

## Surveying migrant households: a comparison of census-based, snowball and intercept point surveys

David J. McKenzie and Johan Mistiaen

*World Bank, Washington DC, USA*

[Received October 2007. Final revision September 2008]

**Summary.** 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 that was 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 that were conducted were households selected randomly from a door-to-door listing using the Brazilian census to select census blocks, a snowball survey using Nikkei community groups to select the seeds and an intercept point survey that was collected at Nikkei community gatherings, ethnic grocery stores, sports clubs and other locations where family members of migrants are likely to congregate. We analyse 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.

**Keywords:** Intercept point; Migration; Snowball survey; Surveying rare elements

### 1. Introduction

The importance of international migration for development has received increasing attention from the research and policy communities (e.g. Global Commission on International Migration (2005) and World Bank (2005)). The focus of much of this research and discussion has been on examining the effects 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 (<http://go.worldbank.org/UK1ETMHBNO>) provide more information about living standards, education and other outcomes of interest, but usually relatively little information on the migration process and, in cases where migration is a rare event, few migrant households. 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 difficult to generalize any conclusions that are reached from them. As Fawcett and Arnold (1987) noted, common approaches that are used by many studies are to choose their samples from individuals who belong to church groups, social organizations or other defined groups, to use snowball samples

*Address for correspondence:* David J. McKenzie, Development Research Group, World Bank, MSN MC3-307, 1818 H Street North West, Washington DC 20433, USA.  
E-mail: [dmckenzie@worldbank.org](mailto:dmckenzie@worldbank.org)

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:

- (a) a stratified sample using the census to sample census tracts randomly, 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;
- (b) a snowball survey in which households are asked to provide referrals to other households with migrant members;
- (c) an intercept point survey (or time-and-space sampling survey), in which individuals are sampled during set time periods at a prespecified 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 among 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 labour markets (Tsuda, 1999; Higuchi, 2006). The estimated 265 000 migrants send approximately \$2 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.

Although our application involves surveying Nikkei households (an ethnic minority), with and without migrants abroad, the methodologies that are 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 various 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, whereas Section 4 describes how the various methodologies were applied in practice. Section 5 compares the results of the three survey methods, and Section 6 provides a cost comparison across the various methods. Section 7 concludes.

## **2. Different methods used for sampling migrants, families of migrants and other rare elements**

Bilborrow *et al.* (1997) noted that, in three-quarters of the countries in the world, the proportion of international migrants was at most 6.5% in the early 1990s. Even in countries in which international migration is more common, finding a household with a migrant who is currently abroad or a recently returned migrant can be a rare event. 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 by using administrative records on new immigrants. Examples include the New Immigrant Survey in the USA (<http://nis.princeton.edu/>), the Longitudinal Survey of Immigrants to Australia (<http://www.immi.gov.au/media/research/lisia/index.htm>) and the

Longitudinal Immigration Survey in New Zealand ([www.immigration.govt.nz/migrant/general/generalinformation/research/lisnz/](http://www.immigration.govt.nz/migrant/general/generalinformation/research/lisnz/)). 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 that we are aware of which does this is the Pacific Island–New Zealand Migration Survey (McKenzie *et al.*, 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. These are stratified sampling using disproportionate sampling fractions with two-phase sampling, snowball sampling and intercept point sampling, also known as time-and-space sampling, location sampling or aggregation point sampling. We discuss each in turn.

The use of stratified sampling with disproportionate sampling fractions is the approach that was recommended by Bilsborrow *et al.* (1997) in their guidelines for improving international migration statistics. They noted 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 recommended the use of the census to select provinces, districts and, if possible, census sectors, with probability proportional to the number of households with migrants. After census sectors have been 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 that are 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 Netherlands Interdisciplinary Demographic Institute–Eurostat surveys in Egypt, Ghana, Morocco, Senegal and Turkey (<http://www.nidi.knaw.nl/web/html/pushpull/>). In most of these applications, surveying is first restricted to certain provinces or districts where migrants are thought to come from, to reduce survey costs. For example in Ghana the survey chose 17 electoral districts and screened 21504 households according to household migration status, to arrive at a target sample of 1980 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 households to identify households with migrants. Fawcett and Arnold (1987) noted also that non-response can be a major problem in immigration surveys, particularly in urban areas. They pointed out that, whereas individuals usually have a legal obligation to answer questions in the census, surveys generally carry no legal sanctions for refusal to respond. There are exceptions with government-run household surveys in some countries, such as labour force surveys in Canada, Australia, New Zealand and the USA. In addition, in urban areas, immigrants who often work long hours, making it difficult to find them at home, and undocumented immigrants may be reluctant to take part in a survey for fear of being found by government authorities.

