexican==1 , m c(m aaerbnm m ar m
anet m nonmatch)
table male [aw=wgt] if generation==1 & regsamp==1 & mexican==1 , m c(m daimlnm m dmolnm m
danfnm rawsum wgt)

```

```

table male [aw=wgt] if generation==1 & regsamp==1 & mexican==0 , m c(m aaeerbnm m ar m
anet m nonmatch)
table male [aw=wgt] if generation==1 & regsamp==1 & mexican==0 , m c(m daimlnm m dmolnm m
danfnm rawsum wgt)

/* Table 8 */
table agegrp [aw=wgt] if generation==1 & regsamp==1 & male==1 , m c(m aaeerbnm m ar m
anet m nonmatch)
table agegrp [aw=wgt] if generation==1 & regsamp==1 & male==1 , m c(m daimlnm m dmolnm m
danfnm rawsum wgt)

/* Table 9 */
table agegrp [aw=wgt] if generation==1 & regsamp==1 & male==0 , m c(m aaeerbnm m ar m
anet m nonmatch)
table agegrp [aw=wgt] if generation==1 & regsamp==1 & male==0 , m c(m daimlnm m dmolnm m
danfnm rawsum wgt)

/* Table 10 */
table agegrpb [aw=wgt] if generation==1 & regsamp==1 & male==1 & mexican==1, m c(m
aaeerbnm m ar m anet m nonmatch)
table agegrpb [aw=wgt] if generation==1 & regsamp==1 & male==1 & mexican==1 , m c(m
daimlnm m dmolnm m danfnm rawsum wgt)

table agegrpb [aw=wgt] if generation==1 & regsamp==1 & male==0 & mexican==1 , m c(m
aaeerbnm m ar m anet m nonmatch)
table agegrpb [aw=wgt] if generation==1 & regsamp==1 & male==0 & mexican==1 , m c(m
daimlnm m dmolnm m danfnm rawsum wgt)

/* Table 11 */
table agegrpb [aw=wgt] if generation==1 & regsamp==1 & male==1 & mexican==0, m c(m
aaeerbnm m ar m anet m nonmatch)
table agegrpb [aw=wgt] if generation==1 & regsamp==1 & male==1 & mexican==0 , m c(m
daimlnm m dmolnm m danfnm rawsum wgt)

table agegrpb [aw=wgt] if generation==1 & regsamp==1 & male==0 & mexican==0 , m c(m
aaeerbnm m ar m anet m nonmatch)
table agegrpb [aw=wgt] if generation==1 & regsamp==1 & male==0 & mexican==0 , m c(m
daimlnm m dmolnm m danfnm rawsum wgt)

/* Table 12 */
table yrarriv [aw=wgt] if generation==1 & regsamp==1 , m c(m aaeerbnm m ar m anet m
nonmatch)
table yrarriv [aw=wgt] if generation==1 & regsamp==1 , m c(m daimlnm m dmolnm m danfnm
rawsum wgt)

table yrarriv [aw=wgt] if generation==1 & regsamp==1 & mexican==1 , m c(m aaeerbnm m ar m
anet m nonmatch)
table yrarriv [aw=wgt] if generation==1 & regsamp==1 & mexican==1 , m c(m daimlnm m
dmolnm m danfnm rawsum wgt)

table yrarriv [aw=wgt] if generation==1 & regsamp==1 & mexican==0 , m c(m aaeerbnm m ar m
anet m nonmatch)
table yrarriv [aw=wgt] if generation==1 & regsamp==1 & mexican==0 , m c(m daimlnm m
dmolnm m danfnm rawsum wgt)

/* Table 13 */

```

```

table rpob [aw=wgt] if generation==1 & regsamp==1 , m c(m aaeerbnm m ar m anet m
nonmatch)
table rpob [aw=wgt] if generation==1 & regsamp==1 , m c(m daimlnm m dmolnm m danfnm
rawsum wgt)

table cob [aw=wgt] if generation==1 & regsamp==1 , m c(m aaeerbnm m ar m anet m nonmatch)
table cob [aw=wgt] if generation==1 & regsamp==1 , m c(m daimlnm m dmolnm m danfnm rawsum
wgt)

