

Two Sources of Error in Data on U.S. Migration in Mexican Household-Based Surveys

ABSTRACT

We examine the nature and degree of two sources of error in data on U.S. migration in Mexican household-based surveys: 1) sampling error that results when whole households migrate and no one is left behind to report their migration; and, 2) reporting errors that result when migrants are not identified by survey respondents. Using data from the first two waves of the Mexican Family Life Survey, which tracked U.S. migrants from 2002 to 2005, our analysis suggests that up to half of migrants are not counted as a result of these two sources of error. Misreporting is the larger source of error, accounting for more than one-third of all migrants. Those who are not counted, especially whole-household migrants, are a unique group of migrants. Their omission results in an underestimate of female migrants, children, and migrants from the Mexican border region, and an overestimate of migrants from the periphery region.

Keywords: Mexico-U.S. migration, sampling error, reporting error, survey data

Two Sources of Error in Data on U.S. Migration in Mexican Household-Based Surveys

In the past three decades, key demographic data sources in Mexico have incorporated household-based reports of migration to the United States, including the decennial census, the intercensal population count, the national demographic survey, and the national employment survey. In these household-based surveys, migration information is collected via household members left behind: survey respondents report the recent migration of household members, including those who have returned and those who are still abroad.¹ These data have been used to estimate and describe migration between Mexico and the U.S. (Bean et al. 1998; Durand et al. 2001; Marcelli and Cornelius 2001; Masferrer and Roberts 2012; Mendoza-Cota 2012; Passel et al. 2012; Passel and Cohn 2009; Rendall et al. 2011; Riosmena and Massey 2012); to assess the educational selectivity of Mexican migrants (Fernández-Huertas Moraga 2011; Ibarra and Lubotsky 2007); and to analyze the causes (Lindstrom and Lauster 2001; Nawrotzki et al. 2013; Villarreal and Blanchard 2013) and consequences (Frank and Hummer 2002; Hildebrandt and McKenzie 2005; McKenzie and Rapoport 2007; Nobles 2013) of migration between Mexico and the U.S.

In this research note we focus on two sources of error in data on migration collected in Mexican household-based surveys. The first is a form of sampling error that arises when whole households migrate and no one is left behind to report their departure. The second is reporting error in which migrants are not reported by the household

¹ Prior to the incorporation of household-based reports of migration, demographic surveys in Mexico included questions on place of birth and place of residence five years prior, which excludes migrants departing and returning within the previous five years and migrants abroad at the time of the survey.

respondent. It is unclear how large these errors are and what sort of bias they present. Several studies have compared Mexican household-based survey data to U.S. data sources, but they are unable to determine whether differences in estimates arise from errors in the Mexican survey data or in the U.S. data source. These studies suggest that the Mexican household-based survey data undercount migrants, especially women and highly educated migrants (Durand et al. 2001; Hill and Wong 2005; Ibarra and Lubotsky 2007).

We bring new data to bear on this issue. The Mexican Family Life Survey (MxFLS) is a national sample of Mexican households interviewed in 2002 and 2005. Crucial for our purposes, the MxFLS made major efforts to track all migrating individuals between the two waves, allowing for the identification of households that migrate as units. Additionally, the MxFLS tracking method allows us to identify U.S. migrants who were not initially reported by the origin household respondent.

Data

The MxFLS is an ongoing, longitudinal survey of 35,677 individuals in 8,440 households in Mexico originally surveyed in 2002 and re-interviewed in 2005 (Rubalcava and Teruel 2007).² The MxFLS was highly successful at tracking households across waves, re-interviewing 90% of the original sample in Wave 2, including 91% of U.S. migrants. The vast majority of those lost to follow up were domestic migrants, whom

² The main data files for the third wave of the MxFLS, collected between 2009-2011, were recently released, but data on U.S. migrants in the third wave had not yet been released at the time of this writing. The data are publically available at <http://www.ennvih-mxfls.org>.

were not tracked if they moved outside of the original MxFLS localities (Velasquez et al. 2010).