A second method that is 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 who are part of the rare population, and each of these is then asked to name  $k$  different individuals who belong to the rare population, and so on. Snowball sampling has been used by

the Mexican Migration Project to sample permanent Mexican migrants in the USA (Massey and Singer, 1995) and was used in part by the Netherlands Interdisciplinary Demographic Institute–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 are that the subjects may not refer friends to protect their privacy, and that contact information is frequently inadequate, so attrition rates can be high. For example, Bilsborrow (2007) reported that, in a 2006 survey of Colombian migrants in Ecuador, the snowballing procedure worked poorly, with fewer than one referral obtained per four interviewed households.

The third method that is 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 (Bustamente *et al.*, 1997), and surveying Latina immigrant women at churches in the USA (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 was 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), proposed a similar methodology for sampling migrants, which was used in the Netherlands Interdisciplinary Demographic Institute–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 centres, 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. 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 and 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 labour to the coffee plantations (Goto, 2007). High rates of migration from Japan to Brazil occurred from 1925 to 1936 as the Japanese Government subsidized emigration, and again from 1955 to 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 1987–1988 (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 \$2 billion in remittances back to Brazil (Beltrão and Sugahara, 2006).

The survey was designed to provide detail on the characteristics of households with and without migrants, to estimate the proportion of households receiving remittances and with migrants in Japan, and to 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 can give similar results to the more expensive census-based survey, and test for the presence of the types of biases that we might expect.

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 their homes, and many are reluctant to be interrupted at home outside 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

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.

##### 4.1. Stratified random sample of Nikkei households in Sao Paulo and Parana

It is estimated that 80% of the Nikkei population live in just two of Brazil's states: 54% in the state of Sao Paulo—population 37.0 million—and 26% in Parana state—population 9.6 million. (Population numbers are taken from the 2000 census: <http://www.ibge.gov.br/english/estatistica/populacao/censo2000/>.) We therefore decided to survey only 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 to determine which households had a Nikkei member. Third, the survey questionnaire was then administered to households that were identified as Nikkei. We now describe the details of each step.

##### 4.1.1. Selection of census tracts

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. On the basis of 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 it will tend to be an overstatement due to misclassifying as Nikkei households comprising individuals of Korean, Taiwanese or Chinese ethnicity who were all born in Brazil and had not been in those countries in the last 10 years. In the 2000 Brazilian census, Japanese born accounted for 74% of all yellow race immigrants, Chinese 11%, Koreans 9% and Taiwanese 5%.

The 2000 census was then used to estimate the number of Nikkei in each municipality, *área de ponderação* (AP) and census tract. An AP is the smallest geographical unit that is 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.

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 sampling with replacement, where size is the number of Nikkei households. Secondly, within each municipality that was selected, APs were sampled with probability proportional to size. Then, finally, census tracts were sampled with probabilit-

ity proportional to size within the APs. 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 that were 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 were chosen from those with 4–15 Nikkei households. We did not include census tracts with three or fewer estimated Nikkei, as they are estimated to cover only 1–3% of the Nikkei population in the two states, and listing such census tracts would increase the cost of the survey 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.

#### 4.1.2. *Listing*

Before the listing operation, letters were sent to approximately 150 Nikkei associations with bases in the areas that were 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 October 13th, 2006, and October 29th, 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 who were 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 because of 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 neighbour or building manager.

Table 1 summarizes the results of the listing process. The listing covered 14239 dwelling units in Sao Paulo state and 8300 units in Parana state, for a total of 22539 dwellings. This was 21% more dwelling units than recorded for these census tracts in the 2000 census, showing the extent of population growth and new construction over the 6 years since the census. Among these 22539 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 more difficult to find at home. Thus 3.7% of the dwelling units that were listed contained Nikkeis.

The census tracts that were listed show a large variation in the number of Nikkei. The mean size of a census tract was 301 households. The mean and median number of Nikkei households in a census tract were 11 and 8 respectively. 59 of the 75 census tracts each had fewer 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.

