Routledge
Taylor & Francis Group

Population Studies
A Journal of Demography

ISSN: 0032-4728 (Print) 1477-4747 (Online) Journal homepage: http://www.tandfonline.com/loi/rpst20

Tracking rural-to-urban migration in China: Lessons from the 2005 inter-census population survey

Avraham Ebenstein & Yaohui Zhao

To cite this article: Avraham Ebenstein & Yaohui Zhao (2015) Tracking rural-to-urban migration in China: Lessons from the 2005 inter-census population survey, Population Studies, 69:3, 337-353, DOI: 10.1080/00324728.2015.1065342

To link to this article: http://dx.doi.org/10.1080/00324728.2015.1065342

Population Studies, 2015
Vol. 69, No. 3, 337–353, http://dx.doi.org/10.1080/00324728.2015.1065342

Routledge
Taylor & Francis Group

Tracking rural-to-urban migration in China: Lessons from the 2005 inter-census population survey

Avraham Ebenstein[1] **and Yaohui Zhao**[2]
[1]Hebrew University of Jerusalem, [2]Peking University

We examined migration in China using the 2005 inter-census population survey, in which migrants were registered at both their place of original (hukou) residence and at their destination. We find evidence that the estimated number of internal migrants in China is extremely sensitive to the enumeration method. We estimate that the traditional destination-based survey method fails to account for more than a third of migrants found using comparable origin-based methods. The 'missing' migrants are disproportionately young, male, and holders of rural hukou. We find that origin-based methods are more effective at capturing migrants who travel short distances for short periods, whereas destination-based methods are more effective when entire households have migrated and no remaining family members are located at the hukou location. We conclude with a set of policy recommendations for the design of population surveys in countries with large migrant populations.

Keywords: China; migration; origin; destination; population surveys

[Submitted March 2013; Final version accepted May 2015]

1. Introduction

In 1958, China established the national *hukou* household registration system that continues to regulate internal migration today. The *hukou* system linked each individual to his or her home for the receipt of government services, and essentially forced rural citizens to live, work, and die where their family was registered. Two types of *hukou* were established, urban and rural, with each individual registered in his or her place of residence. These requirements created a major impediment to migration, with migration from rural to urban areas being particularly difficult. The *hukou* system prevented farmers from relocating to cities in search of higher wages and curtailed freedom of movement for China's massive rural population (Solinger 1997).

In the late 1970s, Chinese policy towards migration and the *hukou* system began to evolve. With the successful de-collectivization of farming, which resulted in massive increases in agricultural productivity, millions of farmers were no longer necessary for the country to meet its agricultural needs (Liang and White 1996). The economic reforms, which gathered momentum in the late 1980s and early 1990s, ushered in a period in which rural-to-urban migration was not only tolerated but promoted as necessary to provide cheap labour for China's growing cities. The marketization of food and other necessities made it possible for temporary migrants to live in cities, even without urban registration (Zhao 2000). While many public goods (e.g., schooling) continue to be available only to residents with local *hukou*, migrants today have greater latitude than those in earlier decades to decide when, where, and how to out-migrate according to labour-market conditions and their own preferences. As a consequence, millions of migrants are living in cities outside of their original *hukou* place, fuelling China's manufacturing boom and spectacular economic growth. However, owing to the hardship faced by rural migrants in cities, most regard themselves as 'temporary' residents of their destination cities (Zhao 2000; Cai et al. 2002). Indeed, they have the official status of 'temporary population' and even have a temporary registration card, since they do not have local household registration (*hukou*).

