Estimation of the Number of Immigrants and Refugees in Lansing (Michigan), 2005–2007

A report prepared for the Family and Community Vitality Focus Area of the Innovation in University-Community Research Collaboration Grants by:

Steven J. Gold, Ph.D.
Wilma Novales Wibert, Ph.D.
Vera Bondartsova, MA
John Melcher, MS
Lori A. Post, Ph.D.
Brian J. Biroscak, MS, MA

DECEMBER 2008

MICHIGAN STATE
U N I V E R S I T Y

Executive Summary

The debate over immigration is one of the most politically charged policy issues in the United States (US). Given the charged nature of this topic, it is vitally important to have reliable data on not only the number of US foreign nationals but also the characteristics of this extremely heterogeneous group—a population comprised not only of immigrants but also refugees.

There exist a small number of data sources for informing policy and practice at the national level. However, such data are often lacking for smaller geographic areas. The present report describes a recent effort to generate serviceable data on the immigrant and refugee population for a medium-sized metropolitan area: Lansing, Michigan.

The authors requested funds from the Family and Community Vitality Focus Area of the Innovation in University-Community Research Collaboration Grants (FACT) Program to support a project intended to acquire the best possible estimates of refugee and immigrant populations in the Lansing area. The goals of this research were two-fold. First, our goal was to provide the IRRC—and Lansing area stakeholders in general—with information to assist them with resettling and obtaining funding for immigrants and refugees. The second goal was to develop better techniques for tabulating diverse refugees and immigrants in a medium-sized community. By comparing and contrasting three data sources—i.e., refugee services, public schools, and a local health plan—we were able to generate estimates of Lansing's refugee and immigrant population.

During the period 2005–2007, we estimate the total number of immigrants and refugees in Lansing to be somewhere between 10,938 and 13,282. Although perhaps a bit on the high end due to methodological assumptions these estimates seem plausible, based on previously cited figures. While such estimates are valuable, a number of shortcomings related to the data prevented us from painting a more complete picture of these populations. We conclude the report with a number of recommendations for future research that will assist Lansing area stakeholders in determining relevant policy and practice.

Table of Contents

Introduction

"The United States, with its large and changing immigrant population, has an enormous stake in international migration. Equally important is the impact that US immigration policies have on the rest of the world. With the interests of so many people hanging in the balance, it is crucial that the US immigration policy debate be guided by reliable data and analysis" (2008b).

As stated above, reliable data should form the basis of policy development and practice, not only with respect to immigration, but in regards to any social issue. During 2006, nearly 1.3 million people were granted lawful permanent resident status in the US (Migration Information Source, 2008c). US law currently provides for three general immigrant categories: family reunification, employment sponsorship, and humanitarian cases (i.e., refugee and asylum adjustments). Family reunification accounted for more than 63 percent of all lawful permanent immigration in 2006; employment-preference immigrants accounted for nearly 13 percent of all lawful permanent immigration; and about 17 percent were status adjusters who entered as refugees or asylees. Mexico, China, the Philippines, India, and Cuba were the top five countries of birth of lawful permanent residents in 2006 (Migration Information Source, 2008c).

Standardized, reliable data (such as the above) can help inform US immigration policy and assist with planning for service provision. In the case of Latin American immigrant groups, for example, Zarrugh (2008) documented differences in US immigration policy and law that impact various groups differently, noting that "a 'one-size-fits-all' approach to the provision of services may be less effective" (p. 49) than a tailored approach. At the national level, one of the primary data sources for informing migration policy and debate is the *Yearbook of Immigration Statistics*: a compendium of tables that provide data on a variety of classes of foreign nationals (US Department of Homeland Security, 2008). US policy makers also have access to *Annual Flow Reports* and *Annual Reports* released by the Office of Immigration Statistics, which provides data on

legal permanent residents, refugees and asylees, nonimmigrant admissions, naturalizations, and enforcement actions.

However, immigrants do not take up *general* residence in the US. That is, they settle in local communities. Therefore, not only are accurate data needed at the national level, they are essential at the local level as well in order to guide policy and plan for services. While there are a number of federal data sources and analysis mechanisms in place, the Lansing (MI) area has no such mechanism. The present report describes a recent effort to generate serviceable data on the immigrant and refugee population for the Lansing area. We begin by describing the impetus for the present research project.

Background

Although precise figures are lacking, the greater Lansing area has become a significant point of settlement for both immigrants and refugees. For example, the Brookings Institution ranked Lansing-East Lansing, MI among the top 10 medium-sized metropolitan areas for number of refugees resettled during the period 1983–2004 (Singer & Wilson, 2006). In addition, diverse arrays of immigrants have come of their own volition. Regardless of the conditions surrounding their arrival, though, these populations encounter many of the same challenges as they build lives in Lansing. For example, they must find a place to live, learn English, acquire employment, enroll their children in school, obtain health care, and fulfill other basic needs.

Area newcomers have arrived with the sponsorship of several resettlement agencies. For example, St. Vincent Catholic Charities (United States Catholic Council of Bishops) is the official sponsor for all such adults and families in the Lansing area. The Refugee Development Center (Christ Lutheran Church) specializes in resettling unaccompanied minors. A group of forward-looking service providers representing religious and secular non-governmental agencies, in cooperation with local government agencies, educational institutions, and other service providers, recently joined together as the

<u>Immigrant and Refugee Resource Collaborative</u> (IRRC). The IRRC exists to provide assistance to Lansing's newly arrived populations. While the challenges of resettling these migrants are considerable and the financial resources available for the task are limited, nevertheless, members of this collaborative—as well as local community leaders—are committed to providing these new arrivals with a positive experience for humane reasons, as well as to create vital human resources for the City of Lansing.

In the words of Lansing's (former) mayor, "Lansing is very fortunate to have immigrants and refugees. They come in and buy houses, fill jobs, open stores, pay taxes. They keep us going " (Range, 2005b). Indeed, Dunlevy (2006) and Co et al (2004) report that immigrants bring about positive effects on state-level exports to their respective home countries. While various estimates of such populations exist, they are only rough approximations and quickly become outdated. Moreover, while resettlement agencies can provide some information about refugees, no local organization oversees the arrival of immigrants. Hence we lack up-to-date information on this important group.

According to the 2000 Census, Lansing lost more than 8,000 residents during the decade 1990 to 2000.[1] That reduction would have been even more severe without the nearly 5,000 refugees estimated to have resettled in Lansing during the decade. During the first half of the present decade, almost 2,000 more refugees made Lansing their home (Range, 2005b). An estimated total of 13,000 refugees have settled in Lansing since 1975. In fact, the Capital Area took more refugees than any other city in Michigan in 2004 (Range, 2005a). Although rough approximations (such as the aforementioned) exist, we know very little about both their accuracy and the characteristics of these populations.

We do know, however, that refugees and immigrants are vital to the area's economic well being. For example, Peckham Vocational Industries, which makes military clothing and parts for General Motors, employs over 200 refugees from 30 countries, paying most about $6-$7 an hour (Range, 2005b). Another

---

[1] More recent figures indicate this decline has continued with an estimated loss of 4,426 residents between 2000 and 2007. Source: American Fact Finder (http://factfinder.census.gov)

important indicator of the significance of refugees and immigrants is the large number attending local schools. The Lansing School District enrolled 2,282 bilingual students during 2005–'06 (about 14% of all 16,838 students in the district), representing a broad number of linguistic and nationality groups (Bilingual Education Department - Lansing School District, 2005). The largest of these include Spanish speakers (1,114), Hmong (336), Vietnamese (188), Somali (136) and Arabic (126). While immigrant and ESL (English as a second language) students have significant needs, they are also an important source of per capita funding for local school districts and without their presence, Lansing School District enrollment—and hence its financial support—would be diminished. In fact, the recently announced loss of more than $3 million in state funding for Lansing schools (Prater, 2008) would have been even greater without the enrollment of immigrants and refugees.

