Check for updates

INTERNATIONAL MIGRATION
International Organization for Migration (IOM)

doi:10.1111/j.1468-2435.2011.00715.x

# A Capture–Recapture Approach to Estimation of Refugee Populations

Steven J. Gold,* Wilma Novales Wibert,** Vera Bondartsova,*
Brian J. Biroscak*** and Lori A. Post***

## ABSTRACT

The debate over immigration continues to be 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 of 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 geographical areas. This paper describes a recent effort to generate serviceable data on the immigrant and refugee population for a medium-sized metropolitan area in the US.

The objectives of this research were twofold. Our first goal was to provide local stakeholders with information to assist them with resettling and obtaining funding for immigrants and refugees. The second aim was to develop better techniques for tabulating diverse refugees and immigrants in a medium-sized community. By comparing and contrasting three data sources – that is, refugee services, public schools and a local health plan – we are able to generate estimates of the local refugee and immigrant population.

During the period from 2005 to 2007, we estimate the total number of immigrants and refugees in the community 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 for the region. While such estimates are valuable, a number of shortcomings related to the data prevent us from painting a more complete picture of these populations. We conclude this paper with a number of recommendations that will assist others in planning research designed to inform migration policy and practice in medium-sized metropolitan areas.

## INTRODUCTION

In the introduction to their jointly authored report, *Immigration: Data Matters*, scholars from the Population Reference Bureau and Migration Policy Institute write "In an era when immigration has become such a polarized topic – and people are trying to understand how

* Department of Sociology, Michigan State University, East Lansing, Michigan, USA.
** Department of Communication, Michigan State University, East Lansing, Michigan, USA.
*** Department of Emergency Medicine, Yale University, New Haven, Connecticut, USA.

© 2012 The Authors
International Migration © 2012 IOM
*International Migration* Vol. 53 (5) 2015
ISSN 0020-7985
Published by John Wiley & Sons Ltd.

immigration is shaping the US population, economy, educational and labor systems, and communities – it is imperative to have access to credible, authoritative and high quality data'' (Batalova et al., 2008: 5). However, those familiar with the study of migrants and refugees know that accurate, up-to-date and detailed data are very difficult to obtain. In the introduction to *Crossing the Border: Research from the Mexican Migration Project*, Jorge Durand and Douglas S. Massey (2004: 2) assert ''As is well known to demographers ... migration data are among the weakest and least reliable statistics routinely collected and published by the federal government, and they were among the last to receive systematic methodological attention from the profession.''

Reliable data should form the basis of policy development and practice, not only with respect to refugee resettlement and immigration, but in regard 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 per cent of all lawful permanent immigration in 2006; employment-preference immigrants accounted for nearly 13 per cent of all lawful permanent immigration; and about 17 per cent 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 can help inform US immigration and refugee resettlement policy and assist with planning for service provision. In the case of Latin American immigrant groups, for example, Zarrugh (2008: 49) 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'' 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 provides data on a variety of classes of foreign nationals (US Department of Homeland Security, 2008). US policymakers 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, non-immigrant admissions, naturalizations and enforcement actions.

However, immigrants and refugees 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, but 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 Greater Lansing area has no such mechanism. This paper works towards fulfilling this need by suggesting the use of Capture–Recapture (CR) methodology to provide a detailed estimate of the numbers and social characteristics of immigrants and refugees in Lansing, Michigan, a medium-sized metropolitan area. We begin by describing the impetus for the present research project.

## Background

Although precise figures are lacking, the Great Lansing area has become a significant point of settlement for both immigrants and refugees. For example, the Brookings Institution ranked Greater Lansing tenth among the top ten medium-sized metropolitan areas for number of refugees resettled during the period from 1983 to 2004 (Singer and Wilson, 2006). In addition, diverse arrays of immigrants have come of their own volition. Regardless of the conditions surrounding their arrival, however, these populations encounter many of the same

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

challenges as they build lives in a new area. For example, they must find a place to live, learn English, acquire employment, enrol their children in school, obtain health care and fulfil 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 Greater 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 Immigrant and Refugee Resource Collaborative (IRRC). The IRRC exists to provide assistance to Greater 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. This study was conducted in response to needs encountered by local city government, including the Lansing IRRC, to provide enumeration of immigrant and refugee populations.