Much ink has been spilt analysing Chinese rural-to-urban migration, and its consequences for Chinese society. China's internal migration has, however, created a set of challenges for census officials charged with accurately enumerating the

country's large and highly mobile population (Chan 2001; Chan 2003; Zhang et al. 2005). Under the *hukou* system, census surveys could accurately record the population by enumerating each individual at the location of his or her *hukou*. However, for the 2000 census, there was a recognition that this survey method would fail to capture the 'floating population'—the millions of Chinese migrants who were living in cities and away from their *hukou* place. The 2000 census was carried out in a manner in which survey officials would record migrants at their 'usual residence', meaning that individuals who were away from their *hukou* place for more than 6 months would be recorded at their destination and migrants who had been away for less than 6 months would be recorded at the location of their *hukou* (Chan 2003). In consequence, the challenges of tracking the population given China's increased internal migration in the 1990s resulted in a nearly 30-fold increase in the census undercount relative to the 1990 census. China's National Bureau of Statistics (NBS) acknowledged 'massive deficiencies' in the execution of the 2000 survey and set out to improve the tracking of the population for the 2005 population survey and the planned 2010 census (Zhang et al. 2005).

One key innovation of the 2005 inter-census survey was the change in focus of the target population from 'usual resident', as had been used in the 2000 census, to 'current resident' (Zhang et al. 2005). This change was specifically designed to deal with inconsistencies in how census personnel decided whether to record migrants in the 2000 census at their current location or at the location of their *hukou*. Since the census personnel were instructed to determine the migrant's length of stay before deciding whether to record the migrant at his or her current location (rather than the location of his or her *hukou*), they often failed to account for migrants who should have been recorded at their current location and were therefore not recorded at either location. Furthermore, NBS officials suspected that many migrants avoided census representatives entirely because they feared that their responses to questions about length of stay would result in fines or expulsion from their city of residence. These challenges in enumerating migrants resulted in a fundamental change in the 2005 population survey: census officials were instructed to engage in full enumeration of migrants at *both* their *hukou* place and their current location, and to pay much greater attention to migrants' concerns over whether responses would have legal consequences.

In this paper, we present evidence that tracking migration in China may be very sensitive to internal decisions made by China's NBS that are not disclosed to the public. The evidence is based on an analysis of a micro-sample of the 2005 population survey of migrants at both their origin (*hukou*) and destination. The official communiqués issued by the NBS provide only a single estimate of the number of migrants, with very limited information on how the NBS calculate these estimates, and how they integrate information collected on migrants at their origin and destination. Using our sample, we estimated the size of the migrant population using several definitions of migration and analysed data collected both at the origin and destination of the migrants. This enabled us to investigate the sensitivity of the size of the migrant population to internal decisions by the NBS that were not disclosed to the public. It is important to note that our focus was on migration as defined by 'living away from one's *hukou*', which is to define it by legal residence status (so *de jure* migration) rather than by 'living in a new location' (*de facto* migration).

While NBS personnel do not reveal precisely how they generate their published figures, the Bureau's official publication states that their method counts all individuals living outside their township *hukou* as migrants, but excludes short-term migration (less than 6 months). Using this official definition of migration and applying it to our micro-data extract of the 2005 population survey, we find that the NBS method yields 167 million migrants found at their destination, whereas origin-based reporting produces 232 million migrants, a staggering difference that highlights the sensitivity of migrant counts to the method used. The official NBS published figure (147 million) for the 2005 survey may understate the true migrant population by roughly a third of the figure that would be yielded by a comparable origin-based estimate. As discussed later in the paper, we attribute this difference (167 million vs. 147 million) to sampling error, since we were using a sample that was roughly one-fifth of the original 2005 population survey. The other possible explanation is that the NBS reports estimates with adjustments that are not publicly disclosed. Its decision to report only a single figure is somewhat puzzling in light of the large difference between origin-based and destination-based measures and the issues and challenges prompted by this calculation (Zhang et al. 2005).

If the origin-based migration totals are to be taken at face value, the scale of omission using destination-based methods is so large that certain presumed facts about Chinese migrants need to be fundamentally reconsidered. For example, many recent academic

articles have cited the 2000 census official communiqué that migrants are made up of similar numbers of males and females. We find that while a traditional destination-based definition yields a 1.08 sex ratio, our origin-based migrant population has a sex ratio of 1.34, suggesting that men are significantly over-represented in the migrant population. Among the missing migrants, the sex ratio is even higher—this population by our calculation has a sex ratio of 1.99, indicating that missing male migrants outnumber missing female migrants by roughly 2:1. Our results also suggest that researchers need to exercise extreme caution when comparing the 2000 and 2005 surveys with earlier surveys. These challenges for analysing migrants will probably be a factor in interpreting the 2010 survey as well, though to our knowledge no copies of the 2010 micro-data have yet been made available for academic research.

