http://www.bgsu.edu/organizations/cfdr/
Phone: (419) 372-7279 cfdr@bgnet.bgsu.edu

Bowling Green State University

Working Paper Series 2005-01

Foreign-Born Emigration:

A New Approach and Estimates Based on Matched CPS File

Jennifer Van Hook
Center for Family and Demographic Research
Bowling Green State University

Weiwei Zhang
Center for Family and Demographic Research
Bowling Green State University

Frank D. Bean
Department of Sociology
University of California--Irvine

Jeffrey S. Passel
The Pew Hispanic Center

FOREIGN-BORN EMIGRATION: A NEW APPROACH AND ESTIMATES BASED ON MATCHED CPS FILES

Jennifer Van Hook*
Center for Family and Demographic Research
Bowling Green State University

Weiwei Zhang
Center for Family and Demographic Research
Bowling Green State University

Frank D. Bean
Department of Sociology
University of California--Irvine

Jeffrey S. Passel
The Pew Hispanic Center

September 2005

Forthcoming in *Demography*

*Jennifer Van Hook, Department of Sociology, Bowling Green State University, Bowling Green, OH 43403, E-mail: vanhook@bgnet.bgsu.edu. This research was supported in part by grants provided by the U.S. Census Bureau and NICHD [HD-39075]. Infrastructure support was provided by a center grant to the Center for Family and Demographic Research from the National Institutes of Health [HD-42831]. We are grateful to Bert Kestenbaum, Samuel Preston, and Elizabeth Grieco for their insightful suggestions and comments.

Foreign-Born Emigration:
A New Approach and Estimates Based on Matched CPS Files

Abstract

The utility of postcensal population estimates depends on the adequate measurement of four major components of demographic change—fertility, mortality, immigration, and emigration. Of the four components, emigration, especially of the foreign-born, has proved the most difficult to gauge. Without "direct" methods (i.e., those identifying individuals who emigrate and when), demographers have relied on indirect approaches, such as residual methods. Residual estimates, however, are sensitive to inaccuracies in their constituent parts, as well as particularly ill-suited for measuring the emigration of recent arrivals. Here we introduce a new method for estimating foreign-born emigration that takes advantage of the sample design of the Current Population Survey (CPS)—repeated interviews of persons in the same housing units over a period of 16 months. Individuals appearing in a first March Supplement to the CPS but not the next include those who died in the intervening year, those who moved within the country, and those who emigrated. We use statistical methods to estimate the proportion of emigrants among those not present in the follow-up interview. Our method produces emigration estimates that are comparable to those from residual methods in the case of longer-term residents (immigrants who arrived more than 10 years ago), but yields higher—and what appear to be more accurate—estimates for recent arrivals. Although somewhat constrained by sample size, we also generate estimates by age, sex, region of birth, and duration of residence.

FOREIGN-BORN EMIGRATION: A NEW APPROACH AND ESTIMATES BASED ON MATCHED CPS FILES

Of the components of demographic change, emigration of the foreign-born has proved the most difficult to gauge. Arising from the lack of information about persons who have moved from the country, 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 using the cohort component method will be too low if emigration is overestimated and too high if it is underestimated. 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 are too high (low) when emigration of the legal foreign born is over (under) estimated" (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 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 data on them. To be certain, the data on the number of legal visas in a given year may not reflect the number of arrivals in that year since the person receiving the visa may have already been living in the country in some capacity other than that of legal permanent resident (either as a non-immigrant [e.g., on a student, work, or some other temporary visa] or as an unauthorized migrant). This can complicate residual estimation since such persons may nonetheless be included in data sources like the Current Population Survey (CPS), either erroneously or through the fulfillment of the survey's residency requirements. Their presence in data provides but one example of the kinds of obstacles that can confront the estimation of annual legal immigration. Despite such difficulties, however, most efforts to measure annual legal immigration appear to have generated reasonably accurate results (Bean, et al, 1998; Woodrow-Lafield 1995, 1998; Passel, Van Hook and Bean, 2004).

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). Jasso and Rosenzweig (1990) analyzed annual changes in the number of legal non-citizens living in the United States (adjusted for the numbers of new legal immigrants and naturalizations) in order to gauge relative levels of and upper limits to emigration. Jasso and Rosenzweig (1990) estimated the number of legal non-citizens in the U.S. using Alien Address Registration data (INS administrative records of the number and locations of legal non-citizens living in the U.S.). However, this approach was no longer possible after 1981 when the INS discontinued the alien address registration program. 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.

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 earlier census. We develop below an alternative method for estimating emigration that takes advantage of the longitudinal nature of the Current Population Survey. By producing new foreign-born emigration estimates with which residual estimates can be compared, we offer a new way to assess and update previous emigration estimates and potentially to improve the production of population estimates, particularly for foreign-born who have recently arrived in the U.S.