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 from 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 decade from 2000–2010, almost 2,000 more refugees made Lansing their home (Range, 2005b). An estimated total of 13,000 refugees have settled in Lansing from 1975 to 2005. In fact, the Capital Area took more refugees than any other city in Michigan in 2004 (Range, 2005a). Although rough approximations (such as the above-mentioned) exist, we know very little about the accuracy of estimates 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 US$6–7 an hour (Range, 2005b). Another 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 2005–2006 (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 enrolment – and hence its financial support – would be diminished. In fact, the recently announced loss of more than US$3 million in state funding for Lansing schools (Prater, 2008) would have been even greater without the enrolment of immigrants and refugees.

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

### 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 the area, 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 and 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.

## The scope of this paper

The challenge 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 – that is, how to count the number of immigrants and refugees. In this paper, we focus on the application of one such method, termed ''capture–recapture'' (Hook and Regal, 1995; International Working Group for Disease Monitoring and Forecasting, 1995a,b). After presenting our population estimates, we discuss the implications in terms of future enumeration efforts.

The present study is limited to an estimation of the Lansing area's immigrant and refugee population, as well as an examination of the available population characteristics. No distinction was made in the present study between legal and undocumented immigrants. The scope of this report is also limited to descriptive analyses – that is, counting and classifying individuals – rather than employment of inferential statistics – for example, 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). 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 the 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 permitted to enter the US because

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 labour market. Asylum seekers include anyone who files an application for asylum in a country other than his or her own. (Their status remains that of asylum seekers until their applications are 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 that some members of this group are not observable (in research jargon) because they are not recorded in a standardized, accessible database. Neither a census nor a sample would capture this invisible population. Furthermore, recent documented or legal migrants may avoid census counts (US Census, 2003). 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 then 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 et al., 1975).

The capture–recapture method is now being used in many epidemiological studies to determine the prevalence of certain diseases or human behaviours of interest to public health, in cases where such populations are also difficult to find and to study (e.g. Lange et al., 2004; Lugardon et al., 2006; Roberts and Brewer, 2006; Snidero et al., 2007; Brewer et al., 2008; Cristiano et al., 2009). However, it is has come to be used for estimates of population data too.

Much of the available population research involves data derived from the Census and other official statistical sources (e.g. Bardsley and Storkey, 2000; Skop, 2001; Newbold, 2002). Other notable studies suggest inventive ways to draw on alternative data sources – for example, public documents such as citizenship applications and marriage licenses – that record both ethnicity and demographic data (e.g. Kitano et al., 1989; Herman, 1994). 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, labour union rosters and health-related institutional 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 that an additional 33,034 individuals were not included in any of the data sets (Monserrat and Cerda', 2002).

The use of capture–recapture data in population studies is undergoing some development and expansion (O'Hara et al., 2009). Recent studies have focused on how to deal with some of the problems encountered when using the capture–recapture method to estimate the size of a population. Such problems include density dependence, sample independence, population movement, population definition and sample heterogeneity (Bloor, 2005; Pledger, 2005; Xu et al., 2007; Bartolucci and Lupparelli, 2008; Böhning, 2008; Malec and Maples, 2008; Schofield and Barker, 2008; Xi et al., 2008). Cadwell et al. (2005) propose a procedure to

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

deal with cases where the number of individuals appearing in both lists of the capture–recapture data lack personal identifiers.

As indicated by the above-mentioned study (Monserrat and Cerda', 2002), the 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 (Portes and Bach, 1985; Haines, 1996; Ong, 2003). 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. Moreover, a positive rapport between members of this research team and the IRRC is critical to facilitate access to the required multiple data sets. Finally, the very large samples that exist in major immigrant gateway cities would have been too massive for the painstaking matching procedures required by our method. However, the number of immigrants and refugees in the Lansing area was appropriate for the application of this technique. This led us to propose a Lansing-based study to address the following research questions.