Migrants are individuals from the 2002 sample who had been in the U.S. for at least one year, or who planned to remain in the U.S. for at least one year, in 2005 (Teruel et al. 2012).³ In Wave 2, pre-printed lists of sociodemographic information of all members of the Wave 1 household were used to identify the original participants. When entire households were no longer at the original address, fieldworkers inquired with neighbors and/or used a re-contact directory developed in Wave 1 to determine the household's new location. For individuals who were absent from original households, remaining household members were asked to identify their new location. When respondents in origin households did not provide information about the migrant's location, the re-contact directory was used and, in some cases, fieldworkers made follow-up visits to the origin household, offering in-kind and monetary incentives. Through this process, the MxFLS identified 854 individuals (2.4% of the original sample) who were U.S. migrants in 2005.

To assess the nature and degree of error presented by whole-household migration and misreporting, we counted the number and compared the characteristics of individuals who migrated to the U.S. with their whole households ("whole-household" migrants) to two categories of "split-household" migrants: those who were not reported or were inaccurately reported by remaining household members ("misreported" migrants) and those who were accurately reported ("reported" migrants). Whole-household migrants are those belonging to households in which all Wave 1 household members were identified

³ We do not analyze individuals who migrated to the U.S. and returned to Mexico between the waves as these individuals would not be subject to the types of errors analyzed here.

as U.S. migrants in Wave 2. Misreported migrants are U.S. migrants whose location was missing, unknown, or incorrectly reported as in Mexico in the Wave 2 household roster, i.e., whose status as U.S. migrants was determined through the tracking procedure.⁴ The logic is that misreported migrants would be uncounted in a typical household-based survey that relies exclusively on the household respondent's information and willingness to report that information. Insofar as the MxFLS Wave 2 respondents' information and willingness to report are similar to respondents in other surveys, misreporting in the MxFLS should give some sense of this error in other data sources.

We compared the social and demographic characteristics of individuals and households in each of these categories, using data from Wave 1. We also estimated two logistic regressions of migration among adults in the sample (age >15 at Wave 1), one including all migrants and one limited to reported migrants, in order to assess bias introduced by these errors.⁵

The MxFLS sampling design was implemented by the National Institute of Geography, Statistics and Information (INEGI). We used household weights to produce national estimates and nationally representative distributions.

Results

Table 1

⁴ All Wave 1 household members were listed in the Wave 2 household roster. Household members who were no longer living in the origin household at Wave 2 were reported as absent and their current location was recorded. This information reflects the result of the initial attempt by fieldworkers to identify the location of U.S. migrants in Wave 2. In the case of absent household members whose location was not reported in the initial visit, fieldworkers use the re-contact directory or made follow-up visits to the household to determine their location, but the Wave 2 roster was not updated with this information.

⁵ We limit the regression and some descriptive analysis to adults because key information such as U.S. networks was only collected of adults.

Of 854 U.S. migrants, 153 (17.9%) were whole-household migrants and 305 (35.7%) were misreported (Table 1). When weighted, these sources of error account for 1,070,647 migrants, half of all migrants who were in Mexico in 2002 and in the U.S. in 2005. Among misreported migrants, 38% were missing location information, for 44% the location was unknown, and for 18% the location was incorrect (not shown).

Table 2

Whole-household migrants have a unique sociodemographic profile (Table 2). More than half are female, compared to just over a third of split-household migrants, and half are children under age 15, compared to a quarter of split-household migrants. Adult whole-household migrants are more likely to be married or previously married and far more likely than other migrants to have a spouse or immediate family member in the U.S. The three groups of migrants do not differ significantly in terms of education.

Households that migrated as a unit are smaller than other migrant households, and half are headed by women, compared to a quarter of all migrant households. Households that migrated as a unit are less likely to own property in Mexico. The geographic distribution of households varies: whole households are more likely to migrate from the Historic and Border regions; no whole households migrated from the Periphery region.