Importance and Relevance to the Community

The accurate enumeration and description of Lansing's immigrant and refugee populations is vital for their effective resettlement, as well as to access the funds required to accomplish this task. Population estimates are especially crucial for refugees in Lansing because funding for many resettlement activities is distributed on a per capita basis. Funding allocations are further limited to very specific populations and social need groups (such as the elderly, the homeless, substance abusers, youth, etc). Accordingly, accurate estimates of population size and social characteristics are essential to determine eligibility for scarce funds.

Yet unique challenges are presented when one attempts to count the populations involved. Official tabulations (such as the US Census) are less than optimal in estimating small, socially marginal populations in local areas and are conducted only every decade (Heer, 1968). As such, they are less than ideal to begin with and quickly become outdated (Anderson & Fienberg, 1999).

A fraction of refugees are formally resettled by agencies in Lansing. Their numbers are known as of their date of arrival. However, immigrants freely join or leave the community. Moreover, many refugees are involved in the process of secondary migration, whereby they leave their place of initial resettlement and move elsewhere (Gold, 1992). The population size is further altered by births and deaths. Given these realities, there currently are no firm estimates of the size and social characteristics of the immigrant and refugee population in Lansing. Hence the need for the research project described herein.

Scope of Report

The problem of enumerating hard-to-count populations has long interested demographers, epidemiologists, and other researchers. A number of methods may be helpful for the type of problem faced here—i.e., the problem of counting the number of immigrants and refugees. In this report, we focus on the application of one such method, termed "capture-recapture" (E.B. Hook & Regal, 1995; International Working Group for Disease Monitoring and Forecasting, 1995a, 1995b). After presenting our population estimates, we will discuss the implications in terms of future enumeration efforts as well as local policy regarding immigrants and refugees.

The present study is limited to an estimation of the Lansing area's immigrant[2] and refugee population, as well as an examination of available population characteristics. The scope of this report is also limited to descriptive analyses—i.e., counting and classifying individuals—rather than employment of inferential statistics—e.g., testing hypotheses. As such, the results presented here should be considered exploratory. Furthermore, while quantitative data analysis is often more efficient, qualitative data collection would allow for a more complete understanding of the population under study (more on this in the Recommendations section).

---

[2] No distinction was made in the present study between legal and illegal immigrants. (Refugees have legal status.) For a discussion of illegal migrant estimation procedures, please see the paper by the *International Centre for Migration Policy Development (Jandl, 2004).*

In review of this report, it is important to remember that there is a lack of high quality data for determining the extent and characteristics of the immigrant and refugee population in Lansing. What this study contributes is to help derive *estimates* and begin to characterize the demographics of these populations. With those words of caution, the following section presents a brief review of the published literature regarding estimation studies of immigrant and refugee populations.

Review of Literature

As stated above, US law provides for three general immigrant categories: family reunification, employment sponsorship, and humanitarian cases (i.e., refugee and asylum adjustments). As defined by the Migration Policy Institute (Migration Information Source, 2008a), migrants admitted for family reunification are individuals admitted because they are the immediate relatives, foreign fiancées, or foreign adopted children of citizens or foreigners already residing in the receiving country. Employment-based settlers are persons selected for long-term settlement because of their qualifications and prospects in the receiving country's labor market. Asylum seekers include anyone who files an application for asylum in a country other than their own. (They remain in the status of asylum-seeker until their application is considered and adjudicated.) A refugee—per the UN Convention Related to the Status of Refugees and the 1967 Protocol—is "any person who, owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country" (Migration Information Source, 2008a).

A review of the literature reveals only a handful of publications where authors have successfully estimated the size of immigrant populations via overlapping and incomplete lists. That is, estimation is required (as opposed to counting) because immigrant populations can be considered a "hidden"

population, meaning some members of this group are not observable (in research speak) because they are not recorded in a standardized, accessible database. Neither a census nor a sample would capture this invisible population. The Department of Natural Resources, responsible for estimating wildlife populations, coined the term capture-recapture because they literally capture their subjects of study (deer and other wildlife), release them, and capture a second sample. Capture-recapture uses the information from the overlapping cases to estimate the invisible cases that do not appear in any of the data sources (Bishop, Fienberg, & Holland, 1975).

Much of the available research involves data derived from the census and other official statistical sources (for example, Bardsley & Storkey, 2000; Newbold, 2002; Skop, 2001). Other notable studies suggest inventive ways to draw on alternative data sources such as public documents like citizenship applications and marriage licenses that record both ethnicity and demographic data (for example, Herman, 1994; Kitano, Yeung, Chai, & Hatanaka, 1989). One of the *most* inventive approaches, however, to estimating the number of migrants in a defined area was employed by Monserrat and Cerda' (2002).

The authors identified economic immigrants via three incomplete and overlapping data sources: police and immigration authority records, labor union rosters, and health-related institution records. Capture-recapture methods were then used to estimate the number of immigrants not documented by any of the overlapping data sources, based on statistical dependencies among the sources. Although the three data sources listed 6,538 economic immigrants during the time period under study, the authors estimated an additional 33,034 individuals were not included in any of the datasets (Monserrat & Cerda', 2002).

As indicated by the aforementioned study, existing literature on immigrant and refugee estimation generally concerns overseas populations or large US populations (of fewer nationalities) in gateway cities such as Los Angeles, Miami, San Francisco, and New York (Haines, 1996; Ong, 2003; Portes & Bach, 1985). In such communities, an organized and specialized bureaucracy—featuring numerous group-specific resettlement agencies, co-ethnic mutual assistance

agencies, and other institutions—is often well equipped for identifying, enumerating, and understanding the population of local refugees and immigrants. The Lansing, Michigan area lacks this infrastructure but would benefit from more timely and accurate information about its newcomers. This led us to propose the following research questions.

Research Questions

The authors requested funds from the Family and Community Vitality Focus Area of the Innovation in University-Community Research Collaboration Grants (FACT) Program to support a project intended to acquire the best possible estimates of refugee and immigrant populations in the Lansing area. The goals of this research were two-fold. First, our goal was to provide the IRRC—and Lansing area stakeholders in general—with information to assist them with resettling and obtaining funding for immigrants and refugees. This research will demonstrate the overlap and dependencies between services of the various agencies. For example, St. Vincent Catholic Charities encourages new arrivals to seek needed health care from the Ingham Health Plan. The results presented here will demonstrate to what extent the newcomers present to the different agencies. The second goal was to develop better techniques for tabulating diverse refugees and immigrants in a medium-sized community. This research has the potential to contribute substantially to the existing literature on immigrants and refugees.

Therefore, the research questions we sought to answer were:

1.) How many refugees and immigrants resided in the Lansing area across the three-year period, 2005–2007?
2.) What combination of data sources provides optimal population estimates?

3.)   What is the nature of refugee and immigrant adaptation and resettlement in a community with a diverse and changing migrant population?

4.)   How can academic researchers and local social service personnel best collaborate to assist immigrant and refugee populations?

Method and Materials

The research project described below was submitted for Institutional Review Board approval at Michigan State University on 05/25/2006. We begin our description of research "Methods and Materials" with a review of the theory behind capture-recapture estimation, followed by a description of each data source employed, along with the procedures required to obtain the data. We then detail the estimation procedures employed before delving into the results of our study.

