

IZA DP No. 3173

Surveying Migrant Households: A Comparison of Census-Based, Snowball, and Intercept Point Surveys

David J. McKenzie
Johan Mistiaen

November 2007

Surveying Migrant Households: A Comparison of Census-Based, Snowball, and Intercept Point Surveys

David J. McKenzie

World Bank and IZA

Johan Mistiaen

World Bank

Discussion Paper No. 3173
November 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Surveying Migrant Households: A Comparison of Census-Based, Snowball, and Intercept Point Surveys^{*}

Few representative surveys of households of migrants exist, limiting our ability to study the effects of international migration on sending families. We report the results of an experiment designed to compare the performance of three alternative survey methods in collecting data from Japanese-Brazilian families, many of whom send migrants to Japan. The three surveys conducted were 1) Households selected randomly from a door-to-door listing using the Brazilian Census to select census blocks; 2) A snowball survey using Nikkei community groups to select the seeds; and 3) An intercept point survey collected at Nikkei community gatherings, ethnic grocery stores, sports clubs, and other locations where family members of migrants are likely to congregate. We analyze how closely well-designed snowball and intercept point surveys can approach the much more expensive census-based method in terms of giving information on the characteristics of migrants, the level of remittances received, and the incidence and determinants of return migration.

JEL Classification: C42, O12

Keywords: migration, surveys, rare elements, intercept-point

Corresponding author:

David McKenzie
MSN MC3-300
The World Bank
1818 H Street NW
Washington, DC 20433
USA
E-mail: dmckenzie@worldbank.org

^{*} The survey on which this research is based was funded by ABN-Amro through the World Bank's Knowledge for Change Program. We thank Ricardo Guedes and the survey team from *Sensus* for their hardwork in implementing the survey, and Alejandrina Salcedo for research assistance. The survey would not have been possible without the collaboration of Yoko Niimi, Claudia Romano, who coordinated activities in Brazil, Kaizô Beltrão and Sonoe Pinheiro who carried out the sampling, and provided valuable expertise and insight on the Nikkei community in Brazil, and Omar Noryo Okino and his team at *Sudameris*, who provided contacts with Nikkei associations. We also thank the Nikkei associations for their support, and especially the survey participants. A preliminary version of this paper was presented at the PAA Meetings 2007. The views expressed here are those of the authors alone, and do not necessarily reflect those of the World Bank Group, its Executive Directors, or the countries they represent.

1. Introduction

The importance of international migration for development has received increasing attention from the research and policy communities (e.g. GCIM, 2005; World Bank 2005), leading to a High-Level Dialogue on International Migration and Development at the United Nations General Assembly in September 2006.¹ The focus of much of this research and discussion has been on examining the impacts of international migration on development in the sending countries, and in identifying policies which can maximize the development benefits of migration. However, very few detailed and representative surveys of households of migrants exist, limiting our ability to study the effects of international migration on sending families.

Public use microdata from national censuses provide representative information, but only for a very limited set of variables. Nationally representative household surveys, such as the World Bank's Living Standards Measurement Surveys provide more information about living standards, education, and other outcomes of interest, but usually relatively little information on the migration process. As a result, answering many questions of interest in the study of migration requires specialized surveys. However, most of these specialized surveys are non-probability samples of unknown representativeness, making it hard to generalize any conclusions reached from them. As Fawcett and Arnold (1987) note, common approaches used by many studies are to choose their samples from individuals who belong to church groups, social organizations, or other defined groups; use snowball samples of individuals referred by friends or acquaintances; and/or to focus exclusively on areas of high out-migration.

This paper reports on the results of an experiment designed to compare the performance of three methodologies for sampling households with migrants: i) a stratified sample using the census to randomly sample census tracts, in which each household is then listed, and screened to determine whether or not it has a migrant, with the full length questionnaire then being applied in a second phase only to the households of interest; ii) a snowball survey in which households are asked to provide referrals to other households

¹ <http://www.un.org/esa/population/hldmigration/> [accessed February 10, 2007].

with migrant members; and iii) an intercept point survey (or time and space sampling survey), in which individuals are sampled during set time periods at a pre-specified set of locations where households in the target group are likely to congregate.

We apply these methods in the context of a survey of Brazilians of Japanese descent (Nikkei). There are approximately 1.2-1.9 million Nikkei amongst Brazil's 170 million population. Many of these Nikkei have migrated to Japan to work after a Japanese law change in 1990 allowed third-generation Nikkei unrestricted access to Japanese labor markets (Tsuda, 1999, Higuchi, 2006). The estimated 265,000 migrants send approximately \$US2 billion in annual remittances (Beltrão and Sugahara, 2006). We compare the performance of the three different survey methods in collecting data from Nikkei households in Brazil with and without migrants in Japan.

Whilst our application involves surveying Nikkei households (an ethnic minority), with and without migrants abroad, the methodologies employed are equally applicable to attempts to survey migrants in their destination countries. More generally, the problem of surveying migrant households is one of surveying "rare elements" (Kish, 1965, Kalton and Anderson, 1986). The results of the survey experiment are therefore also informative for surveys of other rare populations, such as ethnic minorities and the homeless.

The remainder of the paper is structured as follows. Section 2 outlines the different methodologies which have been developed and used in previous studies to survey migrants and their families. Section 3 describes our experiment and the Brazilian setting, while Section 4 describes how the different methodologies were applied in practice. Section 5 compares the results of the three survey methods, and Section 6 provides a cost comparison across the different methods. Section 7 concludes.

2. Different Methods Used for Sampling Migrants, Families of Migrants, and Other Rare Elements

Any attempt to carry out a specialized survey of migrants or of migrant-sending households must face the problem that international migration is a relatively rare event in most countries. Bilsborrow et al. (1997) note that in three-quarters of the countries in the world, the proportion of international migrants was at most 6.5 percent in the early 1990s. Even in countries in which international migration is more common, finding a household with a migrant currently abroad or a recently returned migrant can be a rare event. Therefore carrying out a survey of migrant-sending households is essentially a problem of surveying “rare elements” or “rare populations” (Kish, 1965, Kalton and Anderson, 1986). Our application fits well this description: it is estimated that there are approximately 1.4 million Nikkei households in Brazil, relative to an overall population of over 170 million.

Conducting a probabilistic sample of a rare population presents no problem if a full sample frame is available. Representative samples of legal migrants have thus been recently conducted using administrative records on new immigrants. Examples include the New Immigrant Survey (NIS) in the United States, the Longitudinal Survey of Immigrants to Australia (LSIA) and the Longitudinal Immigration Survey (LISNZ) in New Zealand. This is more difficult to carry out for migrant-sending households, as it requires obtaining from migrants and migrant records the contact details for the remaining household unit. The only application we are aware of which does this is the Pacific Island-New Zealand Migration Survey (McKenzie, Gibson and Stillman, 2006), which links new Tongan migrants in New Zealand to their remaining households in Tonga, and surveys the sending households in Tonga.

The much more common situation is one in which no survey frame is available. Three approaches to sampling rare elements have then been most commonly used in practice to survey migrant-sending households or migrants.² These are stratified sampling using

² Note that surveys of households with migrants will also want to include some households without migrants for comparison purposes. In many circumstances these households are not rare elements, and a

disproportionate sampling fractions with two-phase sampling; snowball sampling; and time and space sampling, also known as intercept point sampling, location sampling, or aggregation point sampling. We discuss each in turn. Other sampling strategies which have been used have been convenience sampling, and identifying migrants through surnames in the telephone book. For example, Osili (2006) used the Chicago phone directory and identified names of the Igbo of South Eastern Nigeria to sample Nigerian migrants in the U.S. We will also discuss the coverage of such a method in our application.

The use of stratified sampling with disproportionate sampling fractions is the approach recommended by Bilsborrow et al. (1997) in their guidelines for improving international migration statistics. They note that most countries have population census data or population registers which can be used to estimate populations and the numbers of international migrants. They therefore recommend using the census to select provinces, districts, and if possible, census sectors, with probability proportional to the number of households with migrants. After census sectors are selected, a two-phase sampling strategy can be used, in which a screening phase is first carried out to identify the respondents of special interest, and then the full questionnaire is administered in a second phase to a sample of households identified in the first phase.

In theory, this approach has the advantage of providing a representative sample of households with and without migrants. It has been used in the NIDI/Eurostat surveys in Egypt, Ghana, Morocco, Senegal and Turkey. In most of these applications, surveying is first restricted to certain provinces or districts where migrants are thought to come from, in order to reduce survey costs. For example in Ghana the survey chose 17 electoral districts, and screened 21,504 households according to household migration status, in order to arrive at a target sample of 1,980 households. 1571 households were then interviewed in the second phase (Groenewold and Bilsborrow, 2004). The disadvantage of this method is that it can be expensive and time-consuming to screen a large number of

sufficient number of such households will be identified using the methods described here to find the rare elements. In our context, it is Nikkei households, both with and without migrants, that are the rare elements.

households in order to identify households with migrants. Fawcett and Arnold (1987) note also that non-response can be a major problem in immigration surveys, particularly in urban areas. They point out that while individuals usually have a legal obligation to answer questions in the census, surveys generally carry no legal sanctions for refusal to respond. In addition, in urban areas, immigrants who work often work long hours, making it difficult to find them at home, while undocumented immigrants may be reluctant to take part in a survey for fear of being found by government authorities.

A second method commonly used to sample rare populations is the chain-referral method, in which an initial sample of individuals is taken, and each of these is asked to provide referrals to other individuals in the population of interest. Snowball sampling (Goodman, 1961) and respondent-driven sampling (Heckathorn, 1997) are the most common examples. In snowball sampling, each individual in the sample is asked to name k different individuals, and each of these is then asked to name k different individuals, and so on. Snowball sampling has been used by the Mexican Migration Project to sample permanent Mexican migrants in the United States (Massey and Singer, 1987), and was used in part by the NIDI/Eurostat survey to survey immigrants in Spain (Groenewold and Bilsborrow, 2004).

A necessary condition for successful application of snowballing is that members of a rare population know each other (Kalton and Anderson, 1986). Such an approach is likely to hold for ethnic minorities, making it appropriate for sampling migrants at destination, and in our case, sampling a rare ethnic group in Brazil. Moreover, recent work by Heckathorn (1997, 2002) has shown that it is possible to obtain a representative sample through chain referral methods, based on the idea of “six degrees of separation”, in which each person in a population is linked to each other person through six intermediaries on average. However, applying this in practice requires that the chain referrals be long, and that adjustments are made for the fact that subjects with larger personal networks are more likely to be oversampled. Other problems which can arise in practice is that the subjects may not refer friends in order to protect their privacy, and that contact information is frequently inadequate, so attrition rates can be high. For example, Bilsborrow (2006)

reports that in a 2006 survey of Colombian migrants in Ecuador, the snowballing procedure worked poorly, with less than one referral obtained per four interviewed households.