#### 4.1.3. *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 that were

**Table 1.** Listing of households in Sao Paulo and Parana states

	<i>Results for Sao Paulo state</i>	<i>Results for Sao Paulo city</i>	<i>Results for Parana state</i>	<i>Results for Curitiba city</i>	<i>Combined sample</i>
Number of <i>municípios</i> 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
Percentage 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 3 visits	24.5	43.2	11.8	7.7	20.3

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 on November 19th, 2006, and all dwellings were visited at least once by December 22nd, 2006. Some of the households identified by proxy reporting 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 could not 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 18th, 2007, to February 2nd, 2007, which was intended to increase the number of households responding. We made some changes to the survey protocol to attempt to obtain a response from households that had not been 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 associations contacted agreed to do this and provided phone numbers and names which could be used in a letter that was 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 that the survey was legitimate.
- (b) The initial round of interviewing used Brazilian interviewers who were not Nikkei, owing to difficulties in hiring Nikkei who were interested in carrying out survey work. More intensive efforts were undertaken to find Nikkei workers, allowing Nikkei fieldworkers to be used in this second wave.
- (c) Prizes were used to try to increase the incentive to participate. Interview subjects were told that a random drawing would be done among 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

**Table 2.** Final survey results of the stratified random sample

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

themselves, and later picked up. This shorter questionnaire was the same as used in the intercept point survey, taking seven minutes on average. The intention with the shorter survey was to provide some data on households that would not answer the full survey because of time constraints, or because they were reluctant to have an interviewer in their house.

This strategy was very successful in increasing response rates, yielding a further 45 full questionnaires and 111 short questionnaires. Table 2 summarizes the final results of this survey process. In total, we could survey 403 out of the 710 Nikkei households, an interview rate of 57%. The refusal rate was 25%, whereas the remaining households were either absent on three attempts or were not surveyed because building managers refused 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.

The group of households that were 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% for the first round, compared with 67% 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.

#### 4.2. *The intercept point surveys*

The intercept survey was designed to carry out interviews at a range of locations that were 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. 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 seven minutes to answer. Respondents had to be 18 years old or older to be interviewed. Interviewing for the Sao Paulo intercept survey took place between December 9th, 2006, and December 20th, 2006, whereas the Curitiba intercept survey took place between March 3rd and March 12th, 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 nine fixed point locations and six events. The nine fixed locations are a sports club, a metro station in the Liberdade neighbourhood, 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 neighbourhood. The six events were an afternoon Japanese film event that was organized by the Sociedade Brasileira de Cultura Japonesa, a large cultural festival with music, dancing and *taiko* drumming that was organized by the Associação Comercial e Assistencial da Liberdade, a Japanese food festival that was organized by the Associação Cultural Esportiva de Santana, a Japanese art exposition that was organized by Fundação Mokiti Okada, a Christmas concert that was organized by Coral do Bunkyo, Paineiras e Silver Boys and a music festival that was 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 that is associated with the Bunkyo and a second Nikkei association, the Associação Brasileira de Dekasseguis.

Interviewers were assigned to visit each location during prespecified blocks of time. Two fieldworkers 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 old or older who passed by each location. For the fixed places, this count was made throughout the prespecified time block. For example, between 2.30 p.m. and 3.30 p.m. 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 that they were not Nikkei. The proportion of falsely identified Nikkei was used to adjust the count that was taken by the fieldworker to obtain an estimate of the number of Nikkei passing the intercept location.

In the case of intercept surveys that were 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 obtain an estimate of the number of adult Nikkei attending the event.

Table 3 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, to make sure that the sample was not too heavily concentrated in only one or two very popular locations. In practice slightly more interviews were taken in several locations, whereas only four 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%, with location-specific refusal rates ranging from only 3% at the food festival to almost 66% at one of the two grocery stores. The last column of Table 3 shows that the total number of Nikkei visiting the 15 locations during the sampling period was almost 14000.

At each location, individuals were asked whether or not they had visited any of the other fixed point locations during the past 2 weeks, and whether they had attended or were planning on attending the six events. Only 19% of individuals had visited only their location, and on average individuals had visited 3.18 of the 15 locations during the 2-week period that was specified. 12% of individuals had visited six or more of the locations, with one individual going to 13 out of the 15.

The number of locations that were visited by an individual during a 2-week period is a count variable. Count data are commonly estimated by using Poisson regression or negative binomial

**Table 3.** Sao Paulo intercept survey

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

models (see Cameron and Trivedi (1998) for an overview). The negative binomial model includes Poisson regression as a special case but in general differs by not assuming that the mean and the variance of the underlying data distribution are the same, allowing more dispersion than the Poisson model. The density function for the negative binomial model with mean  $\mu = \exp(\mathbf{x}'\beta)$ , where  $\mathbf{x}$  is a vector of covariates and  $\beta$  the corresponding coefficient parameters, and variance  $\mu + \alpha\mu^2$  is

$$f(y|\mu, \alpha) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu}\right)^y, \quad \alpha \geq 0, \quad y = 0, 1, 2, \dots,$$

where

$$\Gamma(a) = \int_0^\infty \exp(-t)t^{a-1} dt$$

is the gamma function. This reduces to the Poisson model if  $\alpha = 0$ .