PREVIOUS RESEARCH

Warren and Peck (1980) first developed indirect methods for estimating emigration that occurred during the 1960s. Their technique—referred to here as the residual method—has since served as the major approach for developing and updating foreign-born emigration estimates. Warren and Peck (1980) initially estimated annual emigration of 114,000 for the 1960–70 decade. This figure was later increased to 133,000 on the basis of Warren and Passel's (1987) analysis of INS Alien Address Registration data for 1965-80. The 133,000 figure was used as an official annual "point estimate" by the U.S. Census Bureau until the mid-1990s, when the number was increased to 195,000 based on residual estimates for the 1980s developed by Ahmed and Robinson (1994). The only estimates to our knowledge for the 1990s were developed by

Mulder (Mulder et al. 2002; Mulder 2003), which indicated an estimated annual foreign-born emigration of 225,000. Mulder's estimates were never used in official Census Bureau measures during the 1990s, however, when the Bureau continued to rely on the Ahmed and Robinson (1994) estimates.

The residual method for estimating emigration in the decade between two censuses involves the comparison of two population figures: (1) the "expected" foreign-born population if no emigration had occurred during the decade, and (2) the resident foreign-born population at the end of the decade. For example, Mulder's (2003) residual estimates of emigration during the 1990s were constructed by surviving immigrants who arrived prior to 1990, as revealed in the 1990 Census, forward to 2000 (i.e., by aging all cohorts ten years and subtracting the estimated numbers of deaths) and then comparing the survived population with pre-1990 arrivals in 2000 based on Census 2000. When the former estimate is larger than the latter, the difference is attributed to emigration. The number of emigrants among those arriving between 1990 and 2000 is estimated by applying emigration rates derived from the analysis of earlier arrivals. In other words, emigration rates for recent arrivals are not derived from the data but instead are "borrowed" from earlier arrivals.

Previous emigration estimates based on the residual method as well as other methods have been reviewed in detail elsewhere (Kraly 1998; Woodrow-Lafield 1998; Mulder 2003). To summarize the results of previous work, we compare the estimates of various studies in Table 1 showing figures on the annual number of emigrants together with emigration rates and the rate at which immigration is offset by emigration as would be implied by the various estimates (Table 1). Even though the residual estimates show increases in the annual number of emigrants over time—from 114,000 to 195,000 and then to 225,000, the associated average annual *rates* of emigration appear to have declined from 1.18 percent during the 1960s to 1.15 in the 1980s and then to 0.88 in the 1990s.[1] Similarly, the ratio of emigrants during the decade to immigrants during the period appears to have declined.

The remaining estimates in Table 1 show less consistency across studies, especially for the recently-arrived immigrants and for those from Mexico. Mulder's (2003) estimate of 21,000 emigrants per year among 1990s arrivals implies a much lower emigration rate than prior studies (e.g., Jasso and Rosenzweig 1990). Emigration rates for Mexican immigrants show even more variability, although all of the residual estimates show much lower levels of emigration than the estimates made by Massey and his colleagues (Massey and Singer 1995; Massey et al. 2002). These much higher measures of emigration are derived through the analysis of life histories documenting the number of trips to the United States by Mexican migrants who have returned to Mexico. The detail in the data permit the identification of each separate trip as contributing in- and out-migration, and thus may reflect gross exits more than net emigrants over time. The discrepancy between the residual estimates and those made by Massey demonstrates the importance of distinguishing between net outflows (which is measured by the residual estimates) and gross flows (measured by the estimates of the number of trips out of the United States in Massey's work). We discuss this issue in further detail below.

LIMITATIONS OF THE RESIDUAL APPROACH

A major weakness of the residual method is its inability to estimate emigration for recently-arrived immigrants—i.e., those who arrived between the two censuses. For this group, the earlier census is not available, so emigration rates are not calculated from the data. Rather, in most cases, immigrants who arrived in the decade before the second census are assigned

---

[1] Rates are calculated as emigrants divided by the average population exposed to the risk of emigrating, or the

emigration rates that were calculated for longer-term immigrants. An alternative approach would be to compare the recent cohorts in the second census to estimates of survivors of *legal* immigrants who arrived between the two censuses as recorded in INS immigrant admissions statistics. This is not a viable solution, however, because the estimates of new arrivals do not include unauthorized migrations (e.g., Passel et al. 2004a). One consequence of the lack of residual measures for the recent arrivals is evident from Table 1—the inconsistent, in some cases, unrealistically low residual-based estimates of emigration for recently-arrived immigrants.