The third method used to sample immigrants or ethnic minorities makes use of the fact that immigrants often cluster at certain locations. Simple examples of this type of sampling carried out sampling at only one type of location. Examples include surveying Mexicans at border crossing points in the Encuesta sobre Migración en la Frontera Norte (EMIF) (Bustamente et al., 1997), and surveying Latina immigrant women at churches in the U.S. (Wasserman et al. 2005). However, by sampling at only one type of location, the survey is likely to miss many migrants. Better coverage of the population of interest can be achieved by surveying at multiple locations. An issue which arises here is that individuals can potentially be surveyed more than once, so the survey needs to account for multiple selection possibilities during analysis.

Sampling theory for multiple location samples are provided in Kalsbeek (1986) and Kalton (1991, 2001). The basic survey design involves sampling in both space and time. Primary sampling units are constructed as combinations of locations and time segments where surveying will take place at the location. Then some form of systematic sample is employed to select individuals visiting the location during the specified time period. Such an approach has been used to survey other rare populations, such as visitors to soup kitchens, African nomadic populations by surveying at watering holes, and homosexual men, by surveying at bars, dance clubs and street locations. Blangiardo (1993, cited in Groenewold and Bilsborrow, 2004) proposes a similar methodology for sampling migrants, which was used in the NIDI/Eurostat survey of Ghanaian and Egyptian immigrants in Italy. A listing of popular places, called aggregation points, where migrants tend to meet (such as mosques, health care facilities, telephone calling centers, shelters, and public squares) is made. At each location migrants surveyed are asked how often they visited any of the other aggregation points, allowing *ex-post* selection probabilities to be calculated for each individual surveyed.

Intercepting migrants or rare elements in public places provides a cost-efficient method of surveying, and may allow surveying of individuals who are seldom found in their homes. By ex-post weighting of the sample, one can obtain a sample representative of any person in the reference group who has visited at least one of the locations during the sample period. This method is appealing in that it is likely to offer a sample which is more representative of the underlying population of interest than can be found through the first few referral chains of a snowball sample, with less time and cost than a census-based screening and listing exercise. However, a disadvantage of interviewing in public locations is that individuals will generally have less time to answer the survey than during a home visit. As a result, on location surveys of this type will have to use a much shorter questionnaire, thereby collecting less extensive data on the population of interest.³

A fourth method occasionally used to sample ethnic minorities, including migrants, is to use the telephone book to select individuals with minority names. In our application, this would involve using the Brazilian phone books to select individuals with Japanese surnames. Such an approach presents several potential problems, which we can measure the magnitudes of by identifying which of the households found in the stratified random survey also appear in the phone book with a Japanese surname. The first potential problem is that households may not have a landline telephone. This is likely to be more of an issue with poorer households, and as cellphones increasingly substitute for landlines. 90.1 percent of Nikkei households in our stratified sample have a landline phone – of those without a landline, 79.2 percent have a cellphone. The second potential issue is that not all members of the ethnic minority will have a minority surname, especially with intermarriage. 74.8 percent of Nikkei households have a recognizably Japanese surname. A final problem is that households may opt-out of having their telephone number listed in the phone book. Only 24.1 percent of Nikkei households in our sample appear in the phonebook, and only 20.1 percent appear in the phone book and

³ In some circumstances one may be able to use the intercept location survey to construct a sampling frame of migrants along with contact details, and subsequently follow-up with longer questionnaires at home or in another location. However, it appears likely that many people approached on the street will refuse to provide follow-up contact details, particularly if they are concerned about crime, or are of illegal migration status.

have a Nikkei name. Thus such a method would only cover one-fifth of the target population in our application.

3. The Experiment

Each of the three main methods of sampling migrants or migrant-sending households has its theoretical advantages and disadvantages in terms of cost, time, coverage, and representativeness. However, comparing the practical performance of the three methods is made difficult by the fact that they have all been used in different country contexts, at different times, with different questionnaires and survey teams. Nevertheless, knowing how the different methods perform in practice is a question of large importance for the design of new surveys of migrants or migrant-sending households. We therefore designed an experiment to compare how the three main methods perform in practice. In particular, we compare a census-based stratified random sample, an intercept survey, and a snowball survey.

The context of our experiment is a survey that the World Bank was requested to perform of the Japanese-Brazilian population (Nikkeis) in Brazil. Japanese migration to Brazil began in 1908 with a ship carrying bonded labor to the coffee plantations (Goto, 2006). High rates of migration from Japan to Brazil occurred from 1925-36 as the Japanese government subsidized emigration, and again from 1955-1961 as the Japanese government again promoted emigration during post-war rebuilding. Many of these workers settled in Brazil, and the population of Japanese descent in Brazil was estimated to have reached 1.2 million by 1987-88 (Tsuda, 2003). Following a revision of Japanese immigration law in 1990, many of these Nikkei began migrating back to Japan to work: In 2004 there were 190,000-265,000 Brazilians in Japan, who were estimated to be sending US\$2 billion in remittances back to Brazil (Beltrão and Sugahara, 2006).

The Japanese-Brazilian population is now relatively concentrated in the upper part of Brazil's income distribution, and work in relatively skilled occupations. They migrate to undertake less skilled jobs in Japan than they were doing in Brazil, but for higher wages. McKenzie and Salcedo (2007) use the data from the census-based survey described here

to compare Japanese-Brazilians to the average Brazilian in the PNAD national surveys. A Nikkei worker in the fifth decile of the Nikkei wage distribution would be in the seventh decile of the Sao Paulo-Parana distribution and eighth decile of the overall Brazilian wage distribution. Mean wages of Nikkei wage workers are 1440 reais per month, approximately \$US720.

The survey was designed to provide detail on the characteristics of households with and without migrants, estimate the proportion of households receiving remittances and with migrants in Japan, and examine the consequences of migration and remittances on the sending households. We compare the performance of the three different survey methods in meeting these objectives. The same questionnaire was used for the stratified random sample and snowball surveys, and a shorter version of the questionnaire was used for the intercept surveys. Therefore we can directly compare answers to the same questions across survey methodologies, and determine the extent to which the intercept and snowball surveys are able to give similar results to the more expensive census-based survey, and test for the presence of the types of biases one might expect. For example, we would expect individuals who belong to Nikkei community organizations to have a greater connection to Japan, and therefore to be more likely to migrate. Nikkei who are more integrated into Brazilian society may be harder to observe through snowball and intercept surveys. We will compare across the three surveys the characteristics of migrant sending households, the likelihood of receiving remittances and level of remittances received, and the incidence of return migration.

Several characteristics of the Nikkei population in Brazil present a challenge for surveying. Firstly, the population is predominantly urban, with many living in high-rise apartments secured by building managers or doormen. With crime a general concern in urban Brazil, some building managers are reluctant to allow entry into apartment buildings. Moreover, as is common in urban areas, most individuals work outside of their homes, and many are reluctant to be interpreted at home outside of working hours. Secondly, the Nikkei population in Brazil share the characteristic of many ethnic minorities and migrant groups of being suspicious of outsiders. Furthermore, there have

been incidences of Nikkei returning from working in Japan being targeted for crime. These characteristics are shared by many other migrant groups of interest, such as undocumented migrants and migrants from other urban areas, making this case study an application similar to many other practical applications of interest. In common with common practice in surveys of migrants elsewhere, we made an effort to gain the trust and support of the local community. This was done through communications with Nikkei associations, collaboration with the representatives of the bank *Sudameris* who deal with the Nikkei community, and the use, where possible, of Nikkei interviewers.

4. Implementation of the Three Sampling Methods

This section discusses in detail how the stratified random sample survey, intercept survey, and snowball survey were implemented. All three surveys were implemented by the same survey firm, *Sensus Data World*, an experienced Brazilian survey firm, and were carried out at the same point in time, allowing comparability between the three methods. The same questionnaire was used for both the stratified random sample and snowball surveys, while a much shorter questionnaire with a subset of the questions was used for the intercept survey.

4.1 Stratified Random Sample of Nikkei Households in Sao Paulo and Parana

Brazil's population in the 2000 Census was 169.8 million. However, it is estimated that 80 percent of the Nikkei population lives in just two states: 54 percent in the state of Sao Paulo (population 37.0 million), and 26 percent in Parana state (population 9.6 million).⁴ We therefore decided to only survey these two states, which combined have a population approaching 50 million people. The sampling process then consisted of three stages. First, a stratified random sample of 75 census tracts was selected. Second, interviewers carried out a door-to-door listing within each census tract in order to determine which households had a Nikkei member. Third, the survey questionnaire was then administered to households identified as Nikkei. We now describe the details of each step.

⁴ Population numbers from the 2000 Census are taken from <http://www.ibge.gov.br/english/estatistica/populacao/censo2000/> [accessed February 8, 2007].

*Selection of Census Tracts*⁵

The 2000 Brazilian Census was used to classify households as Nikkei or non-Nikkei. The Brazilian Census does not ask ethnicity, but instead asks questions on race, country of birth, and whether an individual has lived elsewhere in the last 10 years. Based on these questions, a household is classified as (potentially) Nikkei if it has any of the following:

- a) A member born in Japan
- b) A member who is of yellow race, and who has lived in Japan in the last 10 years.
- c) A member who is of yellow race, who was not born in a country other than Japan (predominantly Korea, Taiwan or China), and who did not live in a foreign country other than Japan in the last 10 years.

This procedure provides an approximate estimate of the number of Nikkei households, but will tend to be an overstatement due to misclassifying as Nikkei households comprising of individuals of Korean, Taiwanese or Chinese ethnicity who were all born in Brazil and hadn't been in those countries in the last 10 years.⁶

Table 1 tabulates the number of yellow race immigrants in Brazil in the 1980, 1991 and 2000 Censuses by country of birth. Individuals born in Japan are the second largest immigrant group in Brazil after the Portuguese, accounting for 11 percent of all immigrants in 2000. The number born in Japan has been falling over the last twenty-five years. In 2000, Japanese-born still accounted for 74 percent of all yellow race immigrants, Chinese 11 percent, Koreans 9 percent, and Taiwanese 5 percent. We classify as non-Nikkei yellow race individuals born in other countries, so our concern is with second or latter generation non-Nikkei Asians. If these generations occur in the same proportions as first generation, this would suggest we are overestimating the number of Nikkei by at most 35 percent. However, as Table 1 shows, Japanese were a greater share of yellow race immigrants in the 1980 and 1991 Censuses. This is illustrated further in Figure 1, which uses the 2000 Census to plot the mean year of

⁵ IBGE statisticians Kaizô Beltrão and Sonoe Pinheiro carried out the selection of the sample in consultation with the authors.