As a result of the large differences between origin- and destination-based enumeration methods, we examined *why* these different methods of migrant enumeration generate such divergent results. We first examined the characteristics of the sample of migrants from records at the origin and at the destination, and then inferred who had been missed by census officials. By comparing the information on migrants at their origin and their destination, we were able to describe this population of 'missing migrants' in their demographic composition and their geographic distribution. Our data indicate that the missing migrants are disproportionately male, young, and current holders of rural *hukou*. Using detailed information on county of origin and county of residence, we found that origin-based migrants on average travel shorter distances than destination-based migrants and have been at their destination for shorter periods. This suggests that migrants who travel short distances and stay for short periods at their destination are often not found at the destination and in these circumstances a systematic origin-based system is likely to be more effective at tracking migration than a destination-based approach which only samples certain communities. We also present evidence that origin-based sampling fails to survey migration accurately in certain circumstances, such as when an entire family migrates. In this circumstance, no family member is at the *hukou* location to report the family's absence, and so a destination-based approach will provide more reliable data.

We conclude with an analysis of the geography of migration. We find the largest number of missing migrants in large industrial areas, such as Guangdong, which may have as many as an extra 15 million unrecorded migrants (40 million at the destination, 55 million at the origin). The analysis of migrants at their *hukou* place also reveals that certain 'sending' provinces, such as Anhui and Jiangxi, have experienced a population decline of over 20 per cent owing to net out-migration. In contrast, cities such as Beijing and Shanghai currently have net migration that increased the population by 38 and 25 per cent, respectively. These results indicate that proper urban planning in both sending and receiving communities requires accurate measures of migrant flows, using information collected both at the origin and at the destination. Our results highlight a serious limitation associated with exclusively destination-based or origin-based surveying methods in the measurement of rural-to-urban migration.

While the results indicate a systematic failure of census officials to track migrants at their destination in the 2005 survey, there are several caveats worth noting. First, the 2005 survey may have been particularly vulnerable to the differences generated by origin-based and destination-based sampling, and our 20 per cent subsample of the survey may have been more sensitive to these issues than the 100 per cent sample of the survey. Many of these issues of mismatch of origin-based and destination-based sampling are probably related to the sample size, and would probably be much less significant in a universal census. Second, destination-based sampling may be preferable in situations where officials are charged both with estimating the size of the migrant population and identifying the living and working situations of the population. Third, when an entire household migrates, it may be that there is no alternative to destination-based sampling since no one remains at the origin to complete a survey. In light of these considerations, we conclude the paper with a set of recommendations for tracking migration that adopts a dual approach that blends the strengths of both origin-based and destination-based sampling.

The paper is organized as follows. In Section 2, we present additional background information on origin-based and destination-based methods for surveying migrants. In Section 3, we present our tabulations of the 2005 census and the estimated numbers of migrants and implied migration rates using both definitions. In Section 4, we present a demographic assessment of China's migrants that we obtained using both methods in order to examine why origin-based and destination-based methods produce such different results. We then report the results of using both measures to examine the

geography of China's internal migration. We conclude in Section 5 with a brief discussion of the implications of our findings for survey design, migration policy, and future research on Chinese migration.

2. Background on origin-based and destination-based methods

An important innovation of the 2005 population survey is that the census personnel were instructed to enumerate migrants at both their origin and destination, irrespective of the length of time spent at either location. The Chinese statistical bureau defined the target population for the survey as 'current resident plus absent resident' (Zhang et al. 2005). At the origin, family members were instructed to list all individuals who had their *hukou* registration linked to the household but were currently residing elsewhere. At the destination, individuals filled out the standard census questionnaire and reported their *hukou* of origin if it was different from their place of residence. This double-count procedure was implemented for the 2005 survey because of significant undercounting of migrants in the 2000 census uncovered by the Chinese statistical bureau (Lavely 2001; Zhang et al. 2005).