We estimate the negative binomial model by maximum likelihood using the `nbreg` command in Stata 10.0. This is a non-linear model and, because of the exponential conditional mean, the effect of a 1-unit change in a particular regressor will depend on the values of the other regressors. We therefore use Stata's `mfx` command to report marginal effects, estimated at the mean of the other regressors.

Table 4 then uses the Sao Paulo data to examine the characteristics of individuals who visit more locations among those sampled. Column 1 provides a parsimonious specification, modelling the number of locations that are visited as a function of the covariates—gender, age, marital

**Table 4.** Which individuals go to more intercept locations?†

<i>Characteristic</i>	<i>1, negative binomial</i>	<i>2, negative binomial</i>
Male	-0.288‡ (0.16)	-0.316‡ (0.17)
Age	0.015§ (0.005)	0.005 (0.006)
Married	0.110 (0.18)	0.102 (0.17)
Has university education	0.020 (0.18)	0.077 (0.18)
Has worked or studied in Japan	0.614§ (0.18)	0.306 (0.19)
Works for pay	0.097 (0.19)	0.042 (0.19)
Receives remittances	0.395 (0.29)	0.455 (0.28)
First-generation Nikkei (Issei)		0.538 (0.35)
Second-generation Nikkei (Nissei)		0.484§§ (0.21)
Reads Japanese newspapers		0.682§§ (0.20)
$\alpha$	0.015	0.005
$p$ -value for $\chi^2$ -test of $\alpha = 0$	0.208	0.389
Observations	492	491

†Dependent variable: number of locations visited in past 2 weeks. The coefficients shown are marginal effects of an instantaneous change in the variable on the predicted number of locations, or a discrete change from 0 to 1 for dummy variables on the predicted number of locations, evaluated at the mean of the other variables. Standard errors are given in parentheses. Omitted categories for dummy variables are female, not married, less than university education, does not work for pay, does not receive remittances, third or fourth generation and does not read Japanese newspapers.

‡ $p < 0.1$ .

§ $p < 0.01$ .

§§ $p < 0.05$ .

status, level of education, 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 and whether or not a household member reads Japanese newspapers, variables which are correlated with age and working abroad. As we might expect, individuals who are more connected to Japan, by virtue of being first- ( $p = 0.126$ ) or second-generation Japanese, and being in households where Japanese newspapers are read, are found in more locations. There are also large, but marginally significant, positive associations with working in Japan before ( $p = 0.107$ ) and receiving remittances ( $p = 0.110$ ).

Note that, in both specifications in Table 4, the estimate of  $\alpha$  is close to 0 and statistically insignificant, indicating that the negative binomial model is very similar in this application to the Poisson model. The choice of a negative binomial model rather than a Poisson model allows

the data to decide whether the Poisson model is sufficient, rather than *ex ante* imposing this restriction.

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. To do this, we employ the ‘weight share’ method of Lavalley (1995). Let  $N_i$  be the number of locations that are visited in the past 2 weeks for individual  $i$  and  $f_j$  be the fraction of individuals who are sampled at location  $j$  during the same 2-week period. Then the weight for each observed instance of individual  $i$  is

$$w_i = \frac{1}{N_i} \sum_{j \text{ visited by } i} f_j^{-1}. \quad (1)$$

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% 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% for males and 40.0% 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 years of age. We do not adjust our weighting to account for this difference in refusal rates by age for two reasons. First, we prefer to restrict the sample to be representative of those who agree to answer the survey, rather than assuming that individuals who refuse the survey are more similar apart from age than those who accept the survey. Second, the age ranges that were collected by the interviewers are very crude, based on whether individuals are under or over 50 years old.

#### 4.3. The snowball survey in Sao Paulo state

The final type of survey method that was trialled was that of a snowball survey. The questionnaire that was 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:

- (a) a Nikkei household with a member currently in Japan;
- (b) a Nikkei household with a member who has returned from Japan;
- (c) 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 that 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 that is 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 whom we could interview. 20 of the 25 associations agreed to participate, supplying 67 seed names to us (several gave more than three names). The associations were asked to inform their members about the survey and to obtain their consent. However, many of the individuals appear not to have been informed.

The snowball survey took place from December 5th to 20th, 2006, and experienced two main problems. The first was that some of the households that were 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 households. 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 22nd, 2007, to March 23rd, 2007. More associations were contacted to provide additional seed names (69 more names were obtained) and, as with the stratified sample, an adaptation 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 the long survey, and 25 the short survey. Of those receiving the long survey, only 39% 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 household surveyed—higher than the rate of one referral per four households that was reported by Bilsborrow (2007) in his survey of Colombian migrants in Ecuador, but still much lower than hoped for.