### Research objectives

The aims of this research were twofold. Our first objective was to provide local stakeholders with the information that they requested to assist them with resettling and obtaining funding for immigrants and refugees. This research had the potential to 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. Our second goal was to explore 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.

## METHODS AND MATERIALS

The research project described below was submitted for Institutional Review Board approval on 25 May 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

The capture–recapture method is derived from the practice of sampling a defined wildlife population, marking the captured sample members for later identification purposes and then releasing them back into the population. A second sample (which does not need to be random) is taken from the same population, 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 the method adjusts for any individual

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

source not containing all cases, it generally assumes that the cases that a source does contain truly are 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) for refinement of estimations derived from surveys; (3) for evaluation of data registries; and (4) to derive plausible upper or lower limits on the total affected population via confidence intervals.

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 and Wittes, 1994):

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

For example, if 100 deer are captured at Time 1 and released, then another sample of 100 deer is captured at Time 2. Five of the deer at Time 2 had also been captured at Time 1, meaning that 5 per cent 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 that the 100 deer captured during Time 2 equals 5 per cent 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 by means of which 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 and Deming, 1949), the total population of immigrants and refugees in a defined area (and during a defined time period) can be estimated. If the necessary assumptions are met, as stated below, we can assume ''unity'' and the overlap can be operationalized to estimate the missing population (Fienberg, 2007).

In order to apply 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 iden-

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tifiers that are sufficient to enable the positive matching of all cases between the 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, it was still possible for those persons to be 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 (Hook and Regal, 1992).

### Catchability

With capture–recapture methodology, the probability of ascertainment varies by any of the institutional sources. Catchability can be influenced because of geographical or socio-economic variables (Hook and Regal, 1995). The data must be modelled (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 (Hook and Regal, 1995). It will be difficult to control for probable causes of catchability in the present research, however, because of the dearth of covariates in the data sets.

### Source dependency

Although the final assumption (independence) must be addressed sufficiently, 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 modelling may still be usefully employed to generate estimates of the unobserved members of a population. As stated by Nanan and White (1997: 145), ''... the assumption of independence can be dropped, and interdependence among data sets can be accounted for by using Log-linear modelling techniques to assess source dependencies''. Furthermore, when the number of cases contained on each list is large, the estimates are less sensitive to dependency of samples and catchability (Hook and Regal, 1995). We continue below with a description of data sources used for both dual-source enumeration and log-linear modelling of the population of immigrants and refugees in Lansing.

If the necessary assumptions are not met, the derived capture–recapture population estimates will be invalid, as they produce significant over- and underestimates because they are sensitive to the necessary conditions (Post et al., 2006).

## 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

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 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) – that is, 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), it 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 Michigan State University, 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 analytical sources. An analytical source is ''a list which may be constructed from one or more original sources'' (Hook and Regal, 1995: 244). The original sources – that is, ''a particular list of cases from some institution'' (Hook and Regal, 1995: 244) – have been collapsed into the analytical 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 who were bilingual and enrolled for classes in the years 2005 ($n$ = 2,365), 2006 ($n$ = 2,247) or 2007 ($n$ = 2,246), together with information about their parents/guardians. 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 probably have been connected with schools via referral from refugee services (see below). Because of the nationalities of students listed in this data set, as well as the availability of university housing and local residential patterns, we are confident that very few of the students included in the Lansing Schools data were the children of international faculty or student families associated with Michigan State University. The children of such families commonly attend suburban school districts, not Lansing's.

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 that 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

© 2012 The Authors. International Migration © 2012 IOM

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.[2] We subsequently obtained an additional file from the Lansing School District. After completing an identical process as above – that is, 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 centres that served the area under the auspices of the Ingham Health Plan ($n$ = 779). While not an insurance programme, Ingham Health Plan provides basic medical care to low-income uninsured residents of Ingham County (Health Plan Management Services, 2007–2008).

Again, refugees would probably be referred to the health programme via refugee services. However, a number of immigrants as well as refugees most 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 several training modules designed to provide newcomers with necessary skills, such as 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.