Research design[3]

The capture-recapture method is derived from the practice of randomly sampling a defined wildlife population, marking the captured sample members for later identification purposes, then releasing those captured back into the population. A second random sample from the same population is taken and the proportion of marked individuals from the first sample found in the second sample is assumed to be the same as the proportion of marked sample members in the defined population (Barnes, 1995). Multiplication of the two sample sizes divided by the number found in both samples results in an estimate of the total population of interest (Bishop et al., 1975). While it adjusts for any individual source not containing all cases, it generally assumes that the cases a source does contain are truly cases. According to Hook and Regal (1995), capture-recapture has been applied to human population research for four major applications: (1) to estimate population size when there is clearly incomplete data available from two or more sources; (2) refinement of estimations derived from surveys; (3) evaluation of data registries; and (4) deriving plausible upper or lower limits on the total affected population via confidence intervals.

---

[3] Portions of the following text were taken directly from the second author's PhD dissertation on capture-recapture methodology (Wibert, 1998).

Using the example for estimating the number of deer in a designated area, a sample of n_1 deer are captured, tagged, counted and released back into the woods. A second sample of deer, n_2, is subsequently caught. The number of recaptured deer (m) is counted and is also known as the overlapping population. The proportion of tagged deer in the second sample is m/n_2 which is assumed to be the same as the proportion of the sample to all of the deer in the whole forest, n_1/N. Thus, the number of deer in the whole forest, N, is estimated as (Neugebauer & Wittes, 1994):

$$N = (n_1 * n_2) / m$$

For example, if 100 deer are captured at Time 1 and released, and then another sample of 100 deer are captured at Time 2. Five of the deer at Time 2 also had been captured at Time 1, meaning that 5% of the population from Time 1 was recaptured at Time 2. The proportion of recaptures at Time 2 represents the proportion of the Time 1 sample to the whole deer population. Thus, we can summarize the 100 deer captured during Time 2 equal 5% of the entire deer population. Therefore, we can estimate the total deer population to be 2,000:

$$N = (100 * 100) / 5 = 2,000$$

Members of a population are "captured" by appearing in one data source or "list" and are "recaptured" by appearing again in another list, or by "matching." They are "tagged" by virtue of having an identifier where they can be traced from one "list" to the next. The "lists" are matched against each other using the unique identifier and the matches between lists are then counted. Using these numbers and the formula referenced above for a two-source estimation of population (Sekar & Deming, 1949), the total population of immigrants and refugees in a defined area (and during a defined time period) can be estimated.

For applying the capture-recapture estimation method with two data sources, several assumptions are required:

1. Positive identification of true cases and accurate matching
2. No change to the population during the period under investigation (i.e., population stability)
3. Same chance of each individual being included in each source (catchability)
4. Statistical independence of data sources

True cases and accurate matching

The technique of capture-recapture is only recommended in situations where there are known cases of the topic of interest. Although it is absolute that the data source is incomplete, those cases in the database must be true cases. Furthermore, the cases must have identifiers sufficient for positively matching all cases between data sources.

Population stability

In general, there are no populations that remain completely unchanged during the research period. Therefore, population stability is identified as the degree of "instability." Theoretically, there is constant out-migration and in-migration; however, as long as the overall population size (calculated within a given time frame) remains fairly constant, there is equal opportunity for the loss of immigrants and refugees as well as in-migration of new arrivals. The samples used for this study were collected simultaneously, which reduces the risk of violating the population stability assumption. Therefore, even if there was some entry and exit of captured persons, those persons still had the possibility of being captured in more than one sample. In the case of the present research, three-factor capture-recapture is being used and therefore it is possible to install log-linear controls for sample dependencies (Ernest B Hook & Regal, 1992).

Catchability

With capture-recapture methodology, the probability of ascertainment varies by any of the institutional sources. Catchability can be influenced because of geographic or socioeconomic variables (E.B. Hook & Regal, 1995). The data must be modeled (similar to independence log-linear analysis) and tested for deviations from independence. Overall dependency or relative bias estimates can be explained as though they were calculated from sources with no variable catchability (E.B. Hook & Regal, 1995). It will be difficult to control for likely causes of catchability in the present research, however, because of the dearth of covariates in the datasets.

Source dependency

Although the final assumption (independence) must be sufficiently addressed, the research project described here is using three-source capture-recapture as opposed to solely two-source enumeration. To adjust for assumption violations, log-linear modeling may still be usefully employed to generate estimates of the unobserved members of a population. As stated by Nanan and White (1997): "...the assumption of independence can be dropped, and interdependence among data sets can be accounted for by using Log-linear modeling techniques to assess source dependencies" (p. 145). Furthermore, when the number of cases contained on each list is large, the estimates are less sensitive to dependency of samples and catchability (E.B. Hook & Regal, 1995). We continue below with a description of data sources used for both dual-source enumeration and log-linear modeling of the population of immigrants and refugees in Lansing.

Data sources

As described above, determining the number of immigrants and refugees in a particular area is a difficult task. However, the fact that different institutions in Lansing provide services to these populations was considered an opportunity to

estimate the size of these groups. In order to develop reliable estimates of refugees and immigrants in the Lansing area, we sought first to identify as many bases of enumeration as possible and then compare among them. By comparing and contrasting multiple sources of population data (i.e., triangulation), we hoped to develop viable estimates of population size and create key social and demographic indicators of Lansing's refugee and immigrant populations.

To obtain base estimates, we investigated data sources that could provide for enumeration of local refugee and immigrant populations. These included the Lansing School District, census documents, tabulations from refugee sponsorship agencies, and the Michigan Department of Human Services (which provides refugee cash assistance). We defined a "refugee" in a manner similar to Newbold (2002)—i.e., we assumed that all foreign-born arrivals from a refugee producing country were refugees. While this is clearly a reasonable assumption for some countries (e.g., Afghanistan, Sudan, and Myanmar), this is more problematic for nations where the proportion of refugees is smaller. Because estimates might also be derived through data from organizations having extensive interactions with refugees and immigrants in Lansing, we contacted social service agencies, ethnic organizations, religious congregations, ethnic media, and ethnic businesses. Finally, local surveys and data tabulations—as conducted by the Community Indicators Project of The Capital Area Community, the State of the State Survey from MSU, the Institute of Social Research at the University of Michigan, and Patricia Becker's Southeast Michigan Census Council—were evaluated to determine their potential to provide desired data.

For statistical estimation procedures, we ultimately settled on three analytic sources. An analytic source is "a list which may be constructed from one or more original sources" (E.B. Hook & Regal, 1995, p. 244). The original sources—i.e., "a particular list of cases from some institution" (E.B. Hook & Regal, 1995, p. 244)—have been collapsed into the analytic sources. The three sources for which agencies granted us access were: Lansing School District records, Refugee Service records, and Ingham Health Plan medical records. This after the lead investigator explained our project to agency personnel, provided

documentation and proof of IRB approval, and received consent from the agencies' boards (and, in one case, a law firm representing the agency).

Lansing School District

The Lansing School District provided information on all students, along with information on their parents/guardians, who were bilingual and enrolled for classes in year 2005 (n = 2,365), 2006 (n = 2,247) or 2007 (n = 2,246). The Lansing School District is where any child of immigrant or refugee parents would have enrolled, save for those students enrolled in private school (likely to be a very small number). A number of refugee families would likely have been connected with schools via referral from refugee services (see below).

We deleted records for all families in which the student was US born. After deleting duplicate cases, we were left with a sample of 1,184 records. Next, we created a file with pairs of parents/guardians, because the other data sets to be merged were populated with adults as well. We assumed they had the same race, country of birth, and language as their children. Following the removal of repeated pairs (for siblings) and problematic entries, we had a sample of 760 pairs of parents/guardians. The combination of first and second parents/guardians and deletion of missing and repeated cases resulted in 1,267 records. We combined the student files (n = 1,184) with parents (n = 1,267) for a sample size of 2,450 individuals.[4] We subsequently obtained an additional file from the Lansing School District. After completing an identical process as above—i.e., removing US borns and deleting duplicates—we were left with a final sample size of 2,506 individuals.