Table 5 provides a summary of the households that were surveyed by using the snowball survey.

**Table 5.** Snowball survey summary table

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

**Table 6.** Differences between seed and referral households

<i>Characteristic</i>	<i>Snowball seeds</i>	<i>Snowball referrals</i>
Household size	3.48	3.95
Percentage of households with member who		
Reads Japanese or Nikkei newspapers	48	41
Listens to Japanese or Nikkei radio programmes	16	18
Watches Japanese or Nikkei television programmes	46	67†
Reads Japanese or Nikkei books or magazines	52	51
Reads newspapers from Nikkei associations	59	69
Checks Japanese or Nikkei Web sites on the Internet	21	32
% of households which, for migration,		
Have a member currently in Japan	28	33
Have a member who has returned from work or study in Japan	49	54
% of households which, for remittances,		
Receive remittances from Japan	8	13
Refuse to say whether they receive remittances	8	5
Sample size	61	39

†Referral mean differs from the seed mean at the 5% level of significance.

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 three links.

The seed households were drawn from names that were provided by Nikkei associations, and hence we 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, who are not as closely connected to Japan. However, as Table 6 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 or Nikkei television programmes, 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 much different from the initial seeds.

## 5. Results comparing the various methods

### 5.1. Comparison of samples and estimates of migration and remittance receipt

We expect that the snowball and intercept surveys will oversample individuals who are more connected to Japan and to the Nikkei community in Brazil. This should be especially so for the seed households in the snowball survey, who are all members of Nikkei associations. As discussed above, weighting the intercept survey households helps to 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 results. We therefore wish to test the following hypotheses.

H1: the intercept and snowball households that are 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 oversample *issei and nissei* (first- and second-generation Nikkei) who will be more strongly connected to Japan, and undersample *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 to belong to community associations.

H4: the snowball and intercept surveys will overstate the proportion of households with migrant experience, owing 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 7 compares characteristics of the households that were surveyed by using the various 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, to listen to Nikkei or Japanese radio, to watch Nikkei television programmes and to visit Japanese or Nikkei Web sites than randomly chosen Nikkei households in the stratified sample. For example, 45% of households in the snowball sample have a member who reads Japanese or Nikkei newspapers, compared with 25% in the weighted intercept survey in Sao Paulo, and 13% in the Sao Paulo stratified survey. Both the intercept and the snowball surveys also overestimate the proportion of adults who are 18 years old and over who have worked in Japan, compared with 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 so in both the Sao Paulo and the Parana surveys. As we saw in Table 4, females, older individuals

**Table 7.** Comparison of characteristics of Nikkei across different sampling methods†

Characteristic	Results for Sao Paulo			Results for Parana			
	Stratified survey	Unweighted intercept	Weighted intercept	Snowball survey	Stratified survey	Unweighted intercept	Weighted intercept
<i>Household characteristics</i>							
Household size	3.25	3.29	3.44	3.66‡	3.32	3.42	3.38
Percentage of households with member who							
Reads Japanese or Nikkei newspapers	13	39§	25§§	45§*	12	21§	18§§
Listens to Japanese or Nikkei radio programmes	8	25§	23§	17‡	7	17§	15§
Watches Japanese or Nikkei television programmes	23	43§	38§§	54§**	25	47§	45§
Reads Japanese or Nikkei books or magazines	15	42§	33§	52§*	16	37§	32§
Reads newspapers from Nikkei associations	12	39§	29§	63§*	9	46§	37§
Checks Japanese or Nikkei Web sites on the Internet	5	24§	15§	25§‡‡	8	29§	24§
<i>Characteristics of adults 18 years old and over</i>							
Mean age (years)	47.2	46.6	44.0	49.2*	43.6	48.0§	48.4§
Percentage female	52	50	47	56	51	40§§	40§§
Percentage married	51	48	41	59*	57	60	62
Percentage <i>issei</i>	14	15	8	14‡‡	8	11	10
Percentage <i>nissei</i>	50	47	48	54	38	53§	56§
Percentage <i>sansai</i>	33	36	41	30**	45	35§§	32§
Percentage <i>yonsai</i>	3	2	2	3	5	1§§	1§§
Percentage have worked in Japan	21	45§	32‡	32§§	25	34§§	34§§
Sample size (households/individuals)	204/270	516/516	516/516	100/220	199/330	392/392	392/392

†Stratified sample individual characteristics are only given for individuals who were surveyed with the long questionnaire.  
‡Mean or proportion is significantly different from that in the stratified sample at the 10% level of significance.  
§Mean or proportion is significantly different from that in the stratified sample at the 1% level of significance.  
§§Mean or proportion is significantly different from that in the stratified sample at the 5% level of significance.  
\*Snowball mean or proportion is significantly different from the weighted intercept at the 1% level of significance.  
\*\*Snowball mean or proportion is significantly different from the weighted intercept at the 5% level of significance.  
‡‡Snowball mean or proportion is significantly different from the weighted intercept at the 10% level of significance.