In summary, we relied on data from three sources and excluded cases outside of our analytical time period, 2005–2007. Capture–recapture frequently relies on 3 years of data or more. This eliminates the volatility of year effect, as is consistent with demographic and public health estimates. Public health professionals commonly collapse data across time periods and calculate rates by combining numerators and denominators for two or three years of data rather than using the annual numbers. This approach increases the stability of the resulting rates by increasing the sample size at each time point. We used the same 3-year time period for the three data sources, because having matching time periods is a necessary assumption of using the CR technique.

## 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

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 the 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, he or she still had the potential to be captured in the three sources during his or her 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 some progress still needs to be made, 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 enrol 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 appearing 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 either 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 (Hook and Regal, 1992, 1995).

We anticipated that there should be negative dependence among the data sources – that is, 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 – for example, health care that is easier to obtain than working with the Ingham Health Plan. Thus we expect that 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

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

multi-source analyses (log-linear modelling) to supplement our results and to identify and control for any sample dependencies, both negative or positive.

### 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 (Hook and Regal, 1995; Nanan and 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 population characteristics, to look for differential estimations. An exploration of whether any of these differ between the different sources provides an indicator of variable catchability. One data source may be more sensitive to a particular characteristic and therefore may be more likely to catch a person on that basis.

## RESULTS

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

Examination of the data in rows 1–4 of Table 1 on schools (S), health (H) and refugee services (R) reveals that 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.

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

The Appendix contains information on country of origin. 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%). Lansing area 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%).

Table 2 presents two-source estimates for approximating the missing cell in Table 1 ($x$ = ?). 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 that includes this interaction would be preferable to one that does not.

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 modelling 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).

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

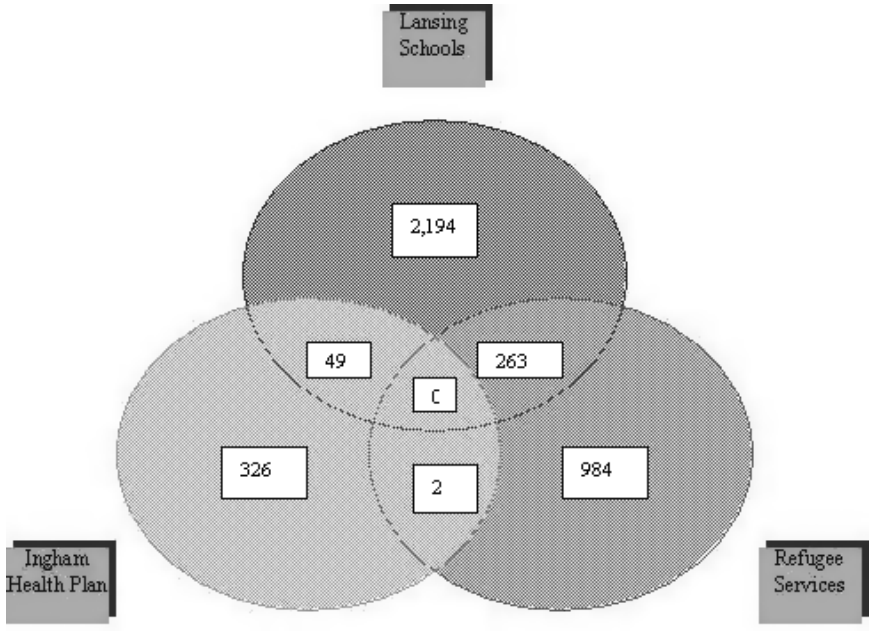
TABLE 1

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 |

FIGURE 1

OVERLAP OF DATA SOURCES, 2005–2007




We estimate the number of missing persons in Table 1 ($x$ = ?) to be about 8,209, with a 95 per cent 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 modelling 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).