/* Table 14 */
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 , m c(m aaeerbnm m ar m anet m
nonmatch)
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 , m c(m daimlnm m dmolnm m danfnm
rawsum ui_wgt)

table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==1 , m c(m aaeerbnm m ar
m anet m nonmatch)
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==1 , m c(m daimlnm m
dmolnm m danfnm rawsum ui_wgt)

table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==0 , m c(m aaeerbnm m ar
m anet m nonmatch)
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==0 , m c(m daimlnm m
dmolnm m danfnm rawsum ui_wgt)

/* Table 15 */
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & (yr == 1996 | yr==1998) , m c(m
aaeerbnm m anet )
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & (yr == 1999 | yr==2000) , m c(m
aaeerbnm m anet )
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & (yr == 2001 | yr==2002) , m c(m
aaeerbnm m anet )
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & (yr == 2003) , m c(m aaeerbnm m
anet )

/* Table 16 */
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==1 & yr <2000, m c(m
aaeerbnm m ar m anet m nonmatch )
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==1 & yr <2000, m c(m
daimlnm m dmolnm m danfnm rawsum ui_wgt)

table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==1 & yr >=2000, m c(m
aaeerbnm m ar m anet m nonmatch )
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==1 & yr >=2000, m c(m
daimlnm m dmolnm m danfnm rawsum ui_wgt)

table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==0 & yr <2000, m c(m
aaeerbnm m ar m anet m nonmatch )
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==0 & yr <2000, m c(m
daimlnm m dmolnm m danfnm rawsum ui_wgt)

table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==0 & yr >=2000, m c(m
aaeerbnm m ar m anet m nonmatch )
table migstat [aw=ui_wgt] if generation==1 & uisamp==1 & mexican==0 & yr >=2000, m c(m
daimlnm m dmolnm m danfnm rawsum ui_wgt)

log close

```

**Appendix C. STATA code, sub-routine 1 (parentvars706.do)**

```

save all, replace
drop if agegrp <=2
save adults, replace

capture gen pintmig = 0
replace pintmig = intmig

capture gen pgoodmatch = 0
replace pgoodmatch = goodmatch

capture gen page = 0
replace page = age
capture gen pen1 = 0
replace pen1 = en1
capture gen pen2 = 0
replace pen2 = en2
capture gen plths = 0
replace plths = lths
capture gen phs = 0
replace phs = hs
capture gen psomecol = 0
replace psomecol = somecol
capture gen hhseq = seq
keep pgoodmatch pintmig page pen1 pen2 plths phs psomecol yr state hhseq lineno
sort yr state hhseq lineno
save temp, replace

use all, clear
gen clineno = lineno
drop lineno
gen lineno=parent
capture gen hhseq = seq
drop if agegrp >2
sort yr state hhseq lineno
merge yr state hhseq lineno using temp
drop if _merge == 1 | _merge==2

gen page1=0
gen page2=0
gen page3=0
gen page4=0
gen page5=0
replace page1=1 if page <30
replace page2=1 if page>=30 & page <35
replace page3=1 if page >=35 & page<45
replace page4=1 if page >=45 & page <55
replace page5=1 if page >=55

drop parent
gen parent = lineno
replace lineno = clineno

append using adults

```

**Appendix D. STATA code, sub-routine 2 (intmodels706v2.do)**

```

global r "h"
global g "1"
global m "1"
    logit intmig $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global m "0"
    logit intmig $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global m "2"
    logit intmig $kvariables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global g "2"
global m "1"
    logit intmig $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global m "0"
    logit intmig $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global m "2"
    logit intmig $kvariables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)

global r "nmex"
global g "1"
global m "1"
    logit intmig $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global m "0"
    logit intmig $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global m "2"
    logit intmig $kvariables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global g "2"
global m "1"
    logit intmig $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global m "0"
    logit intmig $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)
global m "2"
    logit intmig $kvariables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bi$r$g$m=get(_b)

```

**Appendix E. STATA code, sub-routine 3 (gmmodels706v2.do)**