Ingham Health Plan

Public health services provided the numbers of immigrants and refugees who, between 2004 and 2007, had visited any of the health centers that served the

---

[4] The sample size was 2,450 rather than 2,451 (i.e., 1,184 + 1,267) because one repeated record was later found.

area under the auspices of the Ingham Health Plan (n = 779). While not an insurance program, Ingham Health Plan provides basic medical care to low income uninsured residents of Ingham county (Health Plan Management Services, 2007–2008).

Again, refugees would likely be referred to the health program via refugee services. However, a number of immigrants as well as refugees likely sought care at emergency departments. We deleted records of people living in cities other than Lansing, along with year 2004 records and repeats, which resulted in a final sample of 377 cases.

Refugee Services

The Refugee Development Center (RDC)—working in close cooperation with St. Vincent Catholic Charities—provided information on all refugees registered with the organization between 2005 and 2007. The Center provides a number of resettlement services, including a number of training modules designed to provide newcomers with necessary skills, including finding housing, securing employment, and enrolling children in school (Refugee Development Center, 2008). Therefore, we expected to find a large amount of overlap between the number of persons captured with refugee services data and the other two sources.

The initial data set provided by refugee services indicated 2,247 records of families resettled across the 3-year period, 2005–2007. However, these were records of families not individuals. After parsing out individuals who were helped in multiple years (duplicates) and those resettled into communities other than Lansing, a final sample size of 1,249 individuals remained.

Estimation procedures

For statistical estimation, we cross-referenced the three data sets describing the same population. Each data set contained identifiers in order to match cases. For

example, an individual listed in separate data sources with identical first/last names, addresses, and dates of birth would be considered a match. By merging data sources in such a manner, we were able to de-duplicate the total set of cases (i.e., not count the same person more than once). In order to protect respondents' identities, we subsequently replaced identifiers with an anonymous ID code.

As described above, there are four assumptions associated with the use of capture-recapture technique as described by Nanan and White (1997), Hook and Regal (1995), and Fienberg (1998).

Perfect matching

"Individuals identified in one source are perfectly matched in another without error and with zero mismatches or non-matches." There is a reduced risk of matching error in the present study, as the matching was performed manually rather than electronically. Furthermore, a single investigator read and matched the data. In addition, there were several case 'identifiers' such as first name, last name, and race.

Population closure

"The population under study is closed." The closure of population is more crucial in animal ecology studies where the samples are taken sequentially. In the case of this research, data were being recorded simultaneously in all sources. This is less problematic because even if an immigrant moved out of the area, they still had the potential to be captured in the three sources during their tenure. Finally, closure of population is a form of dependency and can be controlled for as is discussed below.

Independence

"The probability of being in one list is not affected by the probability of being on another." This assumption is violated in the case of refugees and immigrants in Lansing, Michigan. Agencies are collaborating with each other more in order to have a coordinated web of services for new arrivals. Although better coordination is still needed, the Refugee Development Center (RDC) encourages newcomers to look into health services offered by the Ingham Health Plan. Similarly, the RDC provides new arrival families with information on how to enroll their children in local schools.

As defined by Hook and Regal (1995), any two data sources (*A* and *B*) are statistically independent if the overall probability of members of a population who appear in their overlap is equal to the product of the average probabilities of appearing in *A* and in *B*. Sources are "dependent" if they are not "independent" as defined above. Dependencies can be thought of as both positive and negative and both can be present within the same model. In general, positive dependence of sources will tend to produce an underestimate of the true population size, and negative dependence will tend to produce an overestimate (Ernest B Hook & Regal, 1992; E.B. Hook & Regal, 1995).

We anticipated that there should be negative dependence among the data sources—i.e., the probability of being in two sources is less than the product of the probabilities of being in any one of them. This assumption is based on the fact that Lansing's network of resources for refugees and immigrants does not yet have a fully developed referral system. Also, other alternatives are sometimes available—e.g., health care that is easier to obtain than working with the Ingham Health Plan. Thus we expect two-source analyses here will (if anything) tend to overestimate the true number of migrants in the Lansing area. This does not, however, invalidate the findings from two-source analyses, and we are still left with multi-source analyses (log-linear modeling) to supplement our results.

Homogeneity

"All subjects must have an equal probability of being captured (also known as catchability)." The degree of homogeneity can bias the results by both overestimating and underestimating the population (E.B. Hook & Regal, 1995; Nanan & White, 1997). Variable catchability is also a form of source dependency that can be reduced through log-linear models. In addition, this assumption can be met by stratifying the sample according to certain characteristics to look for differential estimations. An indicator of variable catchability is to explore if any population characteristics differ between the different sources. One data source may be more sensitive to a certain characteristic and therefore may be more likely to catch a person based on a particular characteristic.

Results

The observed total, unduplicated population reveals 3,818 new arrivals (i.e., immigrants and refugees) in the Lansing area during the period 2005–2007 (Table 1). The overlap of the sources is tabulated below.

Table 1 Overlap of Data Sources

Sources and ascertainment category			Number reported by specified sources and used for capture-recapture analysis
Schools (S)	Health (H)	Refugee (R)	
Yes	Yes	Yes	0
Yes	Yes	No	49
Yes	No	Yes	263
Yes	No	No	2,194
No	Yes	Yes	2
No	Yes	No	326
No	No	Yes	984
No	No	No	x = ?
Total ascertained in sources S, H, R			3,818

Examination of the data in rows 1 to 4 of Table 1 on schools (S), health (H), and refugee services (R) reveals the vast majority of cases (66%) are ascertained by schools (2,506/3,818), and, of these, by schools only (2,194/2,506). The overlap is presented diagrammatically in Figure 1.
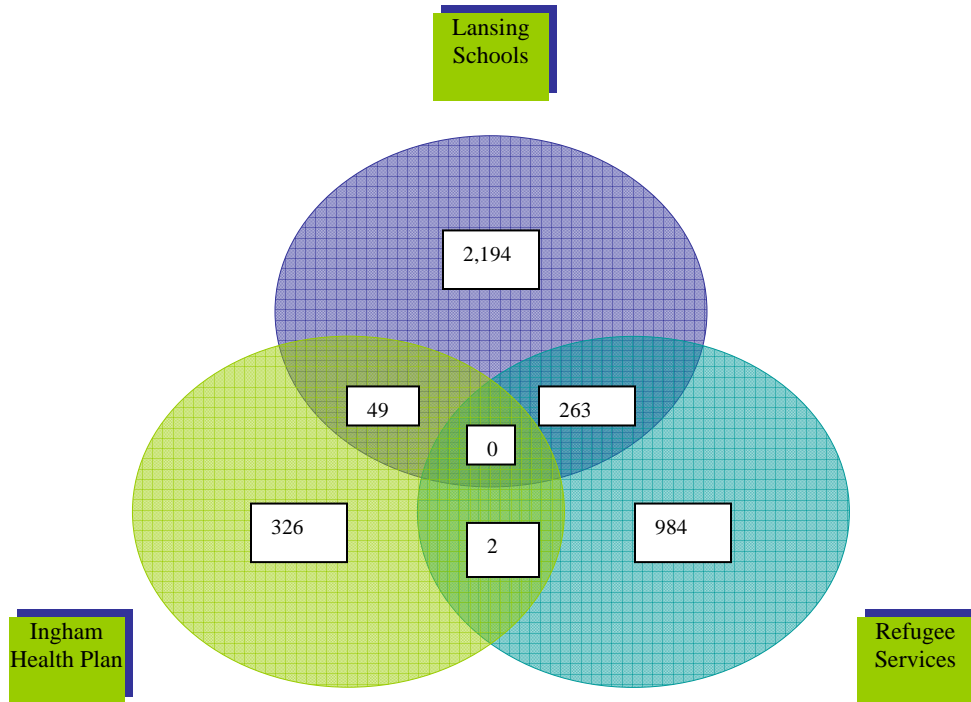
Figure 1. Overlap of Data Sources (2005–2007)

The demographic variables available to us were (unfortunately) missing data for many cases. Data regarding individuals' sex was missing for 32% of cases. In cases where data were available, slightly more newcomers were male (53%) than female. Data regarding individuals' race was missing for 54% of cases.