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2

TWO-SOURCE ANALYSES AND ESTIMATES

| Source | Number unique to listed source, *b* | Number in another source but not in listed source, *c* | Number 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 (first vs. second)** | **Number in first but not in second source, *b*** | **Number in second but not in first source, *c*** | **Number 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 |

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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. Finally, because refugees are entitled to certain government services, policymakers and scholars seek to determine the rates of utilizing such services at the local level. 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 labour-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 enrol 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 (p. 3). Furthermore, the top-ranking countries of origin revealed here for Lansing – for example, Somalia, Myanmar, Burundi and Viet Nam – 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 (Petersen, 1896; Lincoln, 1930), briefly used and forgotten by demographers (Sekar and Deming, 1949), borrowed by epidemiologists (Wittes and Sidel, 1968), improved by statisticians (Fienberg, 1972; Bishop et al., 1975) and employed once again by demographers (Bloor, 2005; Malec and Maples, 2008; O'Hara et al., 2009). The US government is using capture–recapture to control for undercounting in the US Census (Nanan and White, 1997). Even

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 sceptics. There appears to be a polarization of estimation researchers. For example, Laporte (1994) contended that this approach would change how population estimation was conducted in all disciplines. In contrast, Tilling (2001: 13) boldly claimed that ''the assumptions made when using simple capture–recapture methods are unlikely to be true in epidemiological studies''. The pioneers of the method, Hook and Regal, believe that 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 (Hook and 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 continues 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 1 per cent of the 377 foreign nationals who utilized the Ingham Health Plan were also assisted by refugee services. Although persons who utilized health services were probably 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 policymakers in the Lansing area (and in response to a need for such data from the IRRC, a local organization devoted to refugees and immigrant service provision). 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, see Black, 2001). While we believe that our work represents an important starting point in informing the 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, we initially evaluated many more than just the three sources ultimately used here. It proved very difficult to locate data sets with identifiers suitable for record linkage. Additionally, some of the sources that would have proved useful were unavailable to us, either because of the data format or because of 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 that 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 and yet lived in other countries. Children born in refugee camps often had a different nationality from that of 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.

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Finally, we would have liked to have more data regarding the socio-economic status of Lansing's refugees and immigrants – for example, education, ethnicity, age, work experience, employment status and so on. More socio-economic 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 US in terms of size of its "foreign-born population" (Migration Information Source, 2008b), which increased by 14 per cent between the years 2000 and 2006. The term 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 paper, 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 capture–recapture technique. We cannot stress strongly enough, however, that our results should be viewed as exploratory until further research can be conducted.

We hope that researchers and policymakers can draw on this exercise to develop more refined techniques for estimating populations of immigrants and refugees, which have been described by leading demographers as both essential for sound policymaking and yet scarce and poorly developed. The available techniques, including those presented in this paper, have the potential to provide a better understanding of the needs of such populations. At the same time, the resulting data will yield a timely and accurate basis for assessing the social, demographic and economic impact of newcomers within local communities.

## NOTES

1. More recent figures indicate that this decline has continued, with an estimated loss of 4,426 residents between 2000 and 2007 (American Fact Finder; http://factfinder.census.gov).
2. The sample size was 2,450 rather than 2,451 (i.e. 1,184 + 1,267) because one repeated record was later found.

## REFERENCES

Anderson, M.J., and S. Fienberg
1999 *Who Counts? The Politics of Census-taking in Contemporary America*, Russell Sage Foundation, New York.

Bardsley, M., and M. Storkey
2000 "Estimating the numbers of refugees in London", *Journal of Public Health Medicine*, 22(3): 406–412.