```

global r "h"
global g "1"
global m "1"
    logit goodmatch $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bm$r$g$m=get(_b)
global m "0"
    logit goodmatch $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bm$r$g$m=get(_b)
global m "2"
    logit goodmatch $kvariables [pw=wgt] if generation== $g & $r ==1 & amale==
$m
matrix bm$r$g$m=get(_b)
global g "2"
global m "1"
    logit goodmatch $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bm$r$g$m=get(_b)
global m "0"
    logit goodmatch $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bm$r$g$m=get(_b)
global m "2"
    logit goodmatch $kvariables [pw=wgt] if generation== $g & $r ==1 & amale==
$m
matrix bm$r$g$m=get(_b)

global r "nmex"
global g "1"
global m "1"
    logit goodmatch $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bm$r$g$m=get(_b)
global m "0"
    logit goodmatch $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bm$r$g$m=get(_b)
global m "2"
    logit goodmatch $kvariables [pw=wgt] if generation== $g & $r ==1 & amale==
$m
matrix bm$r$g$m=get(_b)
global g "2"
global m "1"
    logit goodmatch $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bm$r$g$m=get(_b)
global m "0"
    logit goodmatch $variables [pw=wgt] if generation== $g & $r ==1 & amale== $m
matrix bm$r$g$m=get(_b)
global m "2"
    logit goodmatch $kvariables [pw=wgt] if generation== $g & $r ==1 & amale==
$m
matrix bm$r$g$m=get(_b)

```

**Appendix F. STATA code, sub-routine 4 (rc706v2.do)**

```

capture drop prm* pri*
global r "h"
global g "1"

global m "1"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))

global m "0"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))
global m "2"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))

global g "2"

global m "1"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))
global m "0"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))

global m "2"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))

global r "nmex"
global g "1"
global m "1"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))

global m "0"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))
global m "2"

```

```

matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))
global g "2"
global m "1"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))
global m "0"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))
global m "2"
matrix score pri$r$g$m =bi$r$g$m
replace pri$r$g$m=exp(pri$r$g$m)/(1+exp(pri$r$g$m))
matrix score prm$r$g$m =bm$r$g$m
replace prm$r$g$m=exp(prm$r$g$m)/(1+exp(prm$r$g$m))

/* non-match */
capture gen nml = 0
capture gen nm2 = 0
replace nml=0
replace nm2=0
replace nml=(1-prmh11) if amale==1 & (mexican==1 )
replace nml=(1-prmh10) if amale==0 & (mexican==1 )
replace nml=(1-prmh12) if amale==2 & (mexican==1 )

replace nml=(1-prmmex11) if amale==1 & (mexican==0 )
replace nml=(1-prmmex10) if amale==0 & (mexican==0 )
replace nml=(1-prmmex12) if amale==2 & (mexican==0 )

replace nm2=(1-prmh21) if amale==1 & (mexican==1 )
replace nm2=(1-prmh20) if amale==0 & (mexican==1 )
replace nm2=(1-prmh22) if amale==2 & (mexican==1 )
replace nm2=(1-prmmex21) if amale==1 & (mexican==0 )
replace nm2=(1-prmmex20) if amale==0 & (mexican==0 )
replace nm2=(1-prmmex22) if amale==2 & (mexican==0 )

/* native emigration */
capture gen naem=0
replace naem=0
replace naem=0.0003319776 if male==1 & agegrp == 0
replace naem=0.0004643357 if male==1 & agegrp ==1
replace naem=0.0003381199 if male==1 & agegrp ==2
replace naem=0.0001483905 if male==1 & agegrp ==3
replace naem=0.0001160999 if male==1 & agegrp ==4
replace naem=0.0002011312 if male==1 & agegrp ==5
replace naem=0.0003694912 if male==1 & agegrp ==6
replace naem=0.0002933757 if male==1 & agegrp ==7
replace naem=0.0001806170 if male==1 & agegrp ==8
replace naem=0.0001423468 if male==1 & agegrp ==9
replace naem=0.0001668840 if male==1 & agegrp ==10
replace naem=0.0000975946 if male==1 & agegrp ==11
replace naem=0.0001278402 if male==1 & agegrp ==12

```