The Appendix contains information on two variables: country of origin and primary language spoken. Data regarding country of origin were stratified according to world region. The top regions of origin were Asia (30%) and Latin America (29%) followed by Africa (25%) (Figure 2). Lansing newcomers from Asia (n = 1,131) hailed primarily from Myanmar (20%) and Viet Nam (17%). Those originating from Latin America (n = 1,093) claimed Cuba (59%) and Mexico (28%) more often than other nations as their country of origin. New arrivals from Africa (n = 965) came primarily from Somalia (40%).
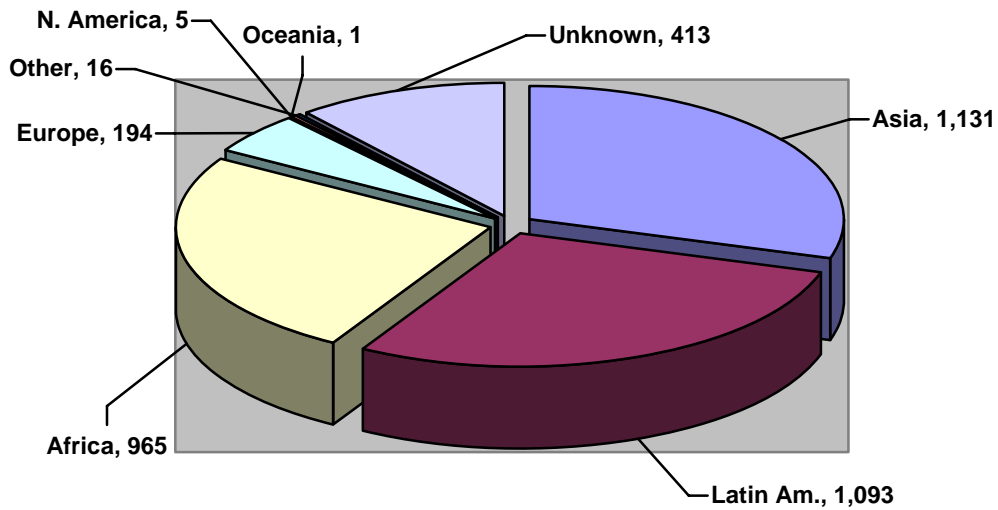
Figure 2. Distribution of countries of origin by world region (2005–2007)

Table 2 presents two-source estimates for approximating the missing cell in Table 1. If our prior expectation of negative dependence among sources is correct then all the two-source estimates in Table 2 tend to be biased high. Examination of the data structure indeed suggests that a negative H,R association exists producing a bias clearly on the high end for this two-source estimate. So the almost certain overestimate from the H,R model suggests that source H and R are at least somewhat negatively dependent in this population; and a log-linear approach which includes this interaction would be preferable to one that does not.

Table 2 Two Source Analyses and Estimates

Source	No. unique to listed source **(b)**	No. in another source but not in listed source **(c)**	No. both in listed source and some other source **(a)**	Estimated number missing $(bc/a) = $ **(x)**	Estimated total population\($n = a + b + c + x$)
S	2,194	1,312	312	9,226	13,044
H	326	3,441	51	21,995	25,813
R	984	2,569	265	9,539	13,357

Analysis (1st vs. 2nd)	No. in 1st but not in 2nd source **(b)**	No. in 2nd but not in 1st source **(c)**	No. in both sources **(a)**	Estimated number missing $(bc/a) = $ **(x)**	Estimated total population\($n = a + b + c + x$)
S vs. H	2,457	328	49	16,447	20,265
S vs. R	2,243	986	263	8,409	12,227
H vs. R	375	1,247	2	233,813	237,631

Table 3 presents the results derived from three sources for estimating the missing cell in Table 1. The last three rows summarize log-linear models with two pairwise interactions. The parameter modeling the interaction between refugee and health lists (R*H) again appears to be very important. Two of the three models that contain it fit the data adequately, based on p-values (Table 3).

We estimate the number of missing persons in Table 1 (x = ?) to be about 8,209 with a 95% confidence interval of 7,120–9,464. Although the last two rows both contain the interaction between refugee and health lists (R*H), the parameter also modeling the interaction between school and health lists (S*H, R*H) appears to provide the best fit, based on confidence interval (C.I.) width. Thus we estimate the total number of immigrants and refugees in Lansing between 2005 and 2007 to be somewhere between 10,938 (= 3,818 + 7,120) and 13,282 (= 3,818 + 9,464).

Table 3 Log-linear models used to estimate the number of immigrants and refugees whose names do not appear on any of the administrative lists

Terms in the log-linear model	X^2	df	p-value	Estimated missing count		C.I. Width
				Point	95% C.I.	
S, H, R	49.48	3	< 0.0001	10,383.8	(9,125.6, 11,815.5)	2,689.9
S*H, R	39.79	2	< 0.0001	9,537.6	(8,282.3, 10,983.2)	2,700.8
S*R, H	22.89	2	< 0.0001	21,938.5	(16,299.1, 29,529.2)	13,230.2
R*H, S	12.30	2	0.0021	9,228.0	(8,100.3, 10,512.7)	2,412.4
S*H, S*R	5.86	1	0.0155	159,851.4	(39,751.4, 642,807.1)	603,055.7
S*H, R*H	0.301	1	0.583	8,209.1	(7,120.3, 9,464.4)	2,344.1
R*H, S*R	0.534	1	0.465	14,559.5	(10,751.4, 19,716.5)	8,965.2

Discussion

As stated above, population estimates are crucial for serving immigrants and refugees because funding for many such activities is distributed on a per capita basis. Funding allocations are further limited to very specific populations and social need groups. In addition to the needs of resettlement agencies, communities are also very interested in acquiring estimates of migrants, because these populations are associated with both positive and negative actions that impact the social environment in which they live. These include basic demographic characteristics—such as age, family composition, health status, and educational level—as well as other concerns associated with criminal justice, local economics, housing, education, the tax base, and labor force participation. Without knowing about the nature of local populations, government, social agencies, employers, schools, and other stakeholders are incapable of developing well-informed service plans. However, by identifying population characteristics, various agencies and stakeholders become able to deal with changing demographics.

This project has allowed for the immigrant and refugee population of Lansing to be counted even when persons did not register with refugee services, utilize a local health plan, or enroll children in public schools. *Counted* here is the hidden population of foreign nationals as well as those persons known to have been present. In sum, based on these data there were an estimated 10,938–13,282 immigrants and refugees in Lansing between January 2005 and December 2007. Although these estimates should be considered upper limits of the true values because of a negative dependence, our estimates seem plausible based on previous figures cited in the Background section (page 3). Furthermore, the top ranking countries of origin revealed here for Lansing—e.g., Somalia, Myanmar, Burundi, Vietnam—are the same as those reported for all of Michigan by the Office of Refugee Resettlement (US Department of Health and Human Services, 2008), which lends further plausibility to our estimates.