**Table 8.** Comparison of migration and remittances of Nikkei across different sampling methods

Characteristic	Results for Sao Paulo			Results for Parana			
	Stratified survey	Unweighted intercept	Weighted intercept	Snowball survey	Stratified survey	Unweighted intercept	Weighted intercept
% of households which, for migration, Have a member currently in Japan	19	35†	40†	30	21	18	16
Have a member who has returned from work or study in Japan	41	65†	59†	51	37	51†	49†
% of households which, for remittances, Receive remittances from Japan	14	10	10	10	16	4†	3†
Refuse to say whether they receive remittances	3	2	3	7	5	0†	0†
Refuse to say how much they receive if receiving	31	73†	77†	50	61	20†	29§
<i>Amount (reales)</i>							
Annual amount received conditional on receiving remittances and reporting amount							
Mean	5404	1483§	1512§	9400	4143	10850†	11528†
Median	2500	1792	1792	2400	3000	9000	10000
Sample size	204	516	516	100	199	392	392

† Mean or proportion is significantly different from that in the stratified sample at the 1% level of significance.  
 ‡ Mean or proportion is significantly different from that in the stratified sample at the 5% level of significance.  
 § Mean or proportion is significantly different from that in the stratified sample at the 10% level of significance.

and individuals who are more connected to Japan visited more locations. Weighting therefore lowers the mean age, the percentage that are female and the mean extent of connection to Japan.

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 not statistically differ in the proportions of adults by generation from the stratified survey. However, in Parana, where the intercept survey visited fewer locations, the intercept survey actually oversamples second-generation relative to third-generation Nikkei and does undersample fourth-generation Nikkei.

Table 8 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 be not statistically significant. From the stratified survey, we estimate that, in both Sao Paulo and Parana, 19–21% of households have a member currently in Japan, and 37–41% 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 who is currently in Japan, compared with the one in five estimate from the stratified sample.

Despite these large differences in rates of migration, 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 with 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 (77%) in the intercept survey than in the household survey (31%). 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 10–50 households.

### *5.2. Do the various 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 that are contained in the intercept point survey, and the fact that data are 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 that are intended to examine the selectivity of migrants.

Table 9 presents the results of probit regressions that were intended to determine which characteristics are associated with being a return migrant. We use the `dprobit` command in Stata 10.0 and report marginal effects, which are interpreted as the change in the probability of being a return migrant associated with a marginal change in the variable of interest, computed at the mean of all the other variables in the model. Since the stratified sample and the snowball sample have multiple individuals per household, we adjust the standard errors for clustering by sandwich estimation, by using the cluster option in Stata.

We estimate two models by using each method. The first uses age, sex, marital status, education and Nikkei generation as controls, whereas the second specification also adds a Nikkei

**Table 9.** Do different sampling methods give different pictures of who return migrants are?†

	Results for stratified survey		Results for intercept survey		Results for snowball survey	
	(1)	(2)	(3)	(4)	(5)	(6)
Age	-0.0000349 (0.0036)	-0.000394 (0.0038)	0.000796 (0.0042)	-0.00302 (0.0045)	-0.00656 (0.0054)	-0.00654 (0.0055)
Female	-0.0835 (0.060)	-0.0800 (0.060)	-0.0688 (0.086)	-0.0143 (0.088)	0.130‡ (0.066)	0.130‡ (0.064)
Married	0.0779 (0.11)	0.100 (0.11)	-0.00625 (0.10)	0.0306 (0.10)	0.0395 (0.15)	0.0389 (0.14)
Undergraduate education	-0.136§ (0.072)	-0.131§ (0.076)	-0.0690 (0.091)	0.00518 (0.093)	0.257§§ (0.078)	0.256§§ (0.079)
Postgraduate education	0.333 (0.30)	0.373 (0.31)	-0.133 (0.10)	-0.111 (0.11)	-0.311§§ (0.090)	-0.311§§ (0.089)
Issei	0.139 (0.23)	0.132 (0.22)	0.428 (0.31)	0.296 (0.35)	0.487§ (0.29)	0.490§ (0.27)
Nissei	-0.0984 (0.13)	-0.106 (0.13)	0.286 (0.20)	0.303 (0.21)	-0.149 (0.25)	-0.146 (0.24)
Sansei	0.00809 (0.15)	-0.00644 (0.16)	0.204 (0.17)	0.177 (0.19)	-0.380‡ (0.17)	-0.378‡ (0.17)
Japanese media use index		0.0156 (0.017)		0.104§§ (0.029)		-0.00179 (0.036)
Sample size	195	195	350	347	140	140