```

replace naem=0.0000702032 if male==1 & agegrp ==13
replace naem=0.0000914651 if male==1 & agegrp ==14
replace naem=0.0002688526 if male==1 & agegrp ==15

replace naem=0.0003263111 if male==0 & agegrp==0
replace naem=0.0004471334 if male==0 & agegrp==1
replace naem=0.0003059633 if male==0 & agegrp==2
replace naem=0.0001266721 if male==0 & agegrp==3
replace naem=0.0001741828 if male==0 & agegrp==4
replace naem=0.0002721120 if male==0 & agegrp==5
replace naem=0.0003155066 if male==0 & agegrp==6
replace naem=0.0001621669 if male==0 & agegrp==7
replace naem=0.0000974636 if male==0 & agegrp==8
replace naem=0.0000733140 if male==0 & agegrp==9
replace naem=0.0000798113 if male==0 & agegrp==10
replace naem=0.0000583690 if male==0 & agegrp==11
replace naem=0.0001009648 if male==0 & agegrp==12
replace naem=0.0000538987 if male==0 & agegrp==13
replace naem=0.0000968418 if male==0 & agegrp==14
replace naem=0.0001761570 if male==0 & agegrp==15

/* mortality */

replace a1 = 1 if agegrp == 3

capture gen mol=-2.0172+0.5172*male+0.3359*mexican+0.0671*hispanic+0.2304*white+0.1763*black-
3.9841*a1-3.7220*a2-3.3239*a3-2.7561*a4-1.8177*a5-0.9944*a6-1.1810*h1-1.1345*h2-
0.9374*h3-0.5822*h4
replace mol=-2.0172+0.5172*male+0.3359*mexican+0.0671*hispanic+0.2304*white+0.1763*black-
3.9841*a1-3.7220*a2-3.3239*a3-2.7561*a4-1.8177*a5-0.9944*a6-1.1810*h1-1.1345*h2-
0.9374*h3-0.5822*h4
replace mol=exp(mol)/((1+exp(mol)))

capture gen mo2=-1.7108+0.5314*male+0.0439*mexican+0.2039*hispanic+0.2048*white+0.2967*black-
4.0421*a1-3.7818*a2-3.1487*a3-2.3296*a4-1.5351*a5-0.8194*a6-1.7112*h1-1.5010*h2-
1.1274*h3-0.6526*h4
replace mo2=-1.7108+0.5314*male+0.0439*mexican+0.2039*hispanic+0.2048*white+0.2967*black-
4.0421*a1-3.7818*a2-3.1487*a3-2.3296*a4-1.5351*a5-0.8194*a6-1.7112*h1-1.5010*h2-
1.1274*h3-0.6526*h4
replace mo2=exp(mo2)/((1+exp(mo2)))
replace mol=0 if agegrp <3
replace mo2=0 if agegrp <3

/* internal migration */
capture gen im1=0
replace im1=0
replace im1=((prih11)) if amale==1 & h==1
replace im1=((prih10)) if amale==0 & h==1
replace im1=((prih12)) if amale==2 & h==1

replace im1=((prinmex11)) if amale==1 & nmex==1
replace im1=((prinmex10)) if amale==0 & nmex==1
replace im1=((prinmex12)) if amale==2 & nmex==1

capture gen im2=0
replace im2=0

```

```

replace im2=((prih21)) if amale==1 & h==1
replace im2=((prih20)) if amale==0 & h==1
replace im2=((prih22)) if amale==2 & h==1
replace im2=((prinmex21)) if amale==1 & nmex==1
replace im2=((prinmex20)) if amale==0 & nmex==1
replace im2=((prinmex22)) if amale==2 & nmex==1

gen asamp = (generation==1 & matchsamp==1 & agegrp >2)