The undercount of the 2000 census provided an important lesson for the design of subsequent surveys, such as the 2005 population survey and the 2010 census. In the 2000 census, in order to avoid possible double counting that came with increased population migration, enumerators were instructed to enumerate people at their current residence (i.e., where they had lived for 6 months or more). They were instructed *not* to enumerate migrants at their home if they had been absent from their *hukou* place for less than 6 months. If a migrant had been away from his or her *hukou* place for 6 months or more, the enumerator was instructed to include this individual at his or her current location, irrespective of length of stay. The method, although theoretically sound, led to bias because the screening necessary to decide whether a person had left the origin for 6 months or more before the survey created a difficult situation for the enumerators who had to decide. In both the 2005 population survey and 2010 census, records for migrants were collected at both their *hukou* place and current location irrespective of duration of absence from the *hukou* place. The 2010 census improved upon the 2000 census by enumerating every migrant without screening for length of absence, but recording the period for which the person had left the origin. This procedure enabled the NBS to deal with this issue in a consistent manner at the centralized data processing facility. Unfortunately, the 2010 census has not been made available for scholarship except in aggregate form.

Destination-based enumeration may under-report migrants for a variety of other reasons as well. First, household registration records are heavily relied upon in the enumeration process, despite the NBS requirement for person-to-person interviews using physical addresses derived from mapping/listing operations. Second, enumerators are prone to miss migrants residing in work sites and slum-like neighbourhoods in city suburbs. Since the majority of migrants maintain their household registration (*hukou*) records at their origin and family members are available to answer questions, the origin-based enumeration method is thought to provide an alternative count of these types of migrants which may capture some migrants missed at their destinations. Third, as described by Zhang et al. (2005), migrants may be wary of meeting census personnel in cities where they have no *hukou*, as they may fear punishment for residing in the city longer than allowed.

The 2005 population survey method of using both origin-based and destination-based sampling represented an important improvement on the existing decennial censuses in 1990 and 2000, in which migrants were generally recorded at their destination. As mentioned, the 2005 population survey accomplished this by breaking with previous surveys that only surveyed migrants at their current location if they had been away from their *hukou* place for a particular period of time. In the 2005 population survey, all migrants were enumerated at their current location, rather than only those that had been absent from their place of registration for more than 6 months (as was the case for the 2000 census) or 1 year (as was the case for the 1990 census). The extensive surveying of migrants at their destination irrespective of duration of absence was then complemented by surveying at the *hukou* location as well. So, when a family reported that a member was currently residing elsewhere, they were asked to report on the missing person's whereabouts. For the individuals who were away, the 2005 survey was designed to achieve a compromise: an extended questionnaire at the *hukou* location required for migrants away from their origin for more than 6 months and more limited information provided for migrants absent for less than 6 months.

3. Origin-based and destination-based migration totals and rates

Data

Our analysis was based on a micro-data extract of the 2005 population survey, which collected a 1.31 per cent sample of China's population (China National Bureau of Statistics 2012). The 2005 survey was nationally representative, drawing respondents from each of China's 2,861 counties. The survey was carried out in a set of 60,000 primary census enumeration districts that were selected using a three-stage cluster sampling method (Zhang et al. 2005). The sample used in our analysis was a subset of the original survey, which included 17.05 million persons; our sample contained 2,084,396 observations. Note that our sample included a double count of migrants, since it retained records of the migrant at origin and destination.

The sample was provided with probability weights generated by the NBS to account for their sampling framework, and all results are presented using these weights. The final sample comprised 2,080,671 individuals. The 2005 population survey, unlike the decennial census, was conducted with a probability weighting scheme that used the Public Security Bureau's 2004 population registry of individuals at their *hukou* place as its reference for sampling. The selection of township units as the Probability Sampling Unit (PSU) was conducted at the national level, using the Proportional Probability Sampling (PPS) method. The survey then selected Small Survey Units (village/residence committees) and Township Survey Units (small survey districts, *diao cha xiao qu*) at the province level, again using the PPS method. The sample design dictated that each small survey district would include 250 persons who had stayed in the district (*de jure* residents). The sampling method may have generated some mismatch between the actual resident population and the sampled population, creating an over-sampling of migrant-sending communities and an under-sampling of migrant-receiving communities.