Table 3

The results of our regression analysis show that the coefficients for sex, household role, home ownership, and Border region are significantly biased by these two sources of error (Table 3). Female sex and Border region coefficients are significantly more negative in Model 2 as a result of women and Border migrants being under-represented among reported migrants. Children, other relations, and home

ownership coefficients are significantly more positive in Model 2 as a result of these characteristics being over-represented among reported migrants.

Discussion

Key Mexican demographic data sources have incorporated household-based reports of U.S. migration in the past three decades. These data are an improvement over previous survey-based measures of migration, they are considered highly reliable, and they have been used extensively. However, they are affected by two types of error: whole-household migration, which leaves behind no one to report the migration, and misreporting when survey respondents are unwilling or lack the information to accurately report U.S. migrants. In this research note, we estimated the nature and degree of these two sources of error using the first and second waves of the MxFLS, a national sample of households in Mexico that were tracked over time.

Our analysis suggests that the number of U.S. migrants not counted as a result of whole-household migration and misreporting is large, totaling as many as half of all migrants who were in Mexico in 2002 and in the U.S. in 2005. Misreporting is the larger source of error, accounting for more than one-third of all migrants.

Migrants who are not counted, particularly whole-household migrants, are not a random sample of migrants. Their sociodemographic profile suggests a unique pathway to migration: these are mostly small families headed by women who are following a spouse or other immediate family member to the U.S. Their migration is facilitated by few economic ties to Mexico, such as land or home ownership, but greater savings. As a result of the omission or miscoding of whole-household and misreported migrants,

regression analysis of U.S. migration is biased, particularly the estimates for sex, household role, home ownership, and region of origin. However, on a whole this bias is not large, and it would not result in substantively incorrect conclusions about the determinants or educational selectivity of U.S. migration using data affected by these sources of error.

These results are mostly consistent with studies comparing Mexican household-based surveys to U.S. data sources, which have suggested that Mexican data sources undercount migrants, particularly women and highly educated migrants. Our results confirm the undercount of women in Mexican household-based surveys; indeed, the MxFLS estimate that 36.7% of migrants are female is more similar to the sex distribution of Mexican immigrants in the U.S. Census (40%) than in the Mexican Census (24%; Ibarra and Lubotsky 2007: Table 5.1). However, we do not find that highly educated migrants are over-represented among the uncounted, suggesting that differences in educational profiles of Mexican immigrants in U.S. and Mexican data sources may arise from the underrepresentation of less educated migrants in U.S. sources.

Two differences between the MxFLS and other Mexican household-based surveys deserve comment. First, because the MxFLS follows a panel of households, migrant misreporting is of specific people who were listed in the household roster in Wave 1 and absent from the household in Wave 2. Respondents' willingness to report the migrant status of their household members may depend on their sense of the risk involved, which may be greater when surveyors are asking about a specific person. Moreover, the MxFLS intent to track these individuals in the U.S. may have increased respondents' sense of risk. These issues would result in an overestimate of misreporting in the MxFLS, as

compared to other surveys, most of which do not ask about the current location of specific people. On the other hand, it might be easier for a respondent to avoid reporting migrants when not asked about specific individuals.

Second, the time frame over which migration is measured in the MxFLS is three years, whereas in most household-based surveys migration is measured over five years. It is plausible that over a longer period, some respondents would be less certain about the whereabouts of migrants, whereas others would be more certain. Over the longer period of time whole-household migrants may represent a greater proportion of all migrants if whole-household migrants are less likely to return to Mexico than split-household migrants. Our results are also not strictly comparable to retrospective reports, which are additionally affected by unobserved geographic mobility and household reconfiguration in the preceding period.