Capture-recapture is a compelling estimation technique first used by wildlife researchers (Lincoln, 1930; Petersen, 1896), briefly used and forgotten by demographers (Sekar & Deming, 1949), borrowed by epidemiologists (Wittes & Sidel, 1968), improved by statisticians (Bishop et al., 1975; Fienberg, 1972) and employed once again by demographers (Nanan & White, 1997). The US government is using capture-recapture to control for undercounting in the US Census (Nanan & White, 1997). Even NASA has employed this technique to count the number of stars in the universe (Fienberg, 1998). Capture-recapture is being used in a creative fashion to estimate some difficult research populations.

Capture-recapture methodology, despite its promise and contributions to decision making, is not without its skeptics. There appears to be a polarization of estimation researchers. For example, Laporte (1994) contended that capture-recapture would change how population estimation was conducted in all disciplines. In contrast, Tilling (2001) boldly claimed that "the assumptions made when using simple capture-recapture methods are unlikely to be true in epidemiological studies" (p. 13). Capture-recapture pioneers, Hook and Regal, believe those on opposite ends of the spectrum may find some common ground in always being careful to interpret the results from the perspective of the eventual intended use of the estimates (Ernest B Hook & Regal, 1999, 2000).

One intended use of the estimation process described here was to evaluate the overlap in service usage among refugees and immigrants in Lansing. For example, access to health care continue to be a major public health concern among immigrant and refugee populations in the US—so much so that the *American Journal of Public Health* recently devoted an entire issue to the topic of "Health Without Borders" (Gonzalez Castro, 2008). However, less than one percent of the 377 foreign nationals who utilized the Ingham Health Plan were also assisted by refugee services. Although persons who utilized health services were likely a mix of immigrants and refugees, it is evident that the degree of overlap between the two services is minimal. This perhaps indicates a breakdown somewhere between intake with refugee services and presentation at a participating health provider.

Another intended use of the estimates derived herein was to serve as the beginning of making reliable data available to policy makers in the Lansing area. The field of refugee and immigrant studies has increased dramatically over the past several decades, although such increases have not necessarily been accompanied by significant policy impacts. (For a review of the growth of the field of refugee studies and its impact on policy, please see the article by Black (2001).) While we believe our work represents an important starting point in informing local policy debate, it also serves to generate many more ideas for much needed research. First, though, we must mention some of the limitations of the current project.

Limitations

As was mentioned in the "Data sources" section (above), we initially evaluated many more than just the three data sources ultimately used here. It proved very difficult to locate data sets with identifiers suitable for record linkage. Additionally, some of the data sources that would have proved useful were unavailable to us either because of the data format or concerns on the part of data holders. It would be extremely useful to note how our estimates might change with the addition of one or more additional data sources. However, as we noted above, the present study should be viewed as exploratory rather than definitive.

As for the data we were able to utilize, some difficulties were uncovered in terms of determining names and relationships between persons. Also, nationality was sometimes difficult to determine since refugees may have had one nationality or ethnicity yet lived in other countries. For children born in refugee camps, they often had a distinct nationality from their parents. Also, as is the case for any record linkage project that relies on surnames, slight misspellings or hyphenations can complicate the linkage process. However, we manually linked records rather than automating the process electronically, so this should have decreased the probability of mismatches.

Finally, we would have liked to have more data regarding the socioeconomic status of Lansing's refugees and immigrants—e.g., education, ethnicity, age, work experience, employment status, etc. More socioeconomic data would help us understand the needs and skills of this population better. Also, we did not ask about legal status for reasons of protection of respondents.

Conclusion

The immigration policy debate can be expected to continue as resource disparities between developed and developing countries increase migration pressures (White, 2007). According to the Migration Policy Institute, Michigan ranks 15th in the United States in terms of size of its foreign-born population (Migration Information Source, 2008b). The foreign-born population in the state increased by 14 percent between years 2000 and 2006. The "foreign-born" population includes (among other classes of individuals) immigrants as well as refugees and asylees. While such data are available at the state level, the present study represents the most comprehensive effort locally to generate population estimates for the Lansing area.

As we stated at the beginning of this report, a thoughtful policy debate can only take place if reliable data are available. This study was able to provide more thoughtful estimates of immigrants and refugees in Lansing, Michigan by using the technique, capture-recapture. We cannot stress strongly enough, however, that our results should be viewed as exploratory until further research can be conducted. Our recommendations for such research are outlined below.

Recommendations

Our present findings seem to generate many more research questions than they answer. Specifically, we recommend that researchers at Michigan State University collaborate with stakeholders in the Lansing community to address the following issues:

- If a research endeavor similar to the one described herein were conducted again, a number of additional variables should be specified including: a variable to denote whether an individual is a refugee or an immigrant; and variables related to socioeconomic status.

- Fund more in depth, qualitative research. A thorough literature review and interviews with local agency staff will help us explore local migrant populations, including their religious and cultural orientations, experiential backgrounds (including proximal hosts and local groups with whom they have conflicts), family and gender patterns, residential locations, economic strategies, and presence of community organizations.

- Research should be conducted locally to examine how social networks among various immigrant groups become activated in new settlement areas. There is a plethora of background literature on social network theory.

- More information is needed about small, emerging refugee populations.

- Typically, research exploring why people leave their home country and move to particular destinations highlight either macro- or micro-level factors. Local research should be conducted to assess the micro-level factors that either "push" or "pull" immigrants to Lansing.

- Immigrant and refugee adjustment to the conditions of the local labor market are generally assessed through a study of earnings. Research should be conducted to assess Lansing immigrant and refugees' adjustment to the labor market.

- Many newcomers to Lansing open businesses. Research should be conducted to assess whether foreign nationals receive the assistance they need in establishing businesses and, if not, how to overcome barriers.

References

Anderson, M. J., & Fienberg, S. (1999). *Who Counts? The Politics of Census-Taking in Contemporary America*. New York: Russell Sage Foundation.

Bardsley, M., & Storkey, M. (2000). Estimating the Numbers of Refugees in London. *Journal of Public Health Medicine, 22*(3), 406–412.

Barnes, S. (1995). *Michigan Heroin Prevalence Study*. Detroit (MI): Wayne State University.

Bilingual Education Department - Lansing School District. (2005). Student Language Count 2005–2006.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge (MA): MIT Press.

Black, R. (2001). Fifty years of refugee studies: From theory to policy. *International Migration Review, 35*(1), 57-78.

Co, C. Y., Euzent, P., & Martin, T. (2004). The export effect of immigration into the USA. *Applied Economics, 36*(6), 573-583.

Dunlevy, J. A. (2006). The influence of corruption and language on the pro-trade effect of immigrants: evidence from the American states. *Review of Economics and Statistics, 88*(1), 182-186.

Fienberg, S. E. (1972). The multiple recapture census for closed populations and incomplete 2k contingency tables. *Biometrika, 59*(3), 591–601.

Fienberg, S. E. (1998, March 31). *Innovative Methodology for Capture Recapture Studies.* Paper presented at the Workshop Notebook from the North Eastern Biometrics Conference, Pittsburgh (PA).

Gold, S. J. (1992). *Refugee Communities: A Comparative Field Study*. Newbury Park (CA): Sage.

Gonzalez Castro, F. (2008). Personal and ecological contexts for understanding the health of immigrants. *American Journal of Public Health, 98*(11), 1933.

Haines, D. W. (Ed.). (1996). *Refugees in America in the 1990s*. Westport (CT): Greenwood Press.

Health Plan Management Services. (2007–2008). County Health Plans. Retrieved October 16, 2008, from http://www.communityhealthplans.org/ihp/healthplans.asp?plan=ingham

Heer, D. (Ed.). (1968). *Social Statistics and the City*. Cambridge (MA): Harvard University Press.