/* residual non-follow-up */
capture gen anf = 0
replace anf=0
replace anf = nm2 - im2*(1-mo2)-mo2 if asamp

capture gen eerb=0
replace eerb=(nm1-im1*(1-mo1)-mo1-anf)/(1-im1) if asamp

gen aim1 = nm1-eerb - mo1 - anf if asamp

/* shift mo, im, and nf values to non-matches */

capture program drop nmonly
program nmonly
version 8
egen m = total($x*wgt*all), by(amale rpob agegrp)
egen mnm = total($x*wgt*nmall), by(amale rpob agegrp)
gen $xnm = 0 if all
replace $xnm = $x*m/mnm if nmall
drop m mnm
end

gen all = (asamp==1)
gen nmall = (asamp==1 & goodmatch==0 )

global x "im1"
global xnm "im1nm"
nmonly

global x "aim1"
global xnm "aim1nm"
nmonly

global x "im2"
global xnm "im2nm"
nmonly

global x "mo1"
global xnm "mo1nm"
nmonly

global x "mo2"
global xnm "mo2nm"

```

```

nmonly

global x "anf"
global xnm "anfnm"
nmonly

global x "eerb"
global xnm "eerbnm"
nmonly

/* pull in out-of-range values */

capture program drop outrangenm
program outrangenm, rclass
version 8
egen xsum = total($x*wgt) if generation==1 & matchsamp==1 & agegrp >2 & goodmatch==0,
by(amale rpob agegrp)
egen xdem = total(wgt) if generation==1 & matchsamp==1 & agegrp >2 & goodmatch==0,
by(amale rpob agegrp)
gen xbar = xsum/xdem if generation==1 & matchsamp==1 & agegrp >2 & goodmatch==0
egen xmax = max($x) if generation==1 & matchsamp==1 & agegrp >2 & goodmatch==0,
by(amale rpob agegrp)
egen xmin = min($x) if generation==1 & matchsamp==1 & agegrp >2 & goodmatch==0,
by(amale rpob agegrp)
gen minsc = min(((1-xbar)/(xmax - xbar)),((xbar)/(xbar-xmin))) if generation==1 &
matchsamp==1 & agegrp >2 & goodmatch==0
gen adjx = 0 if asamp
replace adjx = 0 if asamp==1 & xbar >=0 & xbar <=1
replace adjx = xbar+($x-xbar)*minsc if asamp==1 & xbar >=0 & xbar <=1 & goodmatch==0
replace adjx = 0 if xbar <0 & asamp==1 & goodmatch==0
replace adjx = 1 if xbar >1 & asamp==1 & goodmatch==0
egen rat = total($x*wgt) if asamp
egen arat = total(adjx*wgt) if asamp
gen aa$x = adjx * rat/arar if asamp
replace aa$x = 0 if aa$x <0
replace aa$x = 1 if aa$x >1
drop adjx xsum xdem xbar xmax xmin minsc rat arar
end

global x "eerbnm"
outrangenm

global x "molnm"
outrangenm

global x "anfnm"
outrangenm

global x "aimlnm"
outrangenm

gen sel = aaeerbnm + aamolnm + aanfnm + aaaimlnm if asamp
gen sel2 = aamolnm + aanfnm + aaaimlnm if asamp

gen dmolnm = 0 if asamp
gen danfnm = 0 if asamp

```

```
gen daim1nm = 0 if asamp
replace dmolnm = (aamolnm/sel2) * (1-aaeerbnm) if asamp & goodmatch==0 & aaeerbnm<1
replace danfnm = (aaanfnm/sel2) * (1-aaeerbnm) if asamp & goodmatch==0 & aaeerbnm<1
replace daim1nm = (aaaim1nm/sel2) * (1-aaeerbnm) if asamp & goodmatch==0 & aaeerbnm<1
```

**Appendix G. STATA code, sub-routine 5 (parents706v2.do)**