Despite these concerns, the evidence presented in this paper suggests that researchers using Mexican household-based surveys to study migration should continue to pay attention to how whole-household migration and misreporting may affect their analyses. Understanding the degree and nature of these errors may sharpen estimates and allow researchers to more accurately employ Mexican household-based survey data on migration.

References

- Bean, F. D., Corona, R., Tuirán, R., & Woodrow-Lafield, K. A. (1998). The quantification of migration between Mexico and the United States. *Mexico-United States Binational Migration Study* (Vol. 1, pp. 1-89). Mexico City and Washington, D.C.: Mexican Ministry of Foreign Affairs and U.S. Commission on Immigration Reform. <https://www.utexas.edu/lbj/uscir/binpap-v.html>. Accessed 29 July 2014.
- Durand, J., Massey, D. S., & Zenteno, R. M. (2001). Mexican immigration to the United States: Continuities and changes. *Latin American Research Review*, 36 (1), 107-127.
- Fernández-Huertas Moraga, J. (2011). New evidence on emigrant selection. *The Review of Economics and Statistics*, 93 (1), 72-96.
- Frank, R., & Hummer, R. A. (2002). The other side of the paradox: The risk of low birth weight among infants of migrant and nonmigrant households within Mexico. *International Migration Review*, 36 (3), 746-765.
- Hildebrandt, N., & McKenzie, D. J. (2005). *The effects of migration on child health in Mexico*. Policy Research Working Paper Series. World Bank Policy Research. <http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-3573>. Accessed 29 July 2014.
- Hill, K., & Wong, R. (2005). Mexico–US migration: Views from both sides of the border. *Population and Development Review*, 31 (1), 1-18.
- Ibarraran, P., & Lubotsky, D. (2007). Mexican immigration and self-selection: New evidence from the 2000 Mexican census. In G. J. Borjas (Ed.), *Mexican immigration to the United States* (pp. 159-192). National Bureau of Economic Research. <http://www.nber.org/books/borj06-1>. Accessed 29 July 2014.
- Lindstrom, D. P., & Lauster, N. (2001). Local economic opportunity and the competing risks of internal and U.S. migration in Zacatecas, Mexico. *International Migration Review*, 35 (4), 1232-1256.
- Marcelli, E. A., & Cornelius, W. A. (2001). The changing profile of Mexican migrants to the United States: new evidence from California and Mexico. *Latin American Research Review*, 36 (3), 105-131.
- Masferrer, C., & Roberts, B. R. (2012). Going back home? Changing demography and geography of Mexican return migration. *Population Research and Policy Review*, 31 (4), 465-496.

- McKenzie, D., & Rapoport, H. (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics*, 84 (1), 1-24.
- Mendoza-Cota, J. E. (2012). Características y determinantes de los cambios recientes de los flujos migratorios de trabajadores mexicanos hacia Estados Unidos. *Papeles de Población*, 18 (71), 1-36.
- Nawrotzki, R. J., Riosmena, F., & Hunter, L. M. (2013). Do rainfall deficits predict U.S.-bound migration from rural Mexico? Evidence from the Mexican census. *Population Research and Policy Review*, 32 (1), 129-158.
- Nobles, J. (2013). Migration and father absence: Shifting family structure in Mexico. *Demography*, 50(4), 1303-1314.
- Passel, J., Cohn, D., & Gonzalez-Barrera, A. (2012). Net migration from Mexico falls to zero—and perhaps less (pp. 1-65). Washington, D.C.: Pew Research Center. <http://www.pewhispanic.org/2012/04/23/net-migration-from-mexico-falls-to-zero-and-perhaps-less/>. Accessed 29 July 2014.
- Passel, J. S., & Cohn, D. (2009). Mexican immigrants: How many come? How many leave? (pp. 1-21). Washington, D.C.: Pew Research Center. <http://www.pewhispanic.org/2009/07/22/mexican-immigrants-how-many-come-how-many-leave/>. Accessed 29 July 2014.
- Rendall, M. S., Brownell, P., & Kups, S. (2011). Declining return migration from the United States to Mexico in the late-2000s recession: A research note. *Demography*, 48(3), 1049-1058.
- Riosmena, F., & Massey, D. S. (2012). Pathways to El Norte: Origins, destinations, and characteristics of Mexican migrants to the United States. *International Migration Review*, 46(1), 3-36.
- Rubalcava, L., & Teruel, G. (2007). *User's guide: Mexican Family Life Survey 2005*. <http://www.ennvih-mxfls.org/>. Accessed 28 July 2014.
- Teruel, G., Rubalcava, L., & Arenas, E. (2012). Migration in the Mexican Family Life Survey. In A. Cuecuecha & C. Pederzini (Eds.), *Migration and remittances from Mexico: Trends, impacts, and new challenges*: Landham, MD: Lexington Books.
- Velasquez, A., Genoni, M. E., Rubalcava, L., Teruel, G., & Thomas, D. (2010). *Attrition in longitudinal surveys: Evidence from the Mexican Family Life Survey* (pp. 1-30). http://mitsloan.mit.edu/neudc/papers/paper_322.pdf. Accessed 28 July 2014.
- Villarreal, A., & Blanchard, S. (2013). How job characteristics affect international migration: The role of informality in Mexico. *Demography*, 50 (2), 751-775.