⁶ We will also misclassify as non-Nikkei Nikkei individuals who are married to Chinese or Koreans. Inter-marriage is very low between Japanese and other Asian groups, so this will not induce much error.

arrival in Brazil and mean age of selected yellow race foreign-born. The mean year of arrival is much earlier for Japanese than other races, meaning that a larger proportion of their ethnicity should be second or later generation. Thus our overstatement from misclassifying on race should be considerably less than 35 percent.

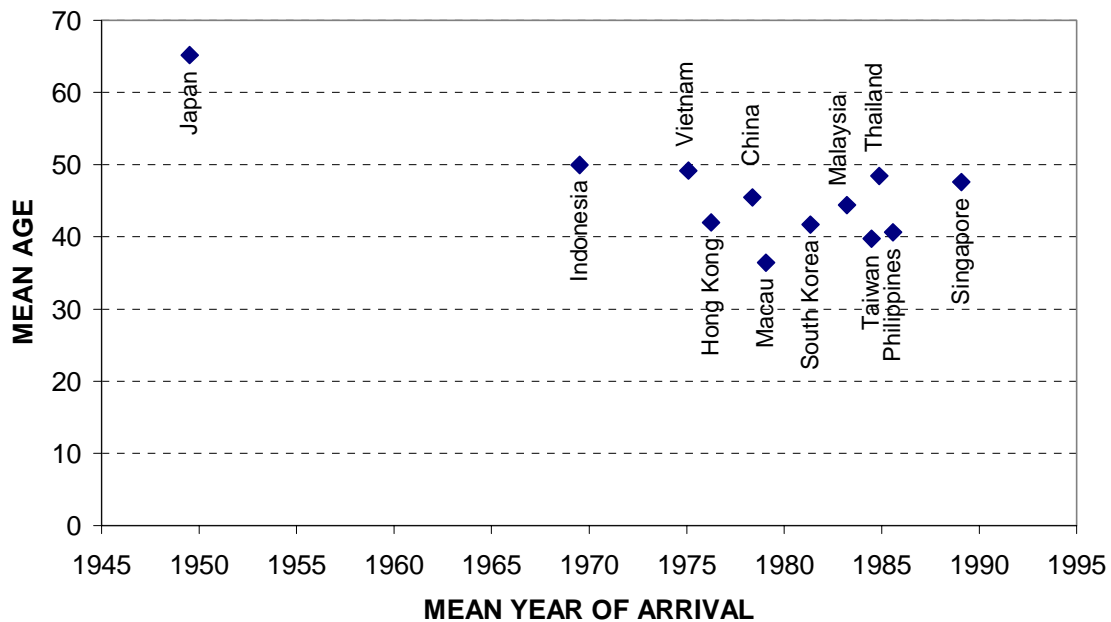
The 2000 Census was then used to estimate the number of Nikkei in each municipality, área de ponderação, and census tract. An área de ponderação (AP) is the smallest geographical unit used for public reporting of the results of the Census, and consists of a grouping of census tracts. There are 1913 APs in Sao Paulo state and 596 in Parana state. A second source of estimation error occurs from the fact that questions on race, birthplace, and migration are only asked on the long form of the Census questionnaire, which is applied to only 10 percent of households in municipalities with more than 15,000 inhabitants, and to 20 percent of households in municipalities with less than 15,000. Therefore an additional source of prediction error arises from this sampling. This sampling error will be small at the level of a municipality and AP, but will be greater at the level of the census tract. An Appendix to the paper provides an example modeled on our survey, where the overestimation of the number of Nikkei households at the census tract level is 50 percent.

Table 1: Yellow Race Immigrants in Brazil

| Country of Birth | Number of foreign-born by country of birth | | | Shares in 2000 Census | |
|------------------|--|-------|-------|---------------------------|------------------------|
| | 1980 | 1991 | 2000 | Percent of all immigrants | Percent of Yellow Race |
| Japan | 139480 | 85572 | 70932 | 10.37 | 73.9 |
| China | 8799 | 8324 | 10301 | 1.51 | 10.7 |
| South Korea | 7258 | 8528 | 8578 | 1.25 | 8.9 |
| Taiwan | 2414 | 2737 | 4536 | 0.66 | 4.7 |
| Indonesia | | | 693 | 0.10 | 0.72 |
| Hong Kong | | | 376 | 0.05 | 0.39 |
| Philippines | | | 360 | 0.05 | 0.38 |
| North Korea | | | 66 | 0.01 | 0.07 |
| Malaysia | | | 60 | 0.009 | 0.06 |
| Macau | | | 27 | 0.004 | 0.03 |

Source: 1980, 1991 and 2000 Brazilian Census

FIGURE 1: MEAN YEAR OF ARRIVAL IN BRAZIL AND MEAN AGE AMONG YELLOW RACE FOREIGN-BORN IN BRAZIL (2000 CENSUS)



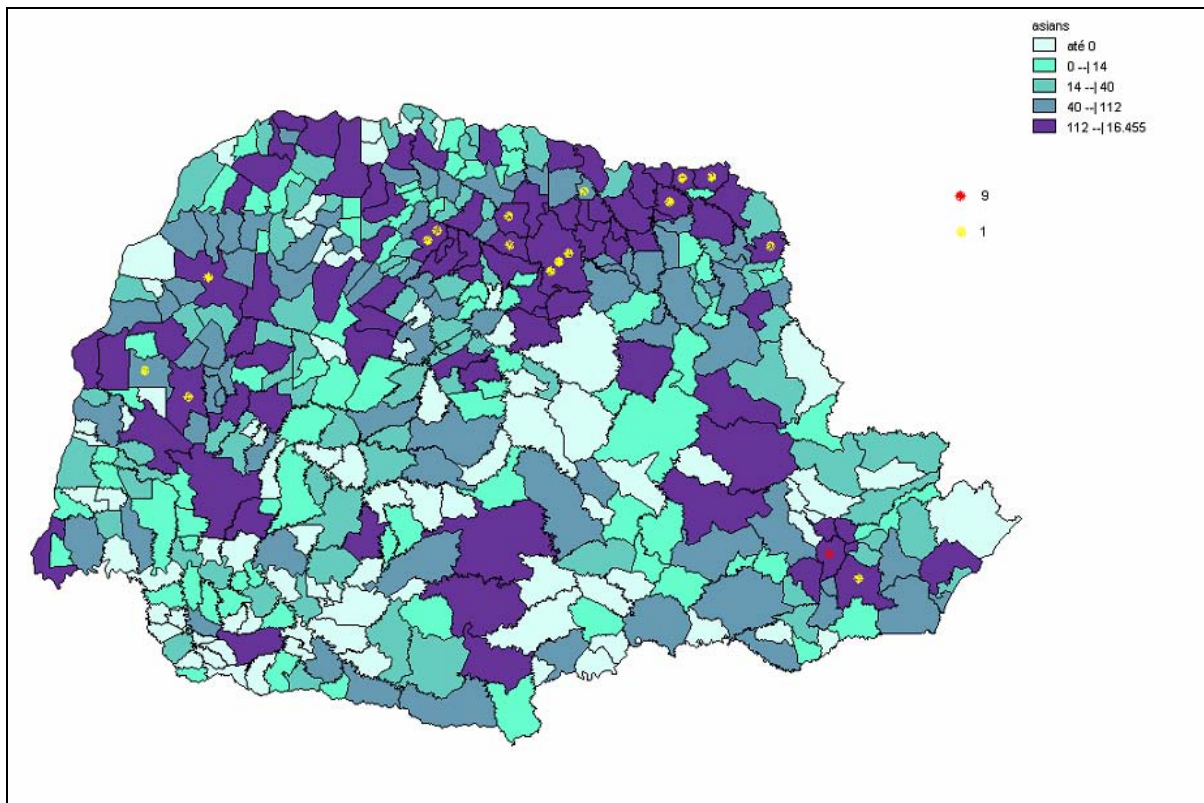
Source: 2000 Brazilian Census

These Nikkei estimates were then used to select 50 census tracts in Sao Paulo state and 25 census tracts in Parana state as follows. First, municipalities were randomly selected according to probability proportional to size (PPS) sampling with replacement, where size is the number of Nikkei households. Secondly, within each municipality selected, APs were sampled with PPS. Then finally, census tracts were sampled with PPS within the APs. In order to ensure coverage of both census tracts with high concentrations of Nikkei, and lower concentrations, we stratified so that in Sao Paulo, 30 out of the 50 census tracts were selected from among census tracts estimated to have 15 or more Nikkei households living in them, and 20 census tracts were estimated to have 4-15 Nikkei households living in them. In Parana, 15 out of the 25 census tracts were chosen from those with 15 or more Nikkei households, and the remaining 10 census tracts chosen from those with 4-15 Nikkei households. We did not include census tracts with 3 or fewer estimated Nikkei, as they are estimated to cover only 1 to 3 percent of the Nikkei population in the two states, and listing such census tracts would increase the survey cost with little additional increase in sample. Survey weights which take account

of the different probabilities of census tracts being sampled will be used in all the analysis.

Figure 2 graphically illustrates the census tracts selected in Parana state. Municipalities are shaded according to the estimated number of Nikkei, with darker shading indicating more Nikkei. The map shows the spatial concentration of Nikkei in certain regions of the state. Nine of the 25 census tracts selected are in one municipality, Curitiba, the state capital with population 1.6 million. The other 16 census tracts are spread over 13 other municipalities, and cover also a couple of municipalities with less than 112 estimated Nikkei households.

Figure 2 – Distribution of Asians (excluding Chinese, Korean and Taiwanese Nationals) by municipality and census tracts in the sample (dots) – Paraná - 2000



Source: Map prepared by Kaizô Beltrão and Sonoe Pinheiro using 2000 Census data.