†Marginal effects from probit estimation of being a return migrant, estimated at the mean of the other variables. The sample is restricted to 18–59-year-olds. Sample weights were used. Robust standard errors are given in parentheses, clustered by using a sandwich estimator. Omitted categories for dummy variables are male, not married, less than university education and *yonse*.

‡ $p < 0.05$ .

§ $p < 0.1$ .

§§ $p < 0.01$ .

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. Less than 5% of highest qualifications were obtained abroad, so in most cases these education variables capture education before migration.

Table 9 shows that the three survey methods do give different pictures of how the return migrants compare with non-migrants. The stratified survey shows little selectivity, with an undergraduate education being the only variable that is significant at the 10% 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. On average, the probability of being a return migrant for a female is 0.13 (13 percentage points) higher than for males in the snowball survey, compared with no discernible difference from males in the stratified survey. The snowball survey

gives strong education selectivity, significant at the 1% level—individuals with postgraduate education are less likely to be return migrants, and with undergraduates 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 from those of 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 various methods

On conclusion of the survey efforts, we asked Sensus to provide us with their updated cost breakdown for the cost of carrying out the survey by using each survey method. The stratified survey and snowball survey have a 36-page questionnaire with just over 1000 variables, taking just over an hour to complete, compared with the three-page intercept questionnaire with 60–70 variables, taking an average of seven minutes to complete. The per household costs here include interviewer time and travel costs, but not an additional 14% for taxes and the administration fee for the survey firm.

Sensus estimated that the listing exercise cost \$2 per dwelling listed, and the follow-up household interviews of households that were identified as Nikkei in the listing cost \$80 each. Combining the listing and surveying, the total variable cost per household interviewed in the stratified survey was \$212, since many households had to be listed to identify one Nikkei household to interview. The snowball survey was estimated to cost \$100 each. Since households in the snowball survey are less geographically clustered than those that were 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 \$30 per questionnaire.

Thus adding on 14% in taxes and 20% 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 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 who are more closely tied to the Nikkei community than randomly sampled individuals who are 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 that are 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 that were used here are often applicable in migrant receiving countries, since migrants tend to cluster in certain geographic locations, and to 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 an intercept point where non-ethnically identified families of migrants congregate, but locations such as festivals, transportation hubs, money transmitting branches, churches and social support networks may be starting points.

So what do we conclude from this experiment? The first conclusion 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 who are more closely connected to the community. Secondly, as the results here and those in Bilsborrow (2007) show, snowball surveys of migrants or their families may be quite ineffective in practice at creating the long referral chains that are needed for this method to capture the target population. In our case, the snowball sample gave quite different (misleading compared with 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, although the intercept method does not provide a representative sample of the whole population, surveying many locations and using reweighting does help to 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 is those who attend community locations. This may be so when policy interventions will rely on these same types of locations to reach migrants. However, there appears to be no very close substitute for the more time-consuming and expensive two-phase stratified sampling in obtaining truly representative surveys.

## **Acknowledgements**

The survey on which this research is based was funded by ABN-Amro through the World Bank's, 'Knowledge for change' programme. We thank Ricardo Guedes and the survey team from Sensus for their hard work 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 co-ordinated 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. Finally, we thank the Joint Editor, Associate Editor and referee for very useful suggestions. The views that are 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 that they represent.