Table 1. Estimates of whole-household migration and reporting error in the MxFLS

	Count	Percent of migrants/ migrant households	Weighted count	Weighted percent of migrants/ migrant households
<i>Individuals</i>				
Migrants	854	100	2,114,613	100
Whole-household migrants	153	17.9	306,072	14.5
Misreported migrants	305	35.7	764,575	36.2
Reported migrants	395	46.3	1,043,432	49.3
<i>Households</i>				
Migrant households	510	100	1,287,372	100
Whole-household migrant households	42	8.2	88,436	6.9
Households with misreported migrants	210	41.2	542,069	42.1
Households with reported migrants	258	50.6	656,867	51.0

Source: Mexican Family Life Survey

Table 2. Characteristics of individuals and households by migration status^a

	Whole- household migrants	Misreported migrants	Reported migrants	All migrants
<i>Migrants all ages</i>				
Sex				
Female	56.2***	37	30.8	36.7
Male	43.8	63	69.2	63.3
Age				
0-14	50.2***	25.8	19.6	26.3
15-24	9.4***	48.2†	56.5	46.7
25-34	17.2	14.1	13.7	14.3
35-44	8.9	9.3	5.7	7.5
45+	14.4*	2.7	4.4	5.2
Household role				
Head/spouse	41.6***	19.9	14.5	20.4
Child of head	52†	63	63.2	61.5
Other	6.4***	17.2	22.3	18.2
Sample	153	305	396	854
<i>Migrants age 15 and older</i>				
Marital status				
Never married	20.5***	61.8	66.7	60.4
Currently married	72.6***	36.5	31.4	37.3
Previously married	6.9	1.7	1.8	2.3
Education				
None	7.0	2.0	3.9	3.5
Primary	40.6	32.7	35.8	35.2
Secondary	28.6	43.6	40.1	40.2
High school	14.3	17.2	14.8	15.7
College+	9.5	3.8	5.2	5.1
Missing	0	0.8	0.1	0.3
Employed	46.3†	63.9	59.5	
U.S. networks				
Spouse in U.S.	21.6**	2.2	3.9	
Immediate family in U.S.	50.3*	31.6	31.3	
Extended family in U.S.	11.7	22.0	19.5	
Sample	80	225	313	618

Table 2 continued on next page

Table 2 cont.