Another major drawback of the residual method is that the estimates are sensitive to differences in census coverage between the two censuses. For example, net undercounts were higher in the 1990 Census than in Census 2000 (Robinson et al. 1993; Hogan 1993; U.S. Census Bureau 2001), so in many cases the "expected" populations in 2000 turned out to be smaller than the enumerated populations in 2000 and thus implied a negative emigration rate (an impossibility). This problem is especially evident for country-of-origin groups that contain large proportions of unauthorized migrants, such as Mexican-origin immigrants. For Mexicans, the "expected" population is significantly lower than the enumerated population for both 1980–1990 and 1990–2000 entrants (Mulder 2003). Ahmed and Robinson (1994) found similar problems in their analysis of the 1980 and 1990 Censuses. In spite of an overall increase in net undercount between 1980 and 1990, they suggest that census coverage of migrants from Mexico and Central America may have improved substantially by the 1990 Census because many former unauthorized migrants had acquired legal status under the provisions of the Immigration Reform and Control Act of 1986 (IRCA). Ahmed and Robinson (1994) handle the problem of differential undercount and negative emigration estimates by computing emigration rates for

---

mid-period foreign-born population.

race/ethnic groups (not country-of-origin groups) while excluding those country groups with negative rates. They then use the race/ethnic-specific rates as proxies to estimate emigration rates for countries that initially had negative rates; they match countries with the races of immigrants, for example, matching Hispanic rates to countries sending high proportions of Hispanic immigrants. Mulder (2003) handles the problem by adjusting the 1990 and 2000 census figures for undercount. This approach results in emigration estimates that are highly sensitive to the coverage estimates on which the adjustments are based.

A third limitation is that residual-based estimates are sensitive to inconsistencies in reporting on or actual changes in social and demographic variables between the two censuses. First, the residual method cannot be used to estimate emigration rates for disaggregations of the foreign born by variables that change over time. One cannot use the residual method, for example, to estimate emigration rates by health status because changes over the decade in the size of health status groups could be due to actual changes in health or to emigration. Second, the accuracy of residual-based emigration estimates depends on the consistent reporting of non-time-varying variables, such as year of birth or year-of-entry, across the two censuses. For example, in the 1990 Census, one-third of immigrants who reported having come to the U.S. between 1985 and 1990 on the year-of-entry question probably lived in the United States prior to 1985 based on their responses to the residence-five-years-ago question (Ellis and Wright 1998). If a significant number of recently-arrived immigrants understated the length of time they have lived in the United States in the earlier census, estimates of recent arrivals would be overstated, as would estimates of the "expected" population ten years later. If reporting on year of entry were more accurate (or biased in a different way) in the later census, it would appear that more recent arrivals emigrated than was in fact the case.

A NEW APPROACH

In this article, we supplement prior residual estimates with emigration estimates based on a method that can be applied to recent arrivals and does not depend on assumptions about differential census coverage or consistent reporting on year-of-entry or other characteristics. We refer to this new approach as the "CPS Matching Method."

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. For addresses sampled in the CPS, interviews are done in four consecutive months (e.g., February through May) in year t and those interviews are assigned month in sample codes 1–4. Interviewers return to those addresses and conduct interviews in the same four months (e.g., February through May) in year t+1. The year t+1 interviews are assigned month-in-sample codes 5–8. With this design, those in month-in-sample 1–4 in year *t* appear in month-in-sample 5–8 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 month-in-sample 1–4 in one year who are successfully interviewed as members of month-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).

METHODOLOGY

Basic Approach

In this section, we describe our basic approach for estimating emigration of the foreign-born. 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:

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

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:

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

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:

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

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[5] 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)[6]. 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:

$$e^f = u^f - m^f - d^f - (u^s - m^s - d^s). \tag{4}$$

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

[5] 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.

[6] 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. bom emigration is so low.

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 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 emigration rates of children, we would almost certainly underestimate emigration rates for foreign-born children. For this reason, we treat children ages 0–14 differently from adults. Assuming that foreign-born children emigrate at the same rate as their parents, we assign children the estimated emigration probability of their parents.

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 is $m^*=m(1 - d)$. 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:

$$e^f = [\ u^f - m^f + m^f d^f - d^f - u^s + m^s - m^s d^s + d^s\ ] / (1 - m^f). \qquad (5)$$

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"[7]). 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.[8]

---

[7] 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.

[8] 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 10th root of the 10-year probability that an immigrant will *not* emigrate. In neither case, however, is annual emigration measured directly.

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,

$$\frac{E}{10} = \overline{E}_y - \overline{R}_y . \qquad (6)$$

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

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:

$$\frac{E/10}{P_t} = \frac{\overline{E}_y}{P_t} - \frac{\overline{R}_y}{P_t} . \qquad (7)$$

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. The predicted probability of non-follow-up for each person can be estimated with logistic regression as shown below, where 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:

$$u_i^f = \frac{\exp(X_i' ß^f)}{1 + \exp(X_i' ß^f)} = F(X_i, ß^f) \qquad \text{(8a—foreign born)}$$

$$u_n^s = \frac{\exp(X_n' ß^s)}{1 + \exp(X_n' ß^s)} = F(X_n, ß^s) \qquad \text{(8b—second generation)}$$

.

[12] The SCHIP CPS March sample was designed to evaluate the State Children's Health Insurance Program (SCHIP). The sample is larger than previous March CPS samples and is representative of children in all 50 states and the District of Colombia.