## References

- Beltrão, K. I. and Sugahara, S. (2006) Permanentemente Temporário: Dekasseguis Brasileiros no Japão. (Available from <http://www.scielo.br/pdf/rbepop/v23n1/v23n1a05.pdf>.)
- Bilsborrow, R. (2007) Designing surveys of international migrants: the needle in the haystack, or finding the right haystack?: a case study of Colombians in Ecuador. *Population Association of America A. Conf., March*. (Available from <http://paa2007.princeton.edu/download.aspx?submissionId=72073>.)
- Bilsborrow, R. E., Hugo, G., Oberai, A. S. and Zlotnik, H. (1997) *International Migration Statistics: Guidelines for Improving Data Collection Systems*. Geneva: International Labour Office.
- 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 *I Fenomeni Migratori nel Bacino Mediterraneo (Migration Movements in the Mediterranean Basin)* (eds L. Di Comite and M. De Candia). Bari: Cacucci.
- Bustamente, J., Jasso, G., Taylor, J. E. and Legarreta, P. T. (1997) Characteristics of Migrants: Mexicans in the United States. In *Mexico-U.S. Binational Migration Study Report*, pp. 91–162. US Commission on Immigration Reform. (Available from <http://www.utexas.edu/lbj/uscir/binational.html>.)
- Cameron, A. C. and Trivadi, P. K. (1998) *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.
- Fawcett, J. T. and Arnold, F. (1987) The role of surveys in the study of international migration: an appraisal. *Int. Migr. Rev.*, **21**, 1523–1540.
- Global Commission on International Migration (2005) Migration in an interconnected world: new directions for action. *Report*. Global Commission on International Migration. (Available from [www.gcim.org](http://www.gcim.org).)
- Goodman, L. (1961) Snowball sampling. *Ann. Math. Statist.*, **32**, 148–170.
- Goto, J. (2007) Latin Americans of Japanese origin (*Nikkeijin*) working in Japan—a survey. *Policy Research Working Paper 4203*. World Bank, Washington DC.
- Groenewold, G. and Bilsborrow, R. (2004) Design of samples for international migration surveys: methodological considerations, practical constraints and lessons learned from a multi-country study in Africa and Europe. *Population Association of America General Conf., Boston*. (Available from [http://www.irpps.cnr.it/ricmob/web\\_conf/download/groenewoldbilsborrow.pdf](http://www.irpps.cnr.it/ricmob/web_conf/download/groenewoldbilsborrow.pdf).)
- Heckathorn, D. D. (1997) Respondent-driven sampling: a new approach to the study of hidden populations. *Soc. Prob.*, **44**, 174–199.
- Heckathorn, D. D. (2002) Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Soc. Prob.*, **49**, 11–34.
- Higuchi, N. (2006) Brazilian migration to Japan: trends, modalities and impact. *United Nations Expert Group Meet. International Migration and Development in Latin America and the Caribbean, Mexico City, Nov. 30th–Dec. 2nd, 2005*. (Available from [http://www.un.org/esa/population/meetings/IttMigLAC/P11\\_Higuchi.pdf](http://www.un.org/esa/population/meetings/IttMigLAC/P11_Higuchi.pdf).)
- Kalsbeek, W. D. (1986) Nomad sampling: an analytic study of alternative design strategies. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 164–169. (Available from [http://www.amstat.org/sections/srms/Proceedings/papers/1986\\_028.pdf](http://www.amstat.org/sections/srms/Proceedings/papers/1986_028.pdf).)
- Kalton, G. (1991) Sampling flows of mobile human populations. *Surv. Methodol.*, **17**, 183–194.
- Kalton, G. (2001) Practical methods for sampling rare and elusive populations. In *Proc. A. Meet. American Statistical Association*. Alexandria: American Statistical Association. (Available from <http://www.amstat.org/sections/srms/Proceedings/y2001/Proceed/00454.pdf>.)
- Kalton, G. and Anderson, D. W. (1986) Sampling rare populations. *J. R. Statist. Soc. A*, **149**, 65–82.
- Kish, L. (1965) *Survey Sampling*. New York: Wiley.
- Lavalle, P. (1995) Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. *Surv. Methodol.*, **21**, 25–32.
- Massey, D. and Singer, A. (1995) New estimates of undocumented Mexican migration and the probability of apprehension. *Demography*, **32**, 203–213.
- McKenzie, D., Gibson, J. and Stillman, S. (2006) How important is selection?: experimental versus non-experimental measures of the income gains from migration. *Policy Research Working Paper 3906*. World Bank, Washington DC.
- Tsuda, T. (1999) The motivation to migrate: the ethnic and sociocultural constitution of the Japanese-Brazilian return-migration system. *Econ. Develpmnt Cult. Change*, **48**, 1–31.
- Tsuda, T. (2003) *Strangers in the Ethnic Homeland: Japanese Brazilian Return Migration in Transnational Perspective*. New York: Columbia University Press.
- Wasserman, M., Bender, D., Kalsbeek, W., Suchindran, C. and Mouw, T. (2005) A church-based sampling design for research with Latina immigrant women. *Popln Res. Poly Rev.*, **24**, 647–671.
- World Bank (2005) *Global Economic Prospects 2006: Economic Implications of Remittances and Migration*. Washington DC: World Bank.