Another issue is that the 2005 survey was conducted on a small sample of full census districts, which meant that the destination-based counts of migrants were based on a limited sample. What made this a problem was that while almost all migrants were drawn to a large city included in the survey, it is likely that many migrants were from villages or townships that were entirely missed by the survey. The possible mismatch is one which could only have been avoided if 100 per cent of the districts had been sampled. While we cannot know how our results would change if all districts had been sampled, the survey was designed to be nationally representative. Although our sample represented only 0.2 per cent of the national sample, provided the survey design was implemented appropriately, the 77,417 primary sampling units included in the 2005 survey meant that the sample was similar in composition to that of a census. Note also that the coverage included all 345 prefectures, and 2,869 counties, suggesting a wide coverage area for the 2005 survey. Nevertheless, the results are subject to the caveat that they may have been affected by the mismatch between surveyed destination districts and the full sample of census districts.

China's decennial census samples had very low undercounts before the increase in internal migration preceding the 2000 census, and it may be that decennial census surveys are somewhat less affected by these issues, since they do not rely on PPS sampling and every district is represented. In fact, the 2010 census added several questions specifically targeting migrants, and was a full 100 per cent sample of the population. This means that some of the challenges we highlight are more relevant to the results of the 2005 survey and other standard population surveys than to those of the 2010 census. However, as mentioned earlier, the 2010 census has not been made available to scholars except in aggregate form. While the 2005 survey has limitations, the issue of sampling bias arises for any population survey using the PPS method to investigate the extensive rural-to-urban migration common to many less developed countries. Our results can therefore be interpreted as the bias that emerges from population surveys of migration, but may overstate the bias in official published migration statistics.

In Table 1 we tabulate our sample of non-migrants and migrants, with migrants stratified by whether they were enumerated by census personnel at their origin or their destination. Row (A) lists local residents (non-migrants) still currently in the township of their *hukou*. Rows (B) and (C) are migrants reported by household members (or others) at the origin. In rows (D) and (E) we report the tabulation of migrants found at their destination.

Note that in this study, we were using *hukou* status to define migrants rather than 'actual' migration status. For example, if migrants had transferred their *hukou* to their destination, they would have been recorded in row (A), in spite of having made a permanent move (Shen 2013). This highlights the

Table 1 Tabulations of population survey using both origin and destination records, China 2005

		All		Males		Females	
		N	Share	*N*	Share	*N*	Share
(A)	*Hukou* here and live here	1,466,685	0.70	740,288	0.69	726,396	0.72
(B)	*Hukou* here and away <6 months	23,268	0.01	13,551	0.01	9,717	0.01
(C)	*Hukou* here and away 6+ months	320,911	0.15	175,969	0.16	144,943	0.14
(D)	Live here and left *hukou* <6 months	38,963	0.02	20,702	0.02	18,261	0.02
(E)	Live here and left *hukou* 6+ months	230,844	0.11	116,680	0.11	114,164	0.11

Notes: The five categories which classify census respondents by migration status are reported in rows (A)–(E). The tabulation uses a micro-sample extract of the original survey in which records were collected from survey respondents at both their current location and at their registration location (*hukou*). In this tabulation, *hukou* refers to the township of the *hukou* location. Note that this sample over-counts China's population because migrants are recorded at both their origin and destination. Results are adjusted for sampling weights.
Source: China Inter-census Population Survey (2005).

problem that arises from the lack of any official NBS definition of migrant. The NBS has instead used such terms as '*liudong renkou*', which refers usually to the 'floating population'. The NBS has never settled on a single definition of migration, a fact that complicates the task of quantifying or characterizing migrants. In our study, we used a range of possible definitions to examine migration in China. We also used both origin-based and destination-based counting to characterize the population of migrants more accurately, with metrics that included distance from *hukou* place and length of absence. Since our data relied on *hukou* status, however, we were essentially examining *hukou* migration. We were also focusing our attention on the stock of *hukou* migrants, data on which were included in the questionnaire, in contrast to studies that examine the flow of migrants who had relocated within the previous 5-year period.