	Whole- household migrants	Misreported migrants	Reported migrants	All migrants
<i>Households</i>				
Size (mean)	3.6***	5.9	5.8	5.6
Female headed	50.1*	18.3	26.0	24.4
Household resources				
Owens home	64.0*	72.4**	85.6	78.6
Has savings	27.5	16.4	13.8	15.8
Owens business	7.7	18.6	16.4	16.8
Owens land	9.5***	32.1	31.3	30.1
Size of locality				
<2,500	45.4	39.6	43.7	42.1
2,500-14,999	17.6	26.4	25.3	25.2
15,000-99,999	12.7	5.5	9.6	8.1
>100,000	24.3	28.5	21.4	24.6
Region^b				
Historic	58.6†	41.9	41.0	42.6
Border	14.4†	10.6**	3.4	7.2
Center	27.0	37.8	41.4	38.9
Periphery	0***	9.7	14.2	11.3
Sample size	42	210	258	510

Source: Mexican Family Life Survey

a. All distributions are weighted. All characteristics were measured in 2002, up to three years prior to the migration.

b. Region is defined by Durand et al. (2001). The Historic region includes Aguascalientes, Colima, Durango, Guanajuato, Jalisco, Michoacán, Nayarit, San Luis Potosí, and Zacatecas; the Border region includes Baja California, Baja California Sur, Coahuila, Chihuahua, Nuevo León, Sinaloa, Sonora, and Tamaulipas; the Center region includes the Federal District, Guerrero, Hidalgo, México, Morelos, Oaxaca, Puebla, Querétaro, and Tlaxcala; and the Periphery region includes Campeche, Chiapas, Quintana Roo, Tabasco, Veracruz, and Yucatán.

†p < 0.1; *p < 0.05; **p < 0.01; ***p < 0.001 based on t-tests of equal means, comparing whole household migrants and misreported migrants each to reported migrants.

Table 3. Logistic regression coefficients of migration among MxFLS adults, comparing migration measured with all migrants to reported migrants^a

	Model 1 <i>All migrants</i>	Model 2 <i>Reported migrants</i>	$X_{m1} = X_{m2}$ ^b
Sex (Male)			
Female	-0.81***	-1.09***	*
Age (15-24)			
25-34	-0.70***	-0.77**	
35-44	-1.05***	-1.26***	
45+	-1.85***	-2.07***	
Marital status (Never married)			
Currently married	0.30	0.36	
Previously married	0.23	0.39	
Household role (Head/spouse)			
Child of head	0.64**	1.12***	*
Other	0.79**	1.42***	**
Education (None)			
Primary	0.03	0.37	
Secondary	0.54†	0.62	
High school	0.59*	0.54	
College+	0.34	0.21	
Employed			
	0.24*	0.25	
Household resources			
Owens home	0.15	0.56*	*
Has savings	-0.16	-0.22	
Owens agricultural land	0.03	-0.10	
Owens non-ag business	-0.13	-0.25	
U.S. networks			
Spouse in U.S.	2.04***	1.86***	
Immediate family in U.S.	1.19***	1.07***	
Extended family in U.S.	0.62***	0.52*	
Missing	0.52**	0.25	†
Region (Historic)			
Border	-1.16***	-1.94***	*
Center	-0.46**	-0.35†	
Periphery	-0.79**	-0.59†	
Size of locality (<2,500)			
2,500-14,999	-0.07	-0.01	
15,000-99,999	-0.49*	-0.25	
>100,000	-1.03***	-1.06***	
Constant			
	-3.68***	-4.88***	
Sample size	23,803	23,543	

Source: Mexican Family Life Survey

a. In Model 1, the dependent variable =1 for all migrants (n=618); in Model 2, the dependent variable =0 for non-migrants and for U.S. migrants inaccurately reported as domestic migrants (n=45) and =1 for reported migrants (n=313). Whole-household migrants (n=80), migrants with missing location information (n=75) and migrants whose location was unknown (n=105) were dropped from the sample in Model 2. Estimates are weighted, and standard errors are adjusted for clustering at the household level. Missing data was mean imputed or dummy substituted.

b. Results from a cross-model Wald test of equal coefficients.

†p < 0.1; *p < 0.05; **p < 0.01; ***p < 0.001