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

Foreign born:	Second Generation:
$m_i^f = F(X_i, \mu^f)$	$m_n^s = F(X_n, \mu^s)$
$d_i^f = F(X_i, d^f)$	$d_n^s = F(X_n, d^s)$

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 3–5 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 (8a) with second generation coefficients:

$$u_i^s = F(X_i, ß^s),$$

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 therefore is expressed as:

$$e_i^f = u_i^f - m_i^f - d_i^f - \left(u_i^s - m_i^s - d_i^s\right).$$

After making adjustments for internal migration being measured retrospectively and rearranging terms (following equation 5), the probability of emigrating is estimated as:

$$e_i^f = \frac{u_i^f - m_i^f + m_i^f d_i^f - d_i^f - u_i^s + m_i^s - m_i^s d_i^s + d_i^s}{1 - m_i^f} \tag{9}$$

Equation (9) is particularly useful because all components can be estimated with CPS and NHIS data. Averaging (9) 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 and Measures

To estimate foreign-born emigration rates for the late 1990s and early 2000s, we use the 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.

To estimate foreign-born emigration rates for the late 1990s and early 2000s, we use data drawn from the 1996, 1997, 1998, 1999, 2000, 2001, 2002, and 2003 March CPS Supplements. During the 1996–2003 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[12] and a research version of the March 2000 Supplement also used Census 2000-based weights. Where possible, we use the 2000-based weights.

The analytical sample used to estimate emigration includes all foreign-born persons in the 1996 through 2002 CPS March samples who were eligible to be followed up in the following year (N = 43,779 foreign born adults and children). This means that for most years, the sample is restricted to those in month-in-sample 1–4. Recall that children are given the emigration probabilities of their parents so it is not necessary to estimate predicted probabilities of non-follow-up and internal migration for them. Therefore, the samples used for modeling non-follow-up and internal migration are limited to first and second generation adults age 15 and older who were eligible to be followed up in the following year. The non-follow-up sample includes those in the 1996 through 2002 CPS March samples (39,980 foreign born and 26,511 second generation persons age 15+), while the internal migration sample includes those in the 1997 through 2003 CPS March samples (N= 38,615 foreign born and 33,918 second generation persons age 15+). Although we use different years of data for examining internal migration and non-follow-up, both refer to the same time period. Internal migration is a retrospective question and refers to moves made by those in the t+1 data that occurred during the period between year t and t+1. Non-follow-up, on the other hand, is measured prospectively and pertains to behavior of those in the CPS in year t for the period between year t and t+1.

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 natives. 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 1987–1997 NDI files. The NHIS did not include a question on place of birth until 1989, so we use the 1989–1994 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.

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

1996–2002 March CPSs with respondents in the following years' CPSs, 1997–2003. In general, households from rotation groups 1–4 in each year t are matched to rotation groups 5–8 in the following year t+1[14]. Then, matching individuals in these households are identified. The matched and unmatched individuals in year t are used to measure follow-up rates. We use the methodology and STATA code developed by Madrian and Lefgren (1999) for linking cases across CPS files, matching on household identification number and 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.[15] 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 (in equation 8) for biases associated with internal migration being measured with retrospective data.

---

[14] We encountered some problems in matching the expanded CPS samples with the March 2001 SCHIP public-use file and the 2002 file. In addition, the match rates were substantially lower for the 2003-2004 match than in previous years. Accordingly, we do not use the 2003-04 matched data at all, and for the 2001–02 match we used the regular March 2001 CPS Supplement; for the 2002–03 matches we used only rotation groups 2 and 3. Also, for those in the Hispanic oversample in most years, month-in-sample is erroneously reverse-coded (personal communication with Census Bureau). For these cases, we select month-in-sample 5-8 as eligible for follow-up.

[15] 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 two years 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 equation (9), we first estimate 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 Appendix Table 1.

The models for non-follow-up and internal migration are estimated for persons age 15+ separately by sex, Mexican/non-Mexican ethnicity, and generational status (1st and 2nd generation) and include as independent variables whether the person was in the CPS oversample, homeownership status, age, year, school enrollment, and education. Because education was not significant in any of the internal migration models, education was included in the models of non-follow-up only[16]. We estimated a total of 8 models of internal migration and 8 models of non-follow-up. We present in Appendix Table 2 the coefficients of the models for Mexican first and second generation men only.[17] The estimates for all models are available from the authors upon request.

For persons age 15+, 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.

The predicted values of non-follow-up, internal migration, and death are then used in equation 8 to estimate an individual-level predicted probability of emigration for foreign-born

---

[16] 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.

[17] The model fit for the second generation versus the foreign born models was generally very similar. The greatest differences occurred for the non-Mexican internal migration models. For men, the pseudo r-square was .083 and .139 for the foreign born and second generation, respectively, and for women, it was .084 and .171.

ages 15+. Children ages 0 to 14 are next assigned the predicted probability of emigration of their parents on the assumption that most children emigrate with their families. 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.