Listing

Prior to the commencement of the listing operation, letters were sent to approximately 150 Nikkei associations with bases in the areas chosen, explaining the purpose of the survey, asking them to encourage their members to answer the survey, and providing a telephone number for any enquiries. A door-to-door listing exercise of the 75 census tracts was then carried out between 13 October 2006 and 29 October 2006. A census tract averaged 301 housing units. Listing used 42 interviewers in Sao Paulo and 24 interviewers in Parana. Interviewers went to each housing unit with a screening questionnaire, which asked whether or not the household had any members who were Nikkei, or Nikkei members currently in Japan. Households with Nikkei were then asked whether they had members who had returned from Japan, whether they had members currently in Japan, and whether they had any members who were third or fourth generation Japanese. Three attempts were made to interview the household in the event that the first or second attempt yielded nobody at home. In the event that an interview could not be made due to refusal, no one at home, or the refusal of apartment building management to allow the survey, the Nikkei status of households was obtained through proxy-reporting from a neighbor or building manager.

Table 2 summarizes the results of the listing process. The listing covered 14,239 dwelling units in Sao Paulo state and 8,300 units in Parana state, for a total of 22,539 dwellings.⁷ This was 21 percent more dwelling units than recorded for these census tracts in the 2000 Census, showing the extent of population growth and new construction over the six years since the Census. Among these 22,539 the listing detected 839 Nikkei households, 528 of which were interviewed in person, and 311 obtained by proxy-reporting. Proxy-reporting was more common in Sao Paulo state, particularly in Sao Paulo city, where household members were harder to find at home. Thus 3.7 percent of the dwelling units listed contained Nikkeis.

⁷ The listing exercise also listed empty dwellings, second-homes, and non-residential buildings so the total number of buildings/dwellings listed was 26,697.

The census tracts listed show a great deal of variation in the number of Nikkei. The mean size of a census tract was 301 households. The mean (median) number of Nikkei households in a census tract was 11 (8). Fifty-nine of the 75 census tracts each had less than 15 Nikkei, including three census tracts with no Nikkei households (with 758 households between them). Two census tracts had more than 50 Nikkei households, one with 58 and the other with 92.

TABLE 2: LISTING OF HOUSEHOLDS IN SAO PAULO AND PARANA STATES

| | Sao Paulo State | Sao Paulo City | Parana State | Curitiba City | Combined Sample |
|---|-----------------|----------------|--------------|---------------|-----------------|
| Number of Municipios surveyed | 28 | 1 | 14 | 1 | 42 |
| Number of Census tracts surveyed | 50 | 16 | 25 | 9 | 75 |
| Average Size of Census tract | 285 | 314 | 332 | 381 | 301 |
| Number of residential units listed | 14239 | 5025 | 8300 | 3425 | 22539 |
| Number of nikkei households listed | 559 | 206 | 280 | 78 | 839 |
| Number of nikkei where household member interviewed | 305 | 86 | 223 | 58 | 528 |
| Percent of household listed that are nikkei | 3.9 | 4.1 | 3.4 | 2.3 | 3.7 |
| Percentage of nikkei households where: | | | | | |
| Interview obtained | 54.6 | 41.8 | 79.6 | 74.4 | 62.9 |
| Interview refused | 18.3 | 13.6 | 7.5 | 18.0 | 14.7 |
| Family was travelling | 2.7 | 1.5 | 1.1 | 0.0 | 2.2 |
| No one was home during three visits | 24.5 | 43.2 | 11.8 | 7.7 | 20.3 |
| <i>Number of households in 2000 Census</i> | 11886 | 3902 | 6698 | 2294 | 18584 |
| <i>Predicted number of Nikkei households in 2000 Census</i> | 1209 | 395 | 532 | 215 | 1741 |
| <i>Listed households/Census households</i> | 1.20 | 1.29 | 1.24 | 1.49 | 1.21 |
| <i>Listed Nikkei/Census predicted Nikkei</i> | 0.46 | 0.52 | 0.53 | 0.36 | 0.48 |

The bottom of Table 2 shows that the number of Nikkei listed was only 48 percent of the number of Nikkei predicted on the basis of the 2000 Census, despite the overall number of households growing 21 percent. There are three main reasons for this difference. First, part of the difference reflects the misclassification of Chinese, Korean and Taiwanese households as Nikkei in predicting the number of Nikkei in the Census. Secondly, given that six years had passed between the Census and our survey, the difference could also partly reflect population dispersion, if Nikkei households are moving out of the more traditional neighborhoods over time. Finally, part of the difference could also be due to the long form of the Census being used only for 10 percent of households, and thus to sampling error in predicting the number of Nikkei in a census tract from the 10 percent sample.

Administration of the Household Survey

Once a list of Nikkei households had been obtained, the final stage of the survey carried out an in-person survey of Nikkei households. Our initial budget planned on surveying 900 households, and so we intended to carry out a stratified sample of the Nikkei households obtained through the listing exercise. However, since only 839 Nikkei households were obtained via listing, all listed Nikkei households were selected for the full survey. Fieldwork began November 19, 2006, and all dwellings were visited at least once by December 22, 2006. Some of the households identified by proxy-reported as being Nikkei were found to be non-Nikkei during this process, reducing the target sample to 710 Nikkei households.

During the initial wave of surveying we were successful in interviewing 247 Nikkei households, 109 in Sao Paulo state and 138 in Parana. The households which were not able to be interviewed during this initial phase were households where no one was home at the time of the survey visit, where the building manager refused access to the building, or where the household refused to answer the survey. A second wave of surveying then took place from January 18, 2007 to February 2, 2007, intended to increase the number of households responding. We made a number of changes to the survey protocol in order to attempt to get a response from households not interviewed in the first wave:

- a) Meetings were held with the presidents of several of the most important Nikkei associations in Sao Paulo city and Curitiba to ask for their direct support. The contacted associations agreed to do this, and provided phone numbers and names which could be used in a letter presented by the interviewer, so that the interview subject could call with any questions about the veracity of the survey. Similarly, additional local contact details were provided for the World Bank, which could again be used by interview subjects to verify the survey was legitimate.
- b) The initial round of interviewing used Brazilian interviewers who were not Nikkei, due to difficulties hiring Nikkei who were interested in carrying out survey work. More intensive efforts were undertaken to find Nikkei workers, allowing Nikkei field workers to be used in this second wave.

- c) Prizes were used to try and increase the incentive to participate. Interview subjects were told that a random drawing would be done amongst completed interviews, with the winners receiving Video iPods.
- d) Finally, if subjects still refused to answer the questionnaire, interviewers would leave a much shorter version of the questionnaire to be completed by the household by themselves, and later picked up. This shorter questionnaire was the same as used in the Intercept point survey, taking 7 minutes on average. The intention with the shorter survey was to provide some data on households that would not answer the full survey due to time constraints, or to them being reluctant to have an interviewer in their house.

This strategy was very successful, yielding a further 45 full questionnaires and 111 short questionnaires. Table 3 summarizes the final results of this survey process. In total, we were able to survey 403 out of the 710 Nikkei households, a 57 percent interview rate. The refusal rate was 25 percent, whilst the remaining households were either absent on three attempts, or were not surveyed due to building managers refusing permission to enter the apartment buildings. Refusal rates were higher in Sao Paulo than in Parana, reflecting greater concerns about crime and a busier urban environment.

Table 3: Final Survey Results of the Stratified Random Sample

| | Sao Paulo | | Parana | | Combined Sample | |
|-------------------------------------|-----------|----|--------|----|-----------------|----|
| | Number | % | Number | % | Number | % |
| <i>Dwellings Screened</i> | 14239 | | 8300 | | 22539 | |
| <i>Nikkei households identified</i> | 450 | | 260 | | 710 | |
| Interviewed | 204 | 45 | 199 | 77 | 403 | 57 |
| Refusals | 145 | 32 | 30 | 12 | 175 | 25 |
| Not allowed to enter building | 31 | 7 | 5 | 2 | 36 | 5 |
| Absent during at least 3 visits | 70 | 16 | 26 | 10 | 96 | 14 |

Table 4 compares the characteristics of households which were surveyed in the first round of survey efforts with households surveyed in the second round of survey efforts who had initially refused or who were not able to be contacted during three attempts in the first round of the survey. The group of households surveyed in the second round is similar in many respects to those surveyed in the first round. The main difference lies in

the percentage of households receiving remittances who refuse to say how much they receive. In Sao Paulo, this is 19 percent for the first round, compared to 67 percent for the second round. This accords with the reports from our survey team that the main reason for refusal was concern about crime and reluctance to discuss financial matters. The extra survey effort was able to convince these households to provide a lot of other important information in the survey, including whether they receive remittances, even though most refuse to say how much they receive.

Table 4: Comparing Households Surveyed in First Round to those who initially refused or were not initially found

| | Sao Paulo round 1 | Sao Paulo round 2 | Parana round 1 | Parana round 2 |
|---|----------------------|----------------------|-------------------|-------------------|
| Household Size | 3.36 | 3.09 | 3.31 | 3.33 |
| <i>Percentage of households with member who:</i> | | | | |
| Reads Japanese/Nikkei newspapers | 13 | 14 | 10 | 15 |
| Listens to Japanese/Nikkei radio programs | 7 | 10 | 8 | 5 |
| Watches Japanese/Nikkei TV programs | 19 | 29 | 27 | 19 |
| Reads Japanese/Nikkei books/magazines | 14 | 16 | 14 | 18 |
| Reads newspapers from Nikkei associations | 12 | 13 | 8 | 9 |
| Checks Japanese/Nikkei websites on the internet | 3 | 8 | 9 | 7 |
| <i>% of households which:</i> | | | | |
| Migration | | | | |
| Have a Member currently in Japan | 21 | 17 | 19 | 25 |
| Have a Member who has returned from work/study in Japan | 38 | 45 | 31 | 51** |
| Remittances | | | | |
| Receive remittances from Japan | 17 | 8 | 14 | 22 |
| Refuse to say if they receive remittances | 2 | 4 | 4 | 6 |
| Refuse to say how much they receive if receiving | 19 | 67** | 47 | 79* |
| Mean amount conditional on receipt and report | 5704 | 3223 | 3302 | 7164 |
| Sample Size | 109 | 95 | 138 | 61 |

*, **, *** indicate significantly different from the 1st round at the 10%, 5% and 1% levels respectively.

4.2 The Intercept Point Surveys

The Intercept survey was designed to carry out interviews at a range of locations frequented by the Nikkei population. It was originally designed to be done in Sao Paulo city only, but a second intercept point survey was later carried out in Curitiba, in Parana.⁸

⁸ The smaller than anticipated sample size from the snowball survey and stratified sample provided the budget resources to finance the additional intercept survey in Curitiba. In addition, an intercept survey of

We designed a short version of the questionnaire to apply at these locations. The questionnaire was four pages in length, consisted of 62 questions, and took a mean time of 7 minutes to answer. Respondents had to be 18 years or older to be interviewed. Interviewing for the Sao Paulo intercept survey took place between December 9, 2006 and December 20, 2006, while the Curitiba intercept survey took place between March 3 and March 12, 2007.