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

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bartolucci, F., and M. Lupparelli
2008 "Focused information criterion for capture–recapture models for closed populations", *Scandinavian Journal of Statistics*, 35(4): 629–649.
Batalova, J., M. Mittelstadt, M. Mather, et al.
2008 *Immigration: Data Matters*, Migration Policy Institute and Population Reference Bureau, Washington, DC.
Bilingual Education Department – Lansing School District
2005 Student Language Count 2005–2006.
Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland
1975 *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, MA.
Black, R.
2001 "Fifty years of refugee studies: from theory to policy", *International Migration Review*, 35(1): 57–78.
Bloor, M.
2005 "Population estimation without censuses or surveys: a discussion of mark-recapture methods illustrated by results from three studies", *Sociology – the Journal of the British Sociological Association*, 39(1): 121–138.
Böhning, D.
2008 "Editorial – recent developments in capture–recapture methods and their applications", *Biometrical Journal*, 50(6): 954–956.
Brewer, D.D., J.M. Roberts, S.Q. Muth, et al.
2008 "Prevalence of male clients of street prostitute women in the United States", *Human Organization*, 67(3): 346–356.
Cadwell, B.L., P.J. Smith, and A.L. Baughman
2005 "Methods for capture–recapture analysis when cases lack personal identifiers", *Statistics in Medicine*, 24(13): 2041–2051.
Co, C.Y., P. Euzent, and T. Martin
2004 "The export effect of immigration into the USA", *Applied Economics*, 36(6): 573–583.
Cristiano, E., L. Patrucco, J.I. Rojas, et al.
2009 "Prevalence of multiple sclerosis in Buenos Aires, Argentina using the capture–recapture method", *European Journal of Neurology*, 16(2): 183–187.
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.
Durand, J., and D.S. Massey (Eds)
2004 *Crossing the Border: Research from the Mexican Migration Project*, Russell Sage Foundation, New York.
Fienberg, S.E.
1972 "The multiple recapture census for closed populations and incomplete 2k contingency tables", *Biometrika*, 59(3): 591–601.
1998 "Innovative methodology for capture recapture studies", Paper presented at the Workshop Notebook from the North Eastern Biometrics Conference, Pittsburgh (PA), 31 March.
2007 *The Analysis of Cross-classified Categorical Data*, Springer, New York.
Gold, S. J.
1992 *Refugee Communities: a Comparative Field Study*, SAGE Publications, Newbury Park, CA.
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*, Greenwood Press, Westport, CT.
Health Plan Management Services
2007–2008 "County health plans", http://www.communityhealthplans.org/ihp/healthplans.asp?plan=ingham (accessed 16 October 2008).
Heer, D. (Ed.)
1968 *Social Statistics and the City*, Harvard University Press, Cambridge, MA.

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 and J. W. Wicks (Eds), *Studies in Applied Demography*, Population and Society Research Center, Bowling Green, OH.
Hook, E.B., and R.R. Regal
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.
1995 ''Capture–recapture methods in epidemiology: methods and limitations'', *Epidemiologic Reviews*, 17: 243–264.
1999 ''Recommendations for presentation and evaluation of capture–recapture estimates in epidemiology'', *Journal of Clinical Epidemiology*, 52(10): 917–926.
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.
Kitano, H., W.T. Yeung, L. Chai, et al.
1989 ''Asian American interracial marriage'', *Journal of Marriage and the Family*, 16: 696–713.
Lange, J.H., Y.-F. Chang, and R.E. LaPorte
2004 ''Use of the capture–recapture method for epidemiological studies in determining prevalence'', *Acta Neurologica Scandinavica*, 109(1): 79–80.
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*, Circular, US Department of Agriculture.
Lugardon, S., K. Desboeuf, P. Fernet, et al.
2006 ''Using a capture–recapture method to assess the frequency of adverse drug reactions in a French university hospital'', *British Journal of Clinical Pharmacology*, 62(2): 225–231.
Malec, D., and J. Maples
2008 ''Small area random effects models for capture/recapture methods with applications to estimating coverage error in the U.S. Decennial Census'', *Statistics in Medicine*, 27(20): 4038–4056.
Migration Information Source
2008a ''Glossary'', http://www.migrationinformation.org/Glossary/ (accessed 9 October 2008).
2008b ''US in focus'', http://www.migrationinformation.com/USfocus/ (accessed 7 October 2008).
2008c ''US in focus: spotlight on legal immigration to the United States'', http://www.migrationinformation.org/USfocus/display.cfm?id = 651 (accessed 7 October 2008).
Monserrat, M.R., and J.C.M. Cerda'
2002 ''Estimation with the capture–recapture method of the number of economic immigrants in Mallorca'', *Atencion Primaria*, 29(8): 463–468.
Nanan, D.J., and F. White
1997 ''Capture–recapture: reconnaissance of a demographic technique in epidemiology'', *Chronic Diseases in Canada*, 18(4): 144–148.
Neugebauer, R., and J. Wittes
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.