Standard Errors

For the estimates of return immigration, we calculate standard errors using the methodology provided in the CPS documentation by applying the "b" factors associated with Hispanics (U.S. Census Bureau 2002b). Because our methodology for producing emigration estimates involves so many computational steps, the calculation of standard errors of the final emigration estimates through the direct application of statistical formulas would be very difficult—if not virtually impossible. We instead use a bootstrapping method (the "bootstrap" command in Stata) to approximate standard errors of the emigration estimates[18]. The bootstrapped standard errors do not take into account error in the model coefficients but instead treat the coefficients as fixed.

RESULTS

The CPS Matching Method yields an estimate for the annual foreign-born emigration rate of 3.8% (with a standard error of 0.06% and 95% confidence interval of ±0.12%) (Table 2). For a population of 29,988 thousand foreign-born (as in the March 2000 CPS), this rate translates into roughly 1,136 thousand emigrants per year (±34 thousand). At the same time, we estimate a return immigration rate of 0.87% (±0.12%), translating into 261 thousand (±38 thousand) return

[18] The bootstrap method repeatedly draws random samples with replacement of size N from the original sample of size N and calculates the emigration estimates using the methodology outlined above. In our application, we draw 100 separate samples. The standard deviation of the estimates across the samples is interpreted as the standard error (Stata Corporation 2003).

immigrants annually. Subtracting return immigration from total emigration yields annual net emigration of 2.9 percent or 875 thousand net emigrants per year (±0.18% or ±52 thousand).

Males are nearly twice as likely to emigrate as females—5.3 percent versus 2.3 percent—and are significantly more likely to be return immigrants, but not enough to offset their significantly higher emigration rates. Net male emigration (4.4%) remains about twice as high as net female migration (1.7%), a difference that is statistically significant. Emigration and return immigration tend to be relatively high for younger foreign born persons and generally decline with age, except that the emigration rate for adults ages 35–44 is higher than all other groups except children. Taking emigration and return immigration together, net emigration appears to be highest for children (ages 0-24) and working-age adults (35–44 years), but dips for young adults (ages 25–34) and older adults (ages 45+). Net emigration rates for those ages 35–44 are statistically significantly higher than those ages 15–34 and ages 45+, suggesting a pattern of returning to countries of origin after having worked or completing higher education in the U.S. during young-adult years.

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. Return immigration rates are higher for recent arrivals (0–9 years in the country) than earlier arrivals (10+ years in the country), suggesting that circular migration is more common among recent arrivals.

There is also considerable variability in emigration by country or region of birth. Much of the variation we find appears to be associated with the composition of the foreign-born population by country. Thus, for example, countries that have high proportions of recent migrants and/or high proportions of unauthorized and legal temporary migrants tend to have higher emigration rates than countries where the foreign-born population contains higher proportions of naturalized citizens and refugees. Foreign born from Mexico, India, and Africa appear to have the highest emigration rates and those from Europe, Canada, the Caribbean, China, and the Philippines the lowest. Of the high-emigration groups, the relatively high return immigration rates for Mexicans stand out as they are 50% higher than any other country or region. This distinction reflects the circular migration patterns that are commonly observed for Mexican migrants, especially the unauthorized group. On balance, *net* emigration rates appear lowest for groups from southeast Asia (China, Philippines, and "Other" Asia), Europe, Canada, and Central or South America, and highest for foreign-born persons from Mexico, south Asia (India), and Africa.

How do the CPS Matching estimates of emigration compare with residual-based estimates? We compare our net emigration rates and estimates with the residual estimates prepared by Ahmed and Robinson (1994) and Mulder (2003) in Table 3. Ahmed and Robinson's estimates pertain to the 1980-1990 period and Mulder's estimates pertain to the 1990–2000 decade. To make a fair comparison, we derive net emigration rates from their results and then apply these rates to the 2000 foreign-born population to obtain estimates of emigration for 2000 to compare with our results.

None of our estimates for any group are lower than the residual-based estimates (although for some groups, they are not significantly higher). Nevertheless, we find that our estimates are closer to Mulder's in the case of recent arrivals than earlier arrivals. In the case of earlier arrivals (i.e., foreign-born persons who have lived in the U.S. for 10 or more years), our net emigration rate of 2.0 percent is 0.6 percentage points higher than Mulder's estimate of 1.4 percent, or a difference of 177,000 annual emigrants. However, we estimate many more net emigrants among recent arrivals—4.3 percent versus only 0.3 percent in Mulder's estimates, or a difference of 531,000 annual emigrants. As we discussed in this article, estimating emigration among recent arrivals is particularly problematic for residual-based methodologies; in particular, they tend to be made by "borrowing" rates from the earlier arrivals. The large differences may arise because many recently-arrived foreign born are here on temporary student or worker visas or are unauthorized workers, many of whom have no intention to stay in the U.S. This suggests that borrowing rates from longer-term residents and applying them to recent arrivals may not be appropriate and that the CPS Matching method offers an improved method for assessing emigration among recent arrivals.