Consultations with Nikkei community organizations, local researchers, and officers of the bank *Sudameris*, which provides remittance services to this community, were used to select a broad range of locations. In Sao Paulo, we chose 9 fixed point locations and 6 events. The 9 fixed locations are: a sports club, a metro station in the Liberdade neighborhood, two *Feiras* (Sunday open markets), a hospital focused on the Nikkei community, two grocery stores specializing in Japanese foods, a Japanese cultural society which offers language classes and evening events, and outside a branch of the Banco Sudameris in the Saúde neighborhood. The 6 events were: an afternoon Japanese film event organized by the Sociedade Brasileira de Cultura Japonesa, a large cultural festival with music, dancing and *taiko*-drumming organized by ACAL (Associação Comercial e Assistencial da Liberdade), a Japanese food festival organized by ACESA (Associação Cultural Esportiva de Santana), a Japanese art exposition organized by Fundação Mokiti Okada, a Christmas concert organized by Coral do Bunkyo, Paineiras e Silver Boys, and a music festival organized by Grupo The Friends.

The Nikkei community is less concentrated in Curitiba, with fewer public places where Nikkei are known to gather. Five fixed point locations were chosen: the municipal market, Clube Nikkei (a sports club), the Bunkyo (a Japanese cultural society), a Japanese language school associated with the Bunkyo, and a second Nikkei association, the Associação Brasileira de Dekasseguis (ABD). The intercept survey was also to include surveying at an event at the Seicho-no-ie church, but this event was canceled.

100 Yonsei (fourth generation Nikkei) was carried out in Sao Paulo. We will not discuss this Yonsei survey in this paper.

Interviewers were assigned to visit each location during pre-specified blocks of time. Two field-workers were assigned to each location. One fieldworker carried out the interviews, while the other carried out a count of the number of people with Nikkei appearance who appeared to be 18 years or older who passed by each location. For the fixed places, this count was made throughout the pre-specified time block. For example, between 2:30pm and 3:30pm at the sports club, the interviewer counted 57 adult Nikkeis. Refusal rates were carefully recorded, along with the sex and approximate age of the person refusing. A note was made of the number of individuals who were asked to answer the questionnaire because they appeared Nikkei, but who replied they were not Nikkei. The proportion of falsely identified Nikkei was used to adjust the count taken by the fieldworker to obtain an estimate of the number of Nikkei passing the intercept location.

In the case of intercept surveys carried out at events, a possible concern was that the same person might circle past the location multiple times, thereby invalidating the count. Therefore the fieldworker instead counted the total number of individuals passing during a 10-minute period, and the number of Nikkei adults passing during this period. Estimates of the total number attending the event were obtained from the event organizers, and adjusted by the sample proportion observed to be adult Nikkei to get an estimate of the number of adult Nikkei attending the event.

Table 5 lists the sample size collected, number of refusals, time spent sampling, and approximate number of Nikkei at each sampling location for the Sao Paulo intercept survey. A target of 34 completed interviews was set for each location, in order to make sure the sample wasn't too heavily concentrated in only one or two very popular locations. In practice slightly more interviews were taken in several locations, while only 4 interviews were completed at the art exposition. In all, 516 intercept interviews were collected, along with 325 refusals. The average refusal rate is thus 39 percent, with location-specific refusal rates ranging from only 3 percent at the food festival to almost 66 percent at one of the two grocery stores. The last column of the Table shows that the

total number of Nikkei visiting the 15 locations during the sampling period was almost 14,000.

Although 11 out of the 15 locations were in two Nikkei neighborhoods: Liberdade and Saúde, only 18 percent of the sample lived in these neighborhoods, with individuals traveling into events, and to work, shop, or visit friends.⁹ In fact, individuals reported living in over 150 distinct neighborhoods, with a few living outside of Sao Paulo state.

Table 5: Sao Paulo Intercept Survey

| Intercept Point | Number of interviews | Number of refusals | Refusal rate (%) | Time spent in location | Approximate number in location |
|---|-----------------------------|---------------------------|-------------------------|-------------------------------|---------------------------------------|
| <i>Fixed point locations</i> | | | | | |
| Cooperçotia Atlético Clube | 34 | 23 | 40.4 | 8.5 hours | 368 |
| Estação Metrô São Joaquim | 49 | 37 | 43.0 | 14 hours | 1436 |
| Feira da Liberdade | 34 | 3 | 8.1 | 5 hours | 1282 |
| Feira Livre da Rua Carneiro | 34 | 3 | 8.1 | 7 hours | 1635 |
| Hospital Santa Cruz | 42 | 12 | 22.2 | 8 hours | 374 |
| Mercearia Marukai | 54 | 76 | 58.5 | 13 hours | 2583 |
| Mercearia Satsuyama | 36 | 69 | 65.7 | 11 hours | 1922 |
| Sociedad Brasileira de Cultura Japonesa-Bunkyo | 34 | 25 | 42.4 | 9 hours | 311 |
| Agencia Sudameris | 34 | 24 | 41.4 | 8 hours | 186 |
| <i>Events</i> | | | | | |
| Cinema Bunkyo | 34 | 19 | 35.8 | 4 hours | 97 |
| ACAL Toyo Matsuri - Festival Oriental | 30 | 22 | 42.3 | 9 hours | 824 |
| ACESA Motitsuki Matsuri (Festival Gastronômico) | 29 | 1 | 3.3 | 8 hours | 424 |
| Fundação Mokiti Okada -Exposição de Obras de Arte | 4 | 2 | 33.3 | 9 hours | 67 |
| Coral Bunkyo - Concerto de Natal | 34 | 3 | 8.1 | 4 hours | 704 |
| Grupo the Friends-Koohaku Utagassen 2006 (Festival musical) | 34 | 6 | 15.0 | 11 hours | 1731 |
| | 516 | 325 | 38.6 | 128.5 hours | 13944 |

At each location, individuals were asked whether or not they had visited any of the other fixed point locations during the past two weeks, and whether they had attended or were planning on attending the six events. Only 19 percent of individuals had visited only their location, and on average individuals had visited 3.18 of the 15 locations during the two week period specified. 12 percent of individuals had visited 6 or more of the locations, with one individual going to 13 out of the 15.

⁹ 19.8 percent of the individuals interviewed in intercept locations in Liberdade and Saúde were from these two neighborhoods.

Table 6 summarizes the results of the Curitiba intercept point survey. The majority of the interviews took place at the municipal market and Clube Nikkei, the sports club. The overall refusal rate was 41 percent, very close to the 39 percent found in Sao Paulo.

Table 6: Curitiba Intercept Point Survey

| | Number of interviews | Refusal rate (%) |
|-------------------------------------|----------------------|------------------|
| <i>Fixed point locations</i> | | |
| Mercado Municipal | 241 | 49.2 |
| Clube Nikkei | 109 | 16.8 |
| Associação Bunkyo | 36 | 28.0 |
| Escola Japonesa | 3 | 0.0 |
| Associação ABD | 3 | 0.0 |
| Total | 392 | 40.7 |

Table 7 uses the Sao Paulo data to examine the characteristics of individuals who visit more locations amongst those sampled. Column 1 carries out a parsimonious OLS regression, of the number of locations as a function of gender, age, marital status, education level, employment status, and two key variables of interest for comparing across surveys: whether or not the individual has ever worked or studied in Japan, and whether or not their household receives remittances from Japan. We see that females and older individuals visit more locations. More importantly, we see that return migrants visit more locations. Column 2 then adds additional controls for generation, whether or not a household member reads Japanese newspapers, and for whether or not employed individuals refuse to give a range for income. As we would expect, individuals who are more connected to Japan, by virtue of being first or second generation Japanese, and being in households where Japanese newspapers are read, are found in more locations. Additionally, we see that individuals who refuse to give their income range are found in less locations. Similar results are seen in columns 3 and 4, which use a negative binomial model, to account for the fact that the number of locations visited is a count variable.

Table 7: Which individuals go to more intercept locations?
 Dependent variable: Number of locations visited in past 2 weeks

| | (1) | (2) | (3) | (4) |
|------------------------------|-----------|-----------|-------------------|-------------------|
| | OLS | OLS | Negative Binomial | Negative Binomial |
| Male | -0.288* | -0.294* | -0.0913* | -0.0902* |
| | (0.17) | (0.17) | (0.053) | (0.052) |
| Age | 0.0151*** | 0.00719 | 0.00467*** | 0.00229 |
| | (0.0052) | (0.0061) | (0.0016) | (0.0019) |
| Married | 0.101 | 0.0965 | 0.0349 | 0.0324 |
| | (0.18) | (0.18) | (0.056) | (0.055) |
| Has University Education | 0.00686 | 0.0842 | 0.00622 | 0.0338 |
| | (0.18) | (0.18) | (0.056) | (0.056) |
| Has worked/studied in Japan | 0.614*** | 0.313 | 0.193*** | 0.0983* |
| | (0.18) | (0.19) | (0.055) | (0.060) |
| Works for Pay | 0.0787 | 0.125 | 0.0310 | 0.0439 |
| | (0.19) | (0.19) | (0.060) | (0.059) |
| Receives Remittances | 0.436 | 0.433 | 0.119 | 0.119 |
| | (0.29) | (0.28) | (0.083) | (0.081) |
| Issei | | 0.485 | | 0.148 |
| | | (0.34) | | (0.099) |
| Nissei | | 0.433** | | 0.145** |
| | | (0.21) | | (0.067) |
| Reads Japanese newspapers | | 0.688*** | | 0.210*** |
| | | (0.20) | | (0.059) |
| Refuses to give income range | | -1.005*** | | -0.377*** |
| | | (0.34) | | (0.12) |
| Constant | 2.209*** | 2.173*** | 0.837*** | 0.811*** |
| | (0.32) | (0.32) | (0.10) | (0.10) |
| Observations | 492 | 491 | 492 | 491 |
| R-squared | 0.07 | 0.13 | | |

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

These results show that individuals who are more strongly linked to the Nikkei community have higher likelihoods of being sampled in the intercept survey. Therefore, to obtain a sample which is representative of anyone who visits any of the different intercept locations, we need to place less weight on individuals who are more likely to be found. In particular, the probability that individual i is sampled is proportional to:

$$p_i = \min \left\{ 1, \sum_{j=1}^{15} \Pr(i \text{ visits } j) * \text{Fraction of individuals sampled at } j \right\} \quad (1)$$

Where j denote the 15 intercept locations.¹⁰ We then weight the sample by the inverse of the probability that each individual was sampled.