```

save all, replace
drop if agegrp <=2
capture gen panf = 0
replace panf = anf
capture gen peerb = 0
replace peerb = eerb

capture gen piml = 0
replace piml = iml
save parents.dta, replace

keep panf peerb par piml yr state hhseq lineno
sort yr state hhseq lineno
save temp, replace

use all, clear
drop if agegrp >2
capture gen clineno = lineno
capture drop lineno
capture drop _merge
gen lineno=parent
capture gen hhseq = seq
drop if lineno ==0 | lineno==.
sort yr state hhseq lineno
merge yr state hhseq lineno using temp
drop if _merge ==2
replace lineno = clineno
gen kiml = iml

drop anf iml eerb aiml all nml iml nm aiml nm im2nm molnm mo2nm anfnm eerbnm aa*nm sel
sel2 d*nm clineno

/* residual non-follow-up */
capture gen anf = 0
replace anf=panf
capture gen iml = 0
replace iml = piml
*replace iml = kiml

gen bsamp = (generation==1 & matchsamp==1 & agegrp <=2 & age>0)

/* residual non-follow-up & internal migration : take from parent */

capture gen eerb=0 if bsamp
replace eerb=(nml-iml-anf)/(1-iml) if bsamp

gen aiml = nml-eerb - mol - anf if bsamp

/* shift mo, im, and nf values to non-matches */

capture program drop nmonly
program nmonly

```

```

version 8
egen m = total($x*wgt*all), by(amale rpob agegrp)
egen mnm = total($x*wgt*nmall), by(amale rpob agegrp)
gen $xnm = 0 if all
replace $xnm = $x*m/mnm if nmall
drop m mnm
end

gen all = (bsamp==1)
gen nmall = (bsamp==1 & goodmatch==0 )

global x "im1"
global xnm "im1nm"
nmonly

global x "aim1"
global xnm "aim1nm"
nmonly

global x "im2"
global xnm "im2nm"
nmonly

global x "anf"
global xnm "anfnm"
nmonly

global x "eerb"
global xnm "eerbnm"
nmonly

/* pull in out-of-range values */

capture program drop outrangenm
program outrangenm, rclass
version 8
egen xsum = total($x*wgt) if bsamp==1 & goodmatch==0, by(amale rpob agegrp)
egen xdem = total(wgt) if bsamp==1 & goodmatch==0, by(amale rpob agegrp)
gen xbar = xsum/xdem if bsamp==1 & goodmatch==0
egen xmax = max($x) if bsamp==1 & goodmatch==0, by(amale rpob agegrp)
egen xmin = min($x) if bsamp==1 & goodmatch==0, by(amale rpob agegrp)
gen minsc = min(((1-xbar)/(xmax - xbar)),((xbar)/(xbar-xmin))) if bsamp==1 &
goodmatch==0
gen adjx = 0 if bsamp
replace adjx = 0 if bsamp==1 & xbar >=0 & xbar <=1
replace adjx = xbar+($x-xbar)*minsc if bsamp==1 & xbar >=0 & xbar <=1 & goodmatch==0
replace adjx = 0 if xbar <0 & bsamp==1 & goodmatch==0
replace adjx = 1 if xbar >1 & bsamp==1 & goodmatch==0
egen rat = total($x*wgt) if bsamp
egen arat = total(adjx*wgt) if bsamp
gen aa$x = adjx * rat/arat if bsamp

```

```
replace aa$x = 0 if aa$x <0
replace aa$x = 1 if aa$x >1
drop adjx xsum xdem xbar xmax xmin minsc rat arat
end

global x "eerbnm"
outrangenm

global x "anfnm"
outrangenm

global x "aimlnm"
outrangenm

gen sel = aaeerbnm + aaanfnm + aaaaimlnm if bsamp
gen sel2 = aaanfnm + aaaaimlnm if bsamp

gen dmolnm = 0 if bsamp
gen danfnm = 0 if bsamp
gen daimlnm = 0 if bsamp
replace danfnm = (aaanfnm/sel2) * (1-aaeerbnm) if bsamp & goodmatch==0 & aaeerbnm<1
replace daimlnm = (aaaaimlnm/sel2) * (1-aaeerbnm) if bsamp & goodmatch==0 & aaeerbnm<1

append using parents
replace bsamp = 0 if bsamp==.
```