Our results also show higher emigration rates for countries or regions that send large numbers of unauthorized, temporary, and recently-arrived foreign born. The CPS Matching estimates are both substantively and statistically significantly higher than Mulder's residual-based estimates for Mexico[19] (a source of unauthorized and circular migrants), and both Ahmed/Robinson's and Mulder's estimates for India (a source of many temporary migrants on H-1B visas[20]) and Africa (a relatively new source of immigration and a source of temporary migrants). In terms of *numbers* of emigrants, the CPS-based results are higher than the residual-based estimates by the largest amounts for Mexicans, Central/South Americans, and those from the Caribbean. Even though the discrepancy in net emigration rates is higher for those from India and Africa, the discrepancy in the number of emigrants is lower because there are fewer Indian and African foreign born in the United States than other nationalities.

[19] The residual methodology failed to produce "direct" estimates for Mexico so that the figures show in Table 3 apply residual-based estimates of rates for Hispanics from other areas to the Mexican-born population.

20 H1-B is a non-immigrant visa that allows a foreign individual in "specialty occupations" (such as those requiring professional or graduate training) to work for a U.S. employer for up to six years.

CONCLUSIONS

We have developed 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 initial estimates using data from the 1996–2003 period. The new estimates tend to be higher that those produced by the residual method, particularly for recent arrivals and Mexicans (in terms of rates and numbers), Central/South American and the Caribbean (in terms of numbers), and Indians, and Africans (in terms of rates). 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.

Further development and assessment 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. But beyond this, estimation of emigration among natives is warranted, and very much needed, in its own right. It is reasonable to conjecture that increases in globalization and transnationalism may be raising U.S. emigration rates, just as such economic and social forces, along with others, seem to be increasing the volume of international migration throughout the world in general (Brown and Bean, 2005). Such changes highlight the substantive and theoretical importance of studying emigration among native-born persons.

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 would further 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 and Karen A. Woodrow-Lafield. 1998. "The Quantification of Migration between Mexico and the United States," pp. 1-90 in *Migration Between Mexico and the United States, Thematic Chapters*, Mexico City and Washington, D.C.: Mexican Ministry of Foreign Affairs and U.S. Commission on Immigration Reform.

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.

Susan K. Brown and Frank D. Bean. 2005. "International Migration." Pp. 347-382 in Dudley L. Poston and Michael Micklin (editors), *Sourcebook on Population* New York: Springer Press.

Duleep, Harriet 1994. "Social Security and the Emigration of Immigrants." *Social Security Bulletin* 57(1): 37–51.

Passel, Jeffrey S., and Barry Edmonston. 1994. "Immigration and Race: Recent Trends in Immigration to the United States." Pp. 31-72 in *Immigration and Ethnicity: The Integration of America's Newest Arrivals*, edited by Barry Edmonston and Jeffrey S. Passel. The Urban Institute Press: Washington, DC.

Ellis, M. and R. Wright. 1998. "When Immigrants are not Immigrants: Counting Arrivals of the Foreign Born Using the U.S. Census." *International Migration Review* 32(1): 127-144.

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).

Hogan, H. 1993. "The 1990 Post-Enumeration Survey: Operations and Results." *Journal of the American Statistical Association* 88 (423): 1,047–1,060.

Jasso, Guillermina and Mark Rosenzweig. 1990. *The New Chosen People: Immigrants in the United States.* Russell Sage Foundation: New York.

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. and Audrey Singer. 1995. "New Estimates of Undocumented Mexican Migration and the Probability of Apprehension." *Demography* 32: 203-213.

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.

Passel, Jeffrey S., Jennifer Van Hook, and Frank D. Bean. 2004 . "Estimates of Legal and Unauthorized Foreign-Born Population for the United States and Selected States, Based on Census 2000." Sabre Systems, Inc.. (http://www.sabresys.com/i_whitepapers.asp).

Robinson, J. Gregory, Bashir Ahmed, Prithis Das Gupta, and Karen A. Woodrow. 1993. "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis," *JASA: Journal of the American Statistical Association* 88(423): 1,061-1,079.

Stata Corporation. 2003. *Stata Base Reference Manual, Volume 4 A-F Release 8.* Stata Corporation: College Station, Texas.

U.S. Census Bureau. 2001. *Report on Adjustment Decision: October 2001*. ESCAP II. Washington, DC. (http://www.census.gov/dmd/www/EscapRep2.html)

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, 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 Jeffrey Passel. 1987. "A Count of the Uncountable." *Demography* 24: 375-393.

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. 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.

Woodrow-Lafield, Karen. 1995. "An Analysis of Net Immigration in Census Coverage Evaluation." *Population Research and Policy Review.* 14(1): 173-204.

Woodrow-Lafield, Karen. 1996. "Emigration from the United States: Multiplicity Survey Evidence." *Population Research and Policy Review.* 15: 171-199.