Table 8 compares the sample means for different variables for the unweighted and weighted intercept sample. Weighting makes more of a difference in Sao Paulo than in Curitiba, since a larger and wider range of locations were used there. Nevertheless, in both cases we see that weighting reduces the proportion of individuals with strong linkages to Japan. In both cities weighting reduces our estimate of the proportion of Nikkei who are first generation (*Issei*), and the proportion who are living in households where someone reads Japanese newspapers, listens to Japanese radio, watches Japanese television programs, or checks Japanese websites. According to the unweighted sample in Sao Paulo, we would estimate that 45 percent of individuals had ever worked or lived in Japan, compared to 35 percent in the weighted sample.

The weighted sample is then representative of anyone who visited any of the different intercept locations *and agreed to answer the survey*. As noted, the refusal rate was 39 percent in Sao Paulo. The gender and approximate age of individuals refusing was collected by our interviewers, enabling us to examine the extent to which refusal varies by these characteristics. Refusal rates for males and females were not statistically different: the refusal rate was 37.1 percent for males and 40.0 percent for females, with a t-test for equality having a p-value of 0.37. In contrast, refusal rates do appear to vary by age, being lower for individuals over 50. The refusal rate is 44.4 percent for individuals 30 or under, 47.0 percent for individuals 31-49, and 27.5 percent for individuals 50 and over. There is no statistically significant difference in refusal rates between 30 and under and 31-49 year olds, but both groups have refusal rates higher than individuals 50 and over at with $p < 0.001$. Since it is likely that the characteristics of young Nikkei who refuse to answer the survey differ from those who agree to answer the survey, we do not attempt to reweight the data to adjust for refusals.

¹⁰ No individuals were interviewed more than once, but 16 out of the 516 individuals interviewed had predicted probabilities of being located of greater than 1, hence the need for imposing the minimum condition in equation (1).

Table 8: Comparison of Means of Unweighted and Weighted Intercept Samples

| | Sao Paulo | | Curitiba | |
|---|----------------------|--------------------|----------------------|--------------------|
| | Unweighted Intercept | Weighted Intercept | Unweighted Intercept | Weighted Intercept |
| Proportion Male | 0.50 | 0.58 | 0.60 | 0.60 |
| Age | 46.6 | 51.6 | 48.0 | 48.4 |
| Proportion Married | 0.48 | 0.43 | 0.60 | 0.62 |
| Proportion with Nikkei Spouse if Married | 0.82 | 0.78 | 0.77 | 0.73 |
| Proportion with some University education | 0.58 | 0.59 | 0.65 | 0.62 |
| Proportion working for pay | 0.67 | 0.69 | 0.59 | 0.61 |
| Household Size | 3.29 | 3.46 | 3.42 | 3.43 |
| Proportion Issei | 0.15 | 0.12 | 0.11 | 0.10 |
| Proportion Nissei | 0.47 | 0.45 | 0.53 | 0.56 |
| Proportion Sansei | 0.36 | 0.42 | 0.35 | 0.32 |
| Proportion Yonsei | 0.02 | 0.02 | 0.01 | 0.01 |
| Proportion who ever studied/worked in Japan | 0.45 | 0.35 | 0.34 | 0.34 |
| Proportion of households with member ever in Japan | 0.65 | 0.59 | 0.51 | 0.49 |
| Proportion with household member currently in Japan | 0.35 | 0.33 | 0.18 | 0.16 |
| Proportion with household member who.. | | | | |
| Reads Japanese/Nikkei newspapers | 0.39 | 0.28 | 0.21 | 0.18 |
| Listens to Japanese/Nikkei radio programs | 0.25 | 0.21 | 0.17 | 0.15 |
| Watches Japanese/Nikkei TV programs | 0.43 | 0.36 | 0.47 | 0.45 |
| Reads Japanese/Nikkei books/magazines | 0.42 | 0.31 | 0.37 | 0.32 |
| Reads newspapers from Nikkei associations | 0.39 | 0.28 | 0.46 | 0.37 |
| Checks Japanese/Nikkei websites on the internet | 0.24 | 0.17 | 0.29 | 0.24 |

4.3 The Snowball Survey in Sao Paulo State

The final type of survey method trialed was that of a snowball survey. The questionnaire used was the same as used for the stratified random sample. Our plan was to begin with a seed list of 75 households, and to aim to reach a total sample of 300 households through referrals from the initial seed households. Each household surveyed was asked to supply the names of three contacts: (i) a Nikkei household with a member currently in Japan; (ii) a Nikkei household with a member who has returned from Japan; and (iii) a Nikkei household without members in Japan and where individuals had not returned from Japan. They were also asked to say the number of households they knew in each category, which could then be used to weight the sample.

The first step was therefore to select the seed households. One approach likely to be followed by researchers attempting a snowball survey is to use ethnic organizations as the source of the seed households. To replicate what a reasonable researcher might do, we therefore decided to use Nikkei associations to obtain the seed households. In collaboration with Sudameris, we therefore contacted 25 associations throughout the state of Sao Paulo, who had prior associations with Sudameris. The purpose of the survey was

explained to each association, and each was asked to supply the names and contact details of three members who we could interview. Twenty of the 25 associations agreed to participate, supplying 67 seed names to us (several gave more than 3 names). The associations were asked to inform their members about the survey and obtain their consent. However, many of the individuals appear not to have been informed.

The snowball survey took place from December 5-20, 2006, and experienced two main problems. The first was that some of the households supplied as seeds by the Nikkei associations refused to answer the survey. The second problem was that among households interviewed, most households did not wish to provide referrals to other Nikkei associations. They noted that the length and content of the questionnaire made them reluctant to give the names of friends who could answer it.

In response to these problems, a second phase of the snowballing survey ran from January 22, 2007 to March 23, 2007. More associations were contacted to provide additional seed names (69 more names were obtained), and as with the stratified sample, an adaption of the intercept survey was used when individuals refused to answer the longer questionnaire. A decision was made to continue the snowball process until a target sample size of 100 had been achieved. 75 households received long survey, and 25 the short survey. Of those receiving the long survey, only 39 percent provided at least one referral. The mean number of referrals per referral-providing household was 1.5. As a result, we obtained 0.57 referrals per surveyed household – higher than the rate of one referral per four households reported by Bilborrow (2006) in his survey of Colombian migrants in Ecuador, but still much lower than hoped for.

Table 9 provides a summary of the households surveyed using the snowball survey. The final sample consists of 60 households who came as seed households from Japanese associations, and 40 households who were chain referrals. The longest chain achieved was 3 links.

Table 9: Snowball Summary Table

| | Names on seed list | Interviews Seed list | 1st reference | 2nd reference | 3rd reference | Total |
|--------------|-----------------------|-------------------------|------------------|------------------|------------------|------------|
| Seed list 1 | 67 | 42 | 19 | 8 | 7 | 76 |
| Seed list 2 | 69 | 18 | 5 | 1 | | 24 |
| Total | 136 | 60 | 24 | 9 | 7 | 100 |

The seed households are drawn from names provided by Nikkei associations, and hence one would expect these households to be more closely connected to Japan than a randomly chosen Nikkei household. The hope with snowball sampling is that the process of chain-referral will lead to coverage of other individuals, not as closely connected to Japan. However, as Table 10 shows, the snowball seed and referral households have very similar characteristics. In fact, the only variable where the means are significantly different is for watching Japanese/Nikkei TV programs, which more of the referral households do than the seed households. Thus the snowballing does not seem to have succeeded in giving households which are that different from the initial seeds.

Table 10: Differences between Seed and Referral households

| | Snowball Seeds | Snowball Referrals |
|--|-------------------|-----------------------|
| Household Size | 3.48 | 3.95 |
| <i>Percentage of households with member who:</i> | | |
| Reads Japanese/Nikkei newspapers | 48 | 41 |
| Listens to Japanese/Nikkei radio programs | 16 | 18 |
| Watches Japanese/Nikkei TV programs | 46 | 67** |
| Reads Japanese/Nikkei books/magazines | 52 | 51 |
| Reads newspapers from Nikkei associations | 59 | 69 |
| Checks Japanese/Nikkei websites on the internet | 21 | 32 |
| <i>% of households which:</i> | | |
| Migration | | |
| Have a Member currently in Japan | 28 | 33 |
| Have a Member who has returned from work/study in Japan | 49 | 54 |
| Remittances | | |
| Receive remittances from Japan | 8 | 13 |
| Refuse to say if they receive remittances | 8 | 5 |
| Sample Size | 61 | 39 |

*, **, *** indicates referral mean differs from seed mean at the 10%, 5% and 1% significance levels respectively.

5. Results Comparing the Different Methods

5.1 Comparison of Samples and Estimates of Migration and Remittance Receipt

We expect that the snowball and intercept surveys will oversample individuals which are more connected to Japan and to the Nikkei community in Brazil. This should be especially the case for the seed households in the snowball survey, who are all members of Nikkei associations. As discussed above, weighting the intercept survey households helps correct for the oversampling of individuals who attend more community events and locations, and therefore should bring the intercept survey results closer to the stratified survey. We therefore wish to test the following hypotheses:

H1: The intercept and snowball households sampled will be more closely connected to the Nikkei community than randomly sampled Nikkei households.

H2: weighting the intercept survey will bring the sample closer to the random sample.

H3: the intercept and snowball samples will over-sample *issei and nissei* (first- and second-generation Nikkei) who will be more strongly connected to Japan, and under-sample *sansei* and *yonsei* (third- and fourth generation Nikkei), who are likely to be more integrated into Brazil and less likely to attend community events or belong to community associations.

H4: the snowball and intercept surveys will overstate the proportion of households with migrant experience, due to oversampling households with more links to Japan.

H5: Refusal rates for questions about remittances will be higher for the intercept survey, since they take place in a public location.

Table 11 compares characteristics of the households surveyed using the different survey methods. Comparing the different samples, we see strong evidence of the first hypothesis. Household members in the intercept and snowball samples are much more likely to read Nikkei newspapers, books and newsletters, listen to Nikkei/Japanese radio, watch Nikkei TV programs and visit Japanese/Nikkei websites than randomly chosen Nikkei households in the stratified sample. For example, 45 percent of households in the snowball sample have a member who reads Japanese/Nikkei newspapers, compared to 28 percent in the weighted intercept survey in Sao Paulo, and 13 percent in the Sao Paulo

stratified survey. Both the intercept and snowball also overestimate the proportion of adults 18 and over who have worked in Japan, compared to the stratified survey.