Herman, P. (1994). A Technique for Estimating a Small Immigrant Population in Small Areas: The Case of Jewish Israelis in the United States. In K. V. Rao & J. W. Wicks (Eds.), *Studies in Applied Demography*. Bowling Green (OH): Population and Society Research Center.

Hook, E. B., & Regal, R. R. (1992). The value of capture-recapture methods even for apparent exhaustive surveys: The need for adjustment for source of ascertainment intersection in attempted complete prevalence studies. *American Journal of Epidemiology, 135*(9), 1060-1067.

Hook, E. B., & Regal, R. R. (1995). Capture-recapture methods in epidemiology: Methods and limitations. *Epidemiologic Reviews, 17*, 243–264.

Hook, E. B., & Regal, R. R. (1999). Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology. *Journal of Clinical Epidemiology, 52*(10), 917-926.

Hook, E. B., & Regal, R. R. (2000). Accuracy of alternative approaches to capture-recapture estimates of disease frequency: Internal validity analysis of data from five sources. *American Journal of Epidemiology, 152*(8), 771-779.

International Working Group for Disease Monitoring and Forecasting. (1995a). Capture-recapture and multiple-record systems estimation. 1. *American Journal of Epidemiology, 142*, 1047–1058.

International Working Group for Disease Monitoring and Forecasting. (1995b). Capture-recapture and multiple-record systems estimation. 2. *American Journal of Epidemiology, 142*, 1059–1068.

Jandl, M. (2004). The Estimation of Illegal Migration in Europe. *Migration Studies, XLI*(153), 141–155.

Kitano, H., Yeung, W. T., Chai, L., & Hatanaka, H. (1989). Asian American Interracial Marriage. *Journal of Marriage and the Family, 16*, 696–713.

LaPorte, R. E. (1994). Assessing the human condition: capture-recapture techniques. *British Medical Journal, 308*, 6920.

Lincoln, F. C. (1930). *Calculating waterfowl abundance on the basis of banking returns*: Cir. U.S. Dept Agric.

Migration Information Source. (2008a). Glossary.   Retrieved October 9, 2008, from http://www.migrationinformation.org/Glossary/

Migration Information Source. (2008b). US in Focus.   Retrieved October 7, 2008, from http://www.migrationinformation.com/USfocus/

Migration Information Source. (2008c). US in Focus: Spotlight on Legal Immigration to the United States.   Retrieved October 7, 2008, from http://www.migrationinformation.org/USfocus/display.cfm?id=651

Monserrat, M. R., & Cerda', J. C. M. (2002). Estimation With the Capture-recapture Method of the Number of Economic Immigrants in Mallorca. *Atencion Primaria, 29*(8), 463–468.

Nanan, D. J., & White, F. (1997). Capture-Recapture: Reconnaissance of a Demographic Technique in Epidemiology. *Chronic Diseases in Canada, 18*(4), 144–148.

Neugebauer, R., & Wittes, J. (1994). Annotation: Voluntary and involuntary capture-recapture samples—Problems in the estimation of hidden and elusive population. *American Journal of Public Health, 84*(7), 1068–1069.

Newbold, B. K. (2002). Estimating the Refugee Population from PUMS Data: Issues and Demographic Implications. *Growth and Change, 33*(3), 370–389.

Ong, A. (2003). *Buddha is Hiding: Refugees, Citizenship, The New America*. Berkeley (CA): University of California Press.

Petersen, C. G. J. (1896). The yearly immigration of young plaice into the Limfjord from the German Sea. *Rep Dan Biol Stan, 6*, 5–84.

Portes, A., & Bach, R. (1985). *Latin Journey: Cuban and Mexican Immigrants in the United States*. Berkeley (CA): University of California Press.

Prater, K. (2008). Lansing schools lose 688 students. *Journal*. Retrieved from
http://www.lansingstatejournal.com/article/20081205/NEWS05/812050327/1006/
NEWS05

Range, S. (2005a, March 28). Finding Their Place. *Lansing State Journal,* p. 1B,

Range, S. (2005b, March 27). Lansing Benefits from Refugees' Strength, Diversity.
Lansing State Journal, p. 1A (8A),

Refugee Development Center. (2008). Current Programs.   Retrieved October 16, 2008,
from http://www.refugeedevelopmentcenter.com/

Sekar, C. C., & Deming, W. E. (1949). On a method of estimating birth and death rates
and the extent of registration. *Journal of American Statistical Association, 44*,
101–115.

Singer, A., & Wilson, J. H. (2006). From 'There' to 'Here': Refugee Resettlement in
Metropolitan America. *Journal, September*. Retrieved from
http://www.brookings.edu/projects/Living-Cities.aspx

Skop, E. (2001). Race and Place in the Adaptation of Mariel Exiles. *International
Migration Review, 35*(1), 449–471.

Tilling, K. (2001). Capture-recapture methods—useful or misleading? *International
Journal of Epidemiology, 30*, 12–14.

US Department of Health and Human Services. (2008, September 23). Refugee Arrival
Data.   Retrieved November 24, 2008, from
http://www.acf.hhs.gov/programs/orr/data/refugee_arrival_data.htm

US Department of Homeland Security. (2008, September 26, 2008). Yearbook of
Immigration Statistics.   Retrieved October 7, 2008, from
http://www.dhs.gov/ximgtn/statistics/publications/yearbook.shtm

White, R. (2007). An Examination of the Danish Immigrant-Trade Link. *International
Migration, 45*(5), 61-86.

Wibert, L. A. (1998). *The Application of Capture-recapture for Indirect Population
Estimation.* Unpublished Dissertation, Michigan State University, East Lansing
(MI).

Wittes, J., & Sidel, V. W. (1968). A generalization of the simple capture-recapture model
with applications to epidemiological research. *Journal of Chronic Diseases, 21*,
287–301.

Zarrugh, L. (2008). The Latinization of the Central Shenandoah Valley. *International
Migration, 46*(1), 19-58.

Appendix

Distribution of countries of origin by UN macro region[5]: Asia (2005–2007)

COUNTRY OF ORIGIN	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
Myanmar	229	20.25%	229	20.25%
Viet Nam	190	16.80%	419	37.05%
Uzbekistan	121	10.70%	540	47.75%
Afghanistan	113	9.99%	653	57.74%
Iraq	111	9.81%	764	67.55%
Thailand	96	8.49%	860	76.04%
India	35	3.09%	895	79.13%
Saudi Arabia	30	2.65%	925	81.79%
Iran	30	2.65%	955	84.44%
Lao People's Democratic Republic	27	2.39%	982	86.83%
China	19	1.68%	1001	88.51%
Philippines	19	1.68%	1020	90.19%
Syria	18	1.59%	1038	91.78%
Korea (unspecified)	13	1.15%	1051	92.93%
Lao/Thailand	12	1.06%	1063	93.99%
Turkey	11	0.97%	1074	94.96%
Japan	8	0.71%	1082	95.67%
Armenia	6	0.53%	1088	96.20%
Georgia	6	0.53%	1094	96.73%
Yemen	6	0.53%	1100	97.26%
Lebanon	4	0.35%	1104	97.61%
Saudi Arabia/Iraq	4	0.35%	1108	97.97%
Singapore	4	0.35%	1112	98.32%
Turkey/Iraq	4	0.35%	1116	98.67%
Azerbaijan	3	0.27%	1119	98.94%
Iraq/Syria	3	0.27%	1122	99.20%
Pakistan	3	0.27%	1125	99.47%
Israel	2	0.18%	1127	99.65%
Pakistan/Afghanistan	2	0.18%	1129	99.82%
Jordan	1	0.09%	1130	99.91%
Viet Nam/Thailand	1	0.09%	1131	100.00%
TOTAL	**1131**	**100.00%**		

[5] United Nations (UN) world macro regions and components available at:
http://www.un.org/depts/dhl/maplib/worldregions.htm. Country of origin was either missing (n = 413) or spread out over more than one region (n = 16) for 429 individuals.