Table 1
Emigration Estimates in Prior Research

Population and Source of Estimate	Time Period of Estimate	Average Annual Emigrants	Emigration Rate[1] (percent)	Emigrants per 100 Arrivals[2]
All Foreign Born				
Warren and Peck (1980)*	1960-1970	114,000	1.18	34.3
Ahmed and Robinson (1994)*	1980-1990	195,000	1.15	27.8
Mulder (2003)*	1990-2000	225,000	0.88	24.7
Recent Arrivals: In U.S. Less Than 10 Years				
Warren and Peck (1980)*	1960-1970	62,000	4.39	18.6
Borjas and Bratsberg (1996)	1970-1980	89,000	3.20	19.7
Mulder (2003)*	1990-2000	21,000	0.32	2.3
In U.S. 10-20 Years				
Mulder (2003)*	1990-2000	142,000	1.66	13.9
Mexican Foreign Born				
Ahmed and Robinson (1994)*	1980-1990	20,000	0.62	12.1
Mulder (2003)*	1990-2000	27,000	0.40	16.1
Massey and Singer (1995)**	1965-1989	1,306,000	51.53	85.9
Recent Arrivals: In U.S. Less Than 10 Years				
Mulder (2003)*	1990-2000	3,000	0.12	1.7
Massey, et al. (2002)[a]**	1965-1985	629,000	28.75	57.5
Massey, et al. (2002)[b]**	1965-1985	182,000	12.50	25.0
In U.S. 10-20 Years				
Mulder (2003)*	1990-2000	15,000	0.68	5.1

* Residual estimate
** Life histories of all international moves
[a] Unauthorized Mexican migrants
[b] Legal Mexican migrants
[1] Average annual emigrants per 100 person years of exposure
[2] Emigrants during the period per 100 immigrants during the period

Table 2

Foreign-Born Emigration Based on the CPS Matching Method, by Age, Sex, Country of Birth, Legal Status, and Time in U.S., Circa 2000

Group or Characteristic	*Annual Rates* (expressed as percents)			*Numbers of Persons* (in 000s)			
	Emigration	Return Immig.	Net Emigration	Emigrants	Return Imm.	Net Emigrants	N (CPS)
All Foreign-Born	3.8	0.9	2.9	1136	261	875	43,779
	(0.06)	(0.06)	(0.09)	(17)	(19)	(26)	
Sex							
Male	5.3	0.9	4.4	798	132	666	21,419
	(0.09)	(0.09)	(0.13)	(14)	(14)	(19)	
Female	2.3	0.6	1.7	342	91	251	22,360
	(0.06)	(0.08)	(0.10)	(10)	(11)	(15)	
Age Group							
0-14 years	5.8	0.9	4.9	119	19	99	3,799
	(0.23)	(0.26)	(0.35)	(5)	(5)	(7)	
15-24 years	5.0	1.4	3.6	219	61	157	5,992
	(0.21)	(0.22)	(0.31)	(9)	(10)	(13)	
25-34 years	3.9	1.2	2.7	268	83	185	10,013
	(0.12)	(0.16)	(0.20)	(8)	(11)	(13)	
35-44 years	5.4	0.7	4.6	347	48	299	9,318
	(0.15)	(0.13)	(0.20)	(10)	(8)	(13)	
45-64 years	1.8	0.5	1.3	127	37	90	10,202
	(0.06)	(0.10)	(0.11)	(4)	(7)	(8)	
65+ years	1.7	0.5	1.2	54	16	37	4,455
	(0.07)	(0.15)	(0.16)	(2)	(5)	(5)	
Years in United States							
0-4 years	6.5	1.5	5.0	301	70	230	8,467
	(0.15)	(0.18)	(0.24)	(7)	(8)	(11)	
5-9 years	5.0	1.4	3.6	474	129	345	8,712
	(0.14)	(0.19)	(0.24)	(13)	(18)	(22)	
10+ years	2.5	0.5	2.0	398	84	313	26,600
	(0.07)	(0.06)	(0.10)	(11)	(10)	(15)	
Country or Region of Birth							
Mexico	5.5	1.2	4.3	472	103	369	15,009
	(0.17)	(0.14)	(0.22)	(14)	(12)	(19)	
Central & S.Amer.	3.1	0.8	2.4	139	33	105	6,818
	(0.12)	(0.16)	(0.20)	(5)	(7)	(9)	
Caribbean	2.5	0.7	1.8	122	36	86	7,699
	(0.12)	(0.15)	(0.19)	(6)	(7)	(9)	
Canada (N.Amer.)	2.7	0.7	2.0	19	5	14	1,136
	(0.23)	(0.34)	(0.41)	(2)	(2)	(3)	
Europe	2.4	0.5	1.8	86	19	67	4,793
	(0.09)	(0.14)	(0.17)	(3)	(5)	(6)	
China	2.9	0.5	2.4	45	8	37	1,672
	(0.19)	(0.23)	(0.29)	(3)	(4)	(5)	
Philippines	2.6	0.8	1.8	36	11	24	1,929
	(0.20)	(0.29)	(0.35)	(3)	(4)	(5)	
India	5.3	0.8	4.5	59	8	51	1,247
	(0.27)	(0.28)	(0.39)	(3)	(3)	(4)	
Other Asia	3.3	0.7	2.6	103	22	81	3,773
	(0.24)	(0.31)	(0.39)	(8)	(10)	(12)	
Africa	5.1	0.8	4.3	36	6	31	897
	(0.13)	(0.19)	(0.23)	(1)	(1)	(2)	

Notes: Rates are based on analysis of pairs of matched CPs files from 1998-99 through 2003-04. Estimates of emigrants and return migrants derived by applying rates to March 2000 CPS population based on Census 2000 weights. Standard errors are shown in parentheses (see text for details).