Secondly, in accordance with the second hypothesis, we see that weighting the intercept sample does bring it closer to the stratified sample, in terms of links to the Nikkei community. This is the case in both the Sao Paulo and Parana surveys. There is also some support for the third hypothesis. The snowball survey picks up more second-generation and less third-generation Nikkei than the stratified survey. The intercept survey in Sao Paulo does get the same proportion of adults by generation as the stratified survey. However, in Parana, where the intercept survey visited less locations, the intercept survey actually oversamples second-generation relative to third generation, and does undersample fourth generation.

Table 11: Comparison of Characteristics of Nikkei Across Different Sampling Methods

| | Sao Paulo | | | | Parana | | |
|--|-------------------|----------------------|--------------------|--------------------|-------------------|----------------------|--------------------|
| | Stratified Survey | Unweighted Intercept | Weighted Intercept | Snowball | Stratified Survey | Unweighted Intercept | Weighted Intercept |
| Household Characteristics | | | | | | | |
| Household Size | 3.25 | 3.29 | 3.45 | 3.66* | 3.32 | 3.42 | 3.38 |
| <i>Percentage of households with member who:</i> | | | | | | | |
| Reads Japanese/Nikkei newspapers | 13 | 39*** | 28*** | 45***, +++ | 12 | 21*** | 18** |
| Listens to Japanese/Nikkei radio programs | 8 | 25*** | 21*** | 17* | 7 | 17*** | 15*** |
| Watches Japanese/Nikkei TV programs | 23 | 43*** | 36** | 54***, +++ | 25 | 47*** | 45*** |
| Reads Japanese/Nikkei books/magazines | 15 | 42*** | 30*** | 52***, +++ | 16 | 37*** | 32*** |
| Reads newspapers from Nikkei associations | 12 | 39*** | 28*** | 63***, +++ | 9 | 46*** | 37*** |
| Checks Japanese/Nikkei websites on the internet | 5 | 24*** | 18*** | 25*** | 8 | 29*** | 24*** |
| Characteristics of Adults 18 and over | | | | | | | |
| Mean Age | 47.2 | 46.6 | 45.0 | 49.2 ⁺⁺ | 43.6 | 48.0*** | 48.4*** |
| Percentage: | | | | | | | |
| Female | 52 | 50 | 46 | 56 ⁺⁺ | 51 | 40** | 40** |
| Married | 51 | 48 | 43 | 59 ⁺⁺⁺ | 57 | 60 | 62 |
| Issei | 14 | 15 | 12 | 14 | 8 | 11 | 10 |
| Nissei | 50 | 47 | 45 | 54 ⁺⁺ | 38 | 53*** | 56*** |
| Sansei | 33 | 36 | 42 | 30 ⁺⁺⁺ | 45 | 35** | 32*** |
| Yonsei | 3 | 2 | 2 | 3 | 5 | 1** | 1** |
| Have worked in Japan | 21 | 45*** | 36*** | 32** | 25 | 34** | 34** |
| Sample Size (households/individuals) | 204/270 | 516/516 | 516/516 | 100/220 | 199/330 | 392/392 | 392/392 |

*, **, and *** denotes that the mean or proportion is significantly different from that in the stratified sample at the 10%, 5% and 1% significance levels respectively.

+, ++, and +++ denotes that the snowball mean or proportion is significantly different from the weighted intercept at the 10%, 5% and 1% significance levels respectively.

Stratified sample individual characteristics are only given for individuals surveyed with the long questionnaire.

Table 12 compares the estimated percentage of Nikkei households with migration experience, and which receive remittances by survey. There is general support for the fourth hypothesis: both intercept surveys significantly overstate the percentage of households with a member who has returned from working or studying abroad. The snowball survey also gives a higher estimated percentage, although the smaller sample size leads this difference to not be statistically significant. From the stratified survey, we estimate that in both Sao Paulo and Parana, 19-21 percent of households have a member currently in Japan, and 37-41 percent have a member who has returned from working or studying in Japan. In contrast, the intercept and snowball surveys in Sao Paulo estimate that one in three Nikkei families have a migrant currently in Japan, compared to the one in five estimate from the stratified sample.

Table 12: Comparison of Migration and Remittances of Nikkei Across Different Sampling Methods

| | Sao Paulo | | | | Parana | | |
|--|-------------------|----------------------|--------------------|-----------------|-------------------|----------------------|--------------------|
| | Stratified Survey | Unweighted Intercept | Weighted Intercept | Snowball | Stratified Survey | Unweighted Intercept | Weighted Intercept |
| <i>% of households which:</i> | | | | | | | |
| Migration | | | | | | | |
| Have a Member currently in Japan | 19 | 35*** | 33** | 30 | 21 | 18 | 16 |
| Have a Member who has returned from work/study in Japan | 41 | 65*** | 59*** | 51 | 37 | 51*** | 49** |
| Remittances | | | | | | | |
| Receive remittances from Japan | 14 | 10 | 10 | 10 | 16 | 4*** | 3*** |
| Refuse to say if they receive remittances | 3 | 2 | 3 | 7 | 5 | 0*** | 0*** |
| Refuse to say how much they receive if receiving | 31 | 73*** | 82*** | 50 ⁺ | 61 | 20*** | 29* |
| Amount (Reales) | | | | | | | |
| Annual Amount Received Conditional on Receiving Remittances and Reporting Amount | | | | | | | |
| Mean | 5404 | 1483* | 1429* | 9400 | 4143 | 10850** | 11528** |
| Median | 2500 | 1792 | 1792 | 2400 | 3000 | 9000 | 10000 |
| Sample Size | 204 | 516 | 516 | 100 | 199 | 392 | 392 |

*, **, and *** denotes that the mean or proportion is significantly different from that in the stratified sample at the 10%, 5% and 1% significance levels respectively.

+, ++, and +++ denotes that the snowball mean or proportion is significantly different from the weighted intercept at the 10%, 5% and 1% significance levels respectively.

Despite these large differences in migration rates, the proportion of households receiving remittances from Japan is similar across the different survey methods in Sao Paulo. However, the Parana intercept survey substantially underestimates the percentage of households receiving remittances, compared to the stratified sample. Finally, we see that

in Sao Paulo, the proportion of those receiving remittances who refuse to report how much they receive is much higher (82 percent) in the intercept survey than in the household survey (31 percent). However, in Parana, the refusal rate for the amount received is lower in the intercept survey. This difference may arise from the fear of crime being higher in Sao Paulo, leading to more reticence in public places there. However, it should be noted that the number of households receiving remittances is small in each sample, so these differences are based on subsamples of 10 to 50 households.

5.2 Do the different sample methods give different results in regressions

Although one goal of representative surveys is to estimate population means, such as the proportion of Nikkei households which have migrant members or which receive remittances, a second goal is to use the individual level data to estimate regression models. The limited number of questions contained in the intercept point survey, and the fact that data is collected for a single member rather than a full household roster, limits the use of the intercept point method for this approach. Nevertheless, we can compare the three methods in terms of how they characterize return migrants. Given that most of the migration from Brazil to Japan is temporary, this can be viewed as also a proxy for regressions intended to examine the selectivity of migrants.

Table 13 presents the results of probit regressions intended to determine which characteristics are associated with being a return migrant. We estimate two models using each method. The first uses age, sex, marital status, education, and Nikkei generation as controls, while the second specification also adds a Nikkei media engagement index, constructed as the first principal component of the six questions on whether household members read Japanese newspapers, watch Japanese television, etc. The education distribution shows three main education levels, corresponding to natural stopping points in schooling: individuals with high school or less, those with undergraduate education, and those with postgraduate education. We therefore include dummy variables for

undergraduate and for postgraduate education, allowing us to investigate the education-selectivity of return migrants.¹¹

Table 13 shows that the three survey methods do give different pictures of how the return migrants compare to non-migrants. The stratified survey shows little selectivity, with an undergraduate education being the only variable significant at the 10 percent level. The intercept survey also shows very little selectivity, except in one respect. Living in a household with strong usage of Japanese media is strongly and positively associated with being a return migrant in the intercept point survey, but not in either of the other two survey methods. This is consistent with the intercept point oversampling both return migrants and those with high usage of Japanese media, perhaps leading to a spurious correlation between the two.

The snowball survey shows even more differences. Females are 13 percentage points more likely to be return migrants in the snowball survey, compared to no difference from males in the stratified survey. The snowball survey gives strong education selectivity, significant at the 1 percent level – individuals with postgraduate education are less likely to be return migrants, and with undergraduate more likely to be return migrants than individuals with high school or less. Sansei (third generation), are surprisingly less likely to be return migrants than the small number of fourth generation in the sample. The snowball survey method therefore gives very different results than the stratified survey or intercept point survey to basic questions of interest such as “are return migrants more likely to be men or women?”, or “are return migrants more or less educated than non-migrants”.

6. Comparison of the Costs of the Different Methods

Upon conclusion of the survey efforts, we asked *Sensus* to provide us with their updated cost breakdown for the cost of carrying out the survey using each survey method.¹² The

¹¹ Of course some of these characteristics may have changed after returning from Japan. The longer stratified survey and snowball survey contain questions on where education was obtained, but the shorter intercept survey does not. Less than five percent of highest qualifications were completed in Japan.

stratified survey and snowball survey has a 36-page questionnaire with just over 1000 variables, taking just over one hour to complete, compared to the 3 page intercept questionnaire with 60-70 variables, taking an average of 7 minutes to complete. The per household costs here include interviewer time and travel costs, but not an additional 14 percent for taxes and the administration fee for the survey firm.

Sensus estimated that the listing exercise cost US\$2 per dwelling listed, and the follow-up household interviews of households identified as Nikkei in the listing cost US\$80 each. Combining the listing and surveying, the total variable cost per household interviewed in the stratified survey was US\$212.¹³ The snowball survey was estimated to cost US\$100 each. Since households in the snowball survey are less geographically clustered than those identified through the listing exercise, the cost of administering the questionnaire was higher than the stratified survey (although the listing phase was not required). The cost of contacting the Nikkei associations and obtaining names from them is not included in this estimate, since it was carried out by a World Bank consultant. The intercept survey was much cheaper, averaging US\$30 per questionnaire.

Thus adding on 14 percent in taxes and 20 percent in administrative fees, the estimated cost of a survey of 500 questionnaires would be: \$142,000 for the random, stratified survey; \$67,000 for a snowball survey; and \$20,100 for an intercept survey. Of course in any given application local wage levels and the costs of transportation will change the levels of these, and could also change the relative ratios. Nevertheless, since few detailed migration surveys are available, and even fewer provide details on their costs, these estimates should be useful to other researchers as a starting point.