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

O'Hara, R.B., S. Lampilla, and M. Orell
2009 "Estimation of rates of births, deaths, and immigration from mark–recapture data", *Biometrics*, 65(1): 275–281.
Ong, A.
2003 *Buddha is Hiding: Refugees, Citizenship, The New America*, University of California Press, Berkeley, CA.
Petersen, C.G.J.
1896 "The yearly immigration of young plaice into the Limfjord from the German Sea", *Reports of the Danish Biological Station*, 6: 5–84.
Pledger, S.
2005 "The performance of mixture models in heterogeneous closed population capture–recapture", *Biometrics*, 61(3): 868–873.
Portes, A., and R. Bach
1985 *Latin Journey: Cuban and Mexican Immigrants in the United States*, University of California Press, Berkeley, CA.
Post, L.A., H. Zhang, G. Barboza, et al.
2006 "Simulations demonstrate feasibility of capture–recapture", Published proceedings from the 2006 International Conference on Statistics, Mathematics and Related Fields, Hawaii.
Prater, K.
2008 "Lansing schools lose 688 students" [electronic version], *Lansing State Journal*, http://www.lansingstatejournal.com/article/20081205/NEWS05/812050327/1006/NEWS05 (accessed 6 December 2008).
Range, S.
2005a "Finding their place", *Lansing State Journal*, 28 March, p. 1B.
2005b "Lansing benefits from refugees' strength, diversity", *Lansing State Journal*, 27 March, p. 1A (8A).
Refugee Development Center
2008 "Current programs", http://www.refugeedevelopmentcenter.com/ (accessed 16 October 2008).
Roberts, J.M.J., and D.D. Brewer
2006 "Estimating the prevalence of male clients of prostitute women in Vancouver with a simple capture–recapture method", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(4): 745–756.
Schofield, M.R., and R.J. Barker
2008 "A unified capture–recapture framework", *Journal of Agricultural Biological and Environmental Statistics*, 13(4): 458–477.
Sekar, C.C., and W.E. Deming
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., and J.H. Wilson
2006 "From 'there' to 'here': refugee resettlement in metropolitan America" [electronic version], *Living Cities Census Series*, September, http://www.brookings.edu/projects/Living-Cities.aspx (accessed 1 November 2008).
Skop, E.
2001 "Race and place in the adaptation of Mariel exiles", *International Migration Review*, 35(1): 449–471.
Snidero, S., B. Morra, R. Corradetti, et al.
2007 "Use of the scale-up methods in injury prevention research: An empirical assessment to the case of choking in children", *Social Networks*, 29(4): 527–538.
Tilling, K.
2001 "Capture–recapture methods – useful or misleading?" *International Journal of Epidemiology*, 30: 12–14.
US Department of Health and Human Services
2008 "Refugee arrival data", 23 September, http://www.acf.hhs.gov/programs/orr/data/refugee_arrival_data.htm (accessed 24 November 2008).

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

US Department of Homeland Security
2008 ''Yearbook of immigration statistics'', 26 September, http://www.dhs.gov/ximgtn/statistics/publications/yearbook.shtm (accessed 7 October 2008).
White, R.
2007 ''An examination of the Danish immigrant-trade link'', *International Migration*, 45(5): 61–86.
Wittes, J., and V.W. Sidel
1968 ''A generalization of the simple capture–recapture model with applications to epidemiological research'', *Journal of Chronic Diseases*, 21: 287–301.
Xi, L., R. Watson, and P.S.F. Yip
2008 ''The minimum capture proportion for reliable estimation in capture–recapture models'', *Biometrics*, 64(1): 242–249.
Xu, Y., L. Liu, N. You, et al.
2007 ''Estimating population size for a continuous time frailty model with covariates in a capture–recapture study'', *Biometrics*, 63(3): 917–921.
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: ASIA, 2005–2007