Distribution of countries of origin by UN macro region: Latin America (2005–2007)

COUNTRY OF ORIGIN	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
Cuba	646	59.10%	646	59.10%
Mexico	304	27.81%	950	86.92%
Haiti	77	7.04%	1027	93.96%
Honduras	16	1.46%	1043	95.43%
Guatemala	9	0.82%	1052	96.25%
Ecuador	8	0.73%	1060	96.98%
Colombia	7	0.64%	1067	97.62%
Brazil	5	0.46%	1072	98.08%
El Salvador	4	0.37%	1076	98.44%
Venezuela	4	0.37%	1080	98.81%
Costa Rica	3	0.27%	1083	99.09%
Dominican Republic	3	0.27%	1086	99.36%
Cuba/Mexico	2	0.18%	1088	99.54%
Jamaica	2	0.18%	1090	99.73%
Puerto Rico	2	0.18%	1092	99.91%
Guatemala/Honduras	1	0.09%	1093	100.00%
TOTAL	**1093**	**100.00%**		

Distribution of countries of origin by UN macro region: Africa (2005–2007)

COUNTRY OF ORIGIN	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
Somalia	388	40.21%	388	40.21%
Liberia	108	11.19%	496	51.40%
Burundi	98	10.16%	594	61.55%
Kenya	94	9.74%	688	71.30%
Sudan	91	9.43%	779	80.73%
Kenya/Somalia	28	2.90%	807	83.63%
Ethiopia	25	2.59%	832	86.22%
Sierra Leone	19	1.97%	851	88.19%
Congo	17	1.76%	868	89.95%
Unspecified	13	1.35%	881	91.30%
Cote d'Ivoire	10	1.04%	891	92.33%
Cameroon	9	0.93%	900	93.26%
Rwanda	9	0.93%	909	94.20%
Nigeria	8	0.83%	917	95.03%
Egypt	6	0.62%	923	95.65%
Eritrea	6	0.62%	929	96.27%
Sudan/Egypt	6	0.62%	935	96.89%
Western Africa (unspecified)	5	0.52%	940	97.41%
Ghana	4	0.41%	944	97.82%
Zimbabwe	4	0.41%	948	98.24%
Guinea	3	0.31%	951	98.55%
Libyan Arab Jamahiriya	3	0.31%	954	98.86%
United Republic of Tanzania	3	0.31%	957	99.17%
Botswana	2	0.21%	959	99.38%
Cote d'Ivoire/Liberia	2	0.21%	961	99.59%
Gambia	2	0.21%	963	99.79%
Congo/Gabon	1	0.10%	964	99.90%
Senegal	1	0.10%	965	100.00%
TOTAL	**965**	**100.00%**		

Distribution of countries of origin by UN macro region: Europe (2005–2007)

COUNTRY OF ORIGIN	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
Bosnia	73	37.63%	73	37.63%
Yugoslavia	26	13.40%	99	51.03%
Germany	21	10.82%	120	61.86%
Russian Federation	21	10.82%	141	72.68%
Romania	14	7.22%	155	79.90%
Croatia	11	5.67%	166	85.57%
Kosovo	9	4.64%	175	90.21%
Ukraine	4	2.06%	179	92.27%
Czech Republic	3	1.55%	182	93.81%
Spain	3	1.55%	185	95.36%
United Kingdom	3	1.55%	188	96.91%
Bosnia/Croatia	2	1.03%	190	97.94%
Italy	2	1.03%	192	98.97%
Yugoslavia/Russia	2	1.03%	194	100.00%
TOTAL	**194**	**100.00%**		

Distribution of countries of origin by UN macro region: North America (2005–2007)

COUNTRY OF ORIGIN	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
Canada	5	100.00%	5	100.00%
TOTAL	**5**	**100.00%**		

Distribution of countries of origin by UN macro region: Oceania (2005–2007)

COUNTRY OF ORIGIN	FREQUENCY	PERCENT	CUMULATIVE FREQUENCY	CUMULATIVE PERCENT
Papua New Guinea	1	100.00%	1	100.00%
TOTAL	**1**	**100.00%**		

PRIMARY LANGUAGE

LANGUAGE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
.	28	1.18	28	1.18
.swahili	2	0.08	30	1.27
Albanian (Kosovo)	29	1.23	59	2.50
Amharic (Ethiopia)	13	0.55	72	3.05
Arabic	175	7.41	247	10.45
Arabic/Spanish	2	0.08	249	10.54
Arabic/farsi	2	0.08	251	10.62
Armenian, Azeri	6	0.25	257	10.88
Bantu, Kikuyua	14	0.59	271	11.47
Bantu, Kikuyua/somali	1	0.04	272	11.51
Bengali, Bihari	3	0.13	275	11.64
Chaldean (Iraq)	19	0.80	294	12.44
Chinese (Kuo Yu)	13	0.55	307	12.99
Czeck, Slovak	3	0.13	310	13.12
English	46	1.95	356	15.07
English/liberiuan/kpelle	1	0.04	357	15.11
Farsi (Iran)	58	2.45	415	17.56
Farsi (Iran)/persian, syrian	2	0.08	417	17.65
French	77	3.26	494	20.91
French/haitian/creole	2	0.08	496	20.99
German	11	0.47	507	21.46
Haitian Creole	59	2.50	566	23.95
Hindi, Marathi	11	0.47	577	24.42
Hmong	135	5.71	712	30.13
Ibo, Idoma	2	0.08	714	30.22
Italian, Romanic	2	0.08	716	30.30
Jamaican Creole	2	0.08	718	30.39
Japanese, Ainu	3	0.13	721	30.51
Korean	15	0.63	736	31.15
Kurdish, Turkish	40	1.69	776	32.84
Kwa, Masai, Dinka	10	0.42	786	33.26
Lao, Miao, Tai Dam	3	0.13	789	33.39
Lebanese	1	0.04	790	33.43
Lebanese/Arabic	2	0.08	792	33.52
Liberian, Kpelle	57	2.41	849	35.93
Liberian, Kpelle, Somali	1	0.04	850	35.97
Liberian, Kpelle, french	1	0.04	851	36.01
Malay	4	0.17	855	36.18
Nyanja (Zambia)	3	0.13	858	36.31
Persian, Syrian	15	0.63	873	36.94
Portuguese	5	0.21	878	37.16
Romanian	14	0.59	892	37.75
Russian, Ukranian	23	0.97	915	38.72
Serbo-Croatian	100	4.23	1015	42.95
Shona, NDebele	4	0.17	1019	43.12
Somali	302	12.78	1321	55.90

PRIMARY LANGUAGE

LANGUAGE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Somali/swahili	8	0.34	1329	56.24
Spanish	762	32.25	2091	88.49
Spanish/Arabic/Somali	1	0.04	2092	88.53
Swahili (Kenya)	26	1.10	2118	89.63
Tagalog (Filipino)	13	0.55	2131	90.18
Teamaie	11	0.47	2142	90.65
Telugu, Tamil	9	0.38	2151	91.03
Tibetan, Burmese	11	0.47	2162	91.49
Twi	2	0.08	2164	91.58
Unknown (Blank)	7	0.30	2171	91.87
Vietnamese	186	7.87	2357	99.75
Yoruba, Hausa	6	0.25	2363	100.00

Frequency Missing = 1455