7. Discussion and Conclusions

Ethnic minorities and households containing migrants tend to be rare elements, making it difficult to obtain representative surveys in many instances. This paper has reported on an

¹² *Sensus* was awarded the contract for this project through competitive bidding based on the quality of the proposal. Since this was a new, experimental survey, with some adjustments made along the way, the ex-post costs per survey were slightly more than *Sensus* had initially anticipated.

¹³ Recall that many households had to be listed to identify one Nikkei household to survey.

experiment which compared three different sampling methods in surveying Japanese-Brazilian households in Brazil. As expected, we find that snowball and intercept point survey methods tend to sample individuals more closely tied to the Nikkei community than randomly sampled individuals identified through a two-phase stratified survey. As a consequence the use of these other methods tends to overestimate the proportion of Nikkei households with migrant experience. Nevertheless, we do find that reweighting the intercept point survey to account for individuals who are more likely to visit multiple locations does bring the results closer to the stratified sample. The different survey methods also give quite different results in probit equations intended to examine the characteristics of return migrants, with the snowball sample in particular giving quite a different picture of the gender- and educational-selectivity of migration.

The three survey methods used here are often applicable in migrant-receiving countries, since migrants tend to cluster in certain geographic locations, and be regular visitors at certain fixed points and community events. The fact that migrants are often ethnically distinguishable from many citizens of the receiving country makes identification of potential migrants easier in intercept point surveys, but is not a necessary condition for the success of this method. In terms of migrant-sending countries, the stratified survey with listing can again be applied without much conceptual difficulty. It may be more difficult to think of intercept point where non-ethnically identified families of migrants congregate, but locations such as festivals, transportation hubs, money-transmitting branches, churches, social support networks may be starting points.

So what do we conclude from this experiment? The first is that, in practice, intercept point and snowball surveys are unlikely to provide a representative sample of the whole population of migrants or migrant-sending families. In particular, they are likely to oversample individuals more closely connected to the community. Secondly, as the results here and those in Bilborrow (2006) show, snowball surveys of migrants or their families may be quite ineffective in practice at creating the long referral chains needed for this method to capture the target population. In our case, the snowball sample gave quite different (misleading compared to the stratified sample) pictures of the gender and

educational selectivity of return migrants. Furthermore, the snowball method is not that much cheaper than a representative sample where the sampling weights are known. Thirdly, while the intercept method does not provide a representative sample of the whole population, surveying many locations and using reweighting does help make it more representative. Moreover, such a survey is much cheaper than the stratified sample, albeit at the cost of much less data being able to be collected. The intercept survey is therefore most likely to be of use for exploratory analysis, and for situations where the target population of interest are those who attend community locations. This may be the case when policy interventions will rely on these same types of locations to reach migrants. However, there appears no very close substitute to the more time-consuming and expensive two-phase stratified sampling in obtaining truly representative surveys.

References

Beltrão, Kaizô Iwakami and Sonoe Sugahara (2006) “Permanente Temporário: Dekasseguis Brasileiros no Japão”, Mimeo. IBGE, Brazil.

Bilsborrow, Richard (2006) “The Design of Samples for International Migration Surveys in Countries of Destination: Methodological Issues and Lessons from NiDi Surveys in Spain and Italy”, Paper prepared for the World Bank.

Bilsborrow, R.E., Graeme Hugo, A.S. Oberai and Hania Zlotnik (1997) *International Migration Statistics: Guidelines for Improving Data Collection Systems*. International Labour Office: Geneva.

Blangiardo, G. (1993) “Una nuova metodologia di campionamento per le indagini sulla presenza straniera. (A new sample methodology for the surveys on foreign immigration)” In: L. Di Comite and M. De Candia (eds.), *I fenomeni migratori nel bacino mediterraneo. (Migration movements in Mediterranean basin)*. Bari: Cacucci.

Bustamente, Jorge, Guillermina Jasso, J. Edward Taylor and Paz Trigueros Legarreta (1997) “Characteristics of Migrants: Mexicans in the United States”, pp. 91-162 in *Mexico-U.S. Binational Migration Study Report*, U.S. Commission on Immigration Reform, <http://www.utexas.edu/lbj/uscir/binational.html>

Fawcett, James T. and Fred Arnold (1987) “The Role of Surveys in the Study of International Migration: An Appraisal”, *International Migration Review* 21(4): 1523-40.

Global Commission on International Migration (2005) *Migration in an interconnected world: New directions for action*, Report of the Global Commission on International Migration, www.gcim.org.

Goodman, Leo (1961) “Snowball Sampling”, *Annals of Mathematical Statistics* 32(1): 148-170.

Goto, Junichi (2006) “Latin Americans of Japanese Origin (*Nikkeijin*) Working in Japan – A Survey”, Mimeo. Kobe University.

Groenewold, George and Richard Bilsborrow (2004) “Design of Samples for International Migration Surveys: Methodological Considerations, Practical Constraints and Lessons Learned from a Multi-Country Study in Africa and Europe”, Paper presented at the Population Association of America 2004 General Conference, Boston, Massachusetts.

- Heckathorn, Douglas D. (1997) "Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations", *Social Problems* 44(2): 174-99.
- Heckathorn, Douglas D. (2002) "Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations", *Social Problems* 49(1): 11-34.
- Higuchi, Naoto (2006) "Brazilian Migration to Japan: Trends, Modalities and Impact", Paper given to the UN Expert Group Meeting on International Migration and Development in Latin America and the Caribbean, Mexico City, 30 November-2 December 2005.
- Kalsbeek, William D. (1986) "Nomad Sampling: An Analytic Study of Alternative Design Strategies," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 164-169.
http://www.amstat.org/sections/srms/Proceedings/papers/1986_028.pdf
- Kalton, Graeme (1991) "Sampling flows of mobile human populations. *Survey Methodology*, 17, 183-194.
- Kalton, Graeme (2001) "Practical Methods for Sampling Rare and Elusive Populations", *Proceedings of the Annual Meeting of the American Statistical Association*,
<http://www.amstat.org/sections/srms/Proceedings/y2001/Proceed/00454.pdf>
- Kalton, Graham and Dallas W. Anderson (1986) "Sampling Rare Populations", *Journal of the Royal Statistical Society, Series A* 149(1): 65-82.
- Kish, Leslie (1965) *Survey Sampling*. New York, Wiley and Sons.
- Massey, Douglas and Audrey Singer (1995) "New Estimates of Undocumented Mexican Migration and the Probability of Apprehension", *Demography* 32(2): 203-213.
- McKenzie, David, John Gibson, and Steven Stillman (2006) "How Important is Selection? Experimental versus Non-experimental Measures of the Income Gains from Migration", *World Bank Policy Research Working Paper* No. 3906.
- McKenzie, David and Alejandrina Salcedo (2007) "Japanese-Brazilians and the Future of Brazilian Migration to Japan", Mimeo. World Bank.
- Osili, Una Okonkwo (2006) "Remittances and Savings from International Migration: Theory and Evidence Using a Matched Sample", *Journal of Development Economics*, forthcoming.

Tsuda, Takeyuki (1999) “The Motivation to Migrate: The Ethnic and Sociocultural Constitution of the Japanese-Brazilian Return-Migration System”, *Economic Development and Cultural Change* 1-31.

Tsuda, Takeyuki (2003) *Strangers in the Ethnic Homeland: Japanese Brazilian Return Migration in Transnational Perspective*. Columbia University Press: New York.

Wasserman, Melanie, Deborah Bender, William Kalsbeek, Chirayath Suchindran and Ted Mouw (2005) “A Church-Based Sampling Design for Research with Latina Immigrant Women”, *Population Research and Policy Review* 24(6): 647-71.

World Bank (2005) *Global Economic Prospects 2006: Economic Implications of Remittances and Migration*. World Bank: Washington D.C.

Appendix: How Selecting Enumeration Areas Based on a 10% Sample from the Census can lead to overestimating the target population

In many countries, including Brazil, the long form of the Census is only administrated to 10 percent of the households. If characteristics needed to identify the minority population of interest are only found in the long form, then this 10 percent sample will have to be used to identify sampling clusters or enumeration areas. However, the result is likely to be an overestimation of the target population if one omits census tracts where zero minority households were found in the subsample, as this example illustrates.

Consider 1000 clusters, each of 300 households (the size of the average census tract in our survey). Suppose that 200 of the clusters have 3 minority households in them (Nikkei households in our application), 500 have 10 minority households, and the remaining 300 each have 25 minority households. For each of these 1000 clusters, the 10% census sample of 30 households will be used to estimate the number of minority households in that cluster. Appendix Table 1 provides the number of clusters according to the number of minority households expected to be found in the 10% subsample.

Appendix Table 1: Expected Distribution of Clusters using a 10 percent Census Sample to Predict the Number of Minority Households

| Number of Minority Households expected in 10% sample | Clusters with total of 3 households | Clusters with total of 10 households | Clusters with total of 25 households | All clusters |
|--|-------------------------------------|--------------------------------------|--------------------------------------|--------------|
| 0 | 146 | 172 | 19 | 337 |
| 1 | 49 | 197 | 58 | 304 |
| 2 | 5 | 98 | 82 | 185 |
| 3 | 0 | 27 | 72 | 99 |
| 4 | 0 | 5 | 42 | 47 |
| 5 | 0 | 1 | 19 | 20 |
| 6 | 0 | 0 | 6 | 6 |
| 7 | 0 | 0 | 2 | 2 |
| 8 | 0 | 0 | 0 | 0 |
| Total # of clusters | 200 | 500 | 300 | 1000 |

The true mean number of minority households per cluster across all 1000 clusters is 13.1. However, Appendix Table 1 shows that 33.7% of clusters would be expected to have no minority households in their 10 percent subsample. In practice it is unlikely that a survey

of minority households would go to the expense of listing households in a census tract where no minority households are found in the subsample. As a result, if one ignores the 337 census tracts above with zero minority households in the 10 percent sample, and calculates the expected number of minority households over the remaining 663 census tracts, the distribution in Appendix 1 would lead to a prediction of 19.7 minority households per census tract. Comparing this to the true mean shows that we would overestimate the number of minority households by 50.4 percent. While the exact amount of overestimation will depend on the distribution of cluster minority-household ratios, the example given here shows that the overestimation can be large in an example modeled on the distribution of Nikkei households found in our work. Together with the misclassification of some other Asian households as Nikkei, this should explain why the number of Nikkei households found during our listing operation was only half that predicted based on the 10 percent Census sample.