| Country of origin[a] | Frequency | Percentage | Cumulative frequency | Cumulative percentage |
|---|---|---|---|---|
| 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 | 1,001 | 88.51 |
| Philippines | 19 | 1.68 | 1,020 | 90.19 |
| Syria | 18 | 1.59 | 1,038 | 91.78 |
| Korea (unspecified) | 13 | 1.15 | 1,051 | 92.93 |
| Lao/Thailand | 12 | 1.06 | 1,063 | 93.99 |
| Turkey | 11 | 0.97 | 1,074 | 94.96 |
| Japan | 8 | 0.71 | 1,082 | 95.67 |
| Armenia | 6 | 0.53 | 1,088 | 96.20 |
| Georgia | 6 | 0.53 | 1,094 | 96.73 |
| Yemen | 6 | 0.53 | 1,100 | 97.26 |
| Lebanon | 4 | 0.35 | 1,104 | 97.61 |
| Saudi Arabia/Iraq | 4 | 0.35 | 1,108 | 97.97 |
| Singapore | 4 | 0.35 | 1,112 | 98.32 |
| Turkey/Iraq | 4 | 0.35 | 1,116 | 98.67 |
| Azerbaijan | 3 | 0.27 | 1,119 | 98.94 |
| Iraq/Syria | 3 | 0.27 | 1,122 | 99.20 |
| Pakistan | 3 | 0.27 | 1,125 | 99.47 |
| Israel | 2 | 0.18 | 1,127 | 99.65 |
| Pakistan/Afghanistan | 2 | 0.18 | 1,129 | 99.82 |
| Jordan | 1 | 0.09 | 1,130 | 99.91 |

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(*CONTINUED*)

| Country of origin[a] | Frequency | Percentage | Cumulative frequency | Cumulative percentage |
|---|---|---|---|---|
| Viet Nam/Thailand | 1 | 0.09 | 1,131 | 100.00 |
| TOTAL | 1,131 | 100.00 | | |

[a]The United Nations (UN) world macro-regions and components are available at http://www.un.org/depts/dhl/maplib/worldregions.htm. The 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 | Percentage | Cumulative frequency | Cumulative percentage |
|---|---|---|---|---|
| Cuba | 646 | 59.10 | 646 | 59.10 |
| Mexico | 304 | 27.81 | 950 | 86.92 |
| Haiti | 77 | 7.04 | 1,027 | 93.96 |
| Honduras | 16 | 1.46 | 1,043 | 95.43 |
| Guatemala | 9 | 0.82 | 1,052 | 96.25 |
| Ecuador | 8 | 0.73 | 1,060 | 96.98 |
| Colombia | 7 | 0.64 | 1,067 | 97.62 |
| Brazil | 5 | 0.46 | 1,072 | 98.08 |
| El Salvador | 4 | 0.37 | 1,076 | 98.44 |
| Venezuela | 4 | 0.37 | 1,080 | 98.81 |
| Costa Rica | 3 | 0.27 | 1,083 | 99.09 |
| Dominican Republic | 3 | 0.27 | 1,086 | 99.36 |
| Cuba/Mexico | 2 | 0.18 | 1,088 | 99.54 |
| Jamaica | 2 | 0.18 | 1,090 | 99.73 |
| Puerto Rico | 2 | 0.18 | 1,092 | 99.91 |
| Guatemala/Honduras | 1 | 0.09 | 1,093 | 100.00 |
| TOTAL | 1,093 | 100.00 | | |

DISTRIBUTION OF COUNTRIES OF ORIGIN BY UN MACRO-REGION: AFRICA, 2005–2007

| Country of origin | Frequency | Percentage | Cumulative frequency | Cumulative percentage |
|---|---|---|---|---|
| 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 |

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(*CONTINUED*)

| Country of origin | Frequency | Percentage | Cumulative frequency | Cumulative percentage |
|---|---|---|---|---|
| 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 | Percentage | Cumulative frequency | Cumulative percentage |
|---|---|---|---|---|
| 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 | Percentage | Cumulative frequency | Cumulative percentage |
|---|---|---|---|---|
| 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 | Percentage | Cumulative frequency | Cumulative percentage |
|---|---|---|---|---|
| Papua New Guinea | 1 | 100.00 | 1 | 100.00 |
| TOTAL | 1 | 100.00 | | |

© 2012 The Authors. International Migration © 2012 IOM

14682435, 2015, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2435.2011.00715.x by Michigan State University, Wiley Online Library on [11/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License