Table 3

Alternative Emigration Estimates Based on the CPS-Matching Method and the Residual Method

	Net Emigration Rate (expressed as percent)					*Annual Net Emigration* (in 000s)				
		Residual Method		*Difference (CPS-Residual)*			*Residual Method*		*Difference (CPS-Residual)*	
Group or Characteristic	CPS Matching Method	Ahmed & Robinson (1994)	Mulder (2003)	Ahmed & Robinson (1994)	Mulder (2003)	CPS Matching Method	Ahmed & Robinson (1994)	Mulder (2003)	Ahmed & Robinson (1994)	Mulder (2003)
All Foreign-Born	2.9 (0.09)	1.2	0.9	1.8	2.0	1136 (17)	346	265	790	870
Years in United States										
Less than 10 years	4.3 (0.24)	---	0.3	---	4.0	576 (10)	---	45	---	531
10+ years	2.0 (0.10)	---	1.4	---	0.6	398 (11)	---	220	---	177
Country or Region of Birth										
Mexico	4.3 (0.22)	0.6	0.4	3.7	3.9	472 (14)	53	35	420	437
Central & S.Amer.	2.4 (0.20)	0.9	1.0	1.5	1.4	139 (5)	41	43	98	96
Caribbean	1.8 (0.19)	1.2	1.0	0.6	0.8	122 (6)	57	47	65	75
Canada (N.Amer.)	2.0 (0.41)	1.4	1.2	0.6	0.9	19 (2)	10	8	9	11
Europe	1.8 (0.17)	1.3	1.0	0.5	0.8	86 (3)	49	38	37	48
China	2.4 (0.29)	1.8	0.9	0.6	1.5	45 (3)	28	14	17	31
Philippines	1.8 (0.35)	1.6	0.6	0.2	1.1	36 (3)	21	9	14	27
India	4.5 (0.39)	1.6	0.4	2.9	4.1	59 (3)	18	4	41	55
Other Asia	2.6 (0.39)	0.7	1.8	1.9	0.8	103 (8)	23	57	80	46
Africa	4.3 (0.23)	2.4	1.1	1.9	3.2	36 (1)	17	8	19	28

Sources: CPS Matching Method -- Table 2; Residual Method -- Ahmed and Robinson (1994); Mulder (2003).
Standard errors are shown in parentheses (see text for details). Mulder (2003) and Ahmed and Robinson (1994) did not report standard errors.

Appendix Table 1

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

	Immigrants	Natives
Intercept	-2.017 ***	-1.711 ***
Male	0.517 ***	0.531 ***
Race/Ethnicity		
Mexican	0.336 ***	0.044
Other Hispanic	0.067	0.204 ***
NH White	0.230 ***	0.205 ***
Black	0.176	0.297 ***
(Other)		
Age		
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	-0.994 ***	-0.819 ***
(75+)		
General Health Status		
Excellent	-1.181 ***	-1.711 ***
Very Good	-1.135 ***	-1.501 ***
Good	-0.937 ***	-1.127 ***
Fair	-0.586 ***	-0.653 ***
(Poor)		
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.

Appendix Table 2

Logistic Regression Models of Internal Migration and Non-Follow-Up Among First and Second Generation Mexican Origin Males

	Internal Migration		Non-Follow-up	
	1st Generation	2nd Generation	1st Generation	2nd Generation
CPS Oversample	-0.859 ***	-0.835 ***	0.025	-0.094
Homeowner	-0.815 ***	-1.437 ***	-1.176 ***	-1.179 ***
Age				
15-24	1.195 ***	0.684 *	1.258 ***	0.426 *
25-34	0.744 *	0.543	0.528 **	0.320
35-44	0.482	0.854 *	0.246	-0.165
45-54	0.067	-0.326	-0.038	-0.241
CPS Year				
2001	-0.551 ***	-0.733 **	---	---
2002	-0.493 ***	-0.390	0.106	0.225
2003	-0.164	-0.210	---	---
School Enrollment				
High School	-0.401	-0.543 *	-0.942 ***	-0.252
College	-0.922 *	-0.432	-0.835 **	-0.171
Education Attainment				
Less than High School	---	---	-0.093	-0.439
High School	---	---	-0.091	-0.194
Some College	---	---	-0.122	-0.193
Intercept	-1.338 ***	-0.925 **	-0.354	0.091
N	7,164	6,298	7,038	2,795
-2LL Chi-Square	379	398	660	194
Pseudo R-square	0.087	0.107	0.096	0.073

Source: 1998-2004 March Current Population Survey (see text for description of sample)
*** p<.001 ** p<.01 *p<.05
Reference categories are noted in parentheses.