Our estimate of the number of migrants in China was contingent on the enumeration method employed, and the calculation of the adjustment factor to inflate our sample size to correspond to China's population was contingent on choosing which migrant enumeration method was more trustworthy. The interpretation of the number of migrants also required a decision on which measure of migration was accurate and how long a stay away from the *hukou* qualified an individual to be described as a migrant. In the migration totals in the next section we present all our results based on the origin-based population measure, in which migration is defined as being away from the *hukou* for 6 months. In consequence, each of our observations represented 722 persons, which was the ratio of China's official population in 2005 (1.3067 billion) to the 1,810,864 residents to whom our preferred definition of migration applied. The issue of what makes a definition preferable is discussed in the next section.

Estimating the number of migrants

A fundamental question in tracking migration is how to define a migrant. Boundaries crossed, length of stay at destination, and legal registration status all have relevance to the definition. Table 2 shows how the number of migrants we studied changes according to the definition used. Panel 1 includes migrants who have migrated for any period of time. Panel 2 excludes those who were away from their *hukou* for less than 6 months, to match the definition of migrants in NBS census communiqués. The columns show the number of migrants by whether the measure was origin-based or destination-based, and whether they migrated within or beyond city districts.

Table 2 shows that the most inclusive definition of migration yields 248 million migrants (origin-based, any duration, living outside of township of *hukou*) while the most exclusive definition yields 68 million migrants (destination-based, more than 6 months, living outside of province of *hukou*). Based on the definition of migration used by the NBS for the 2005 population survey (destination-based, at least 6 months, living outside of township of *hukou*), we estimate from our sample that China has 167 million migrants, which is 20 million more than the 147 million migrants reported by the NBS in the official communiqué of the survey's results. We assume that this difference is attributable to our use of only a 20 per cent subsample of the original data, and to discrepancies in method that cannot be ascertained from the information reported by the NBS. Note that while the NBS chooses to exclude migrants who have been living away from their *hukou* location for less than 6 months, it chooses to include any instances of intra-county migration in the migration total. When migration is defined by travelling

Table 2 Migration totals (millions) using alternative definitions of migration, China 2005

	Including migration within city districts		Excluding migration within city districts	
	Origin-based (1)	Destination-based (2)	Origin-based (3)	Destination-based (4)
Panel 1: Totals using broad inclusion of migrants of any duration away from hukou				
Living outside of township	248	195	203	142
Living outside of county	172	135	160	114
Living outside of prefecture	148	109	148	109
Living outside of province	98	82	98	82
Panel 2: Totals excluding migration of duration less than 6 months away from hukou				
Living outside of township	232	167	189	120
Living outside of county	162	115	151	96
Living outside of prefecture	140	91	140	91
Living outside of province	93	68	93	68

Notes: Sample comprises persons defined as migrants either interviewed at their destination or reported away from home by family members. In Panel 1 we report migration totals of any duration, stratified by whether the calculation was performed using origin-based or destination-based migration, and whether within-city migration is recorded as migration. In Panel 2 we restrict our sample to migration durations of longer than 6 months. Official NBS calculations include those who leave their township and intra-city movements, but exclude migration where the duration away from the *hukou* is less than 6 months. Results are adjusted for sampling weights.
Source: As for Table 1.

longer distances (e.g., those who have left their *hukou* county), the number of migrants they report shrinks to 115 million. Thus, while the NBS chooses to use the lower-count destination-based estimates for counting migrants, they also choose to use a relatively inclusive definition of how far migrants need to travel to be considered migrants. If migration is counted when it refers to inter-province movements only, the number of migrants shrinks to 68 million using a destination-based approach, and 93 million using an origin-based approach.

In the remainder of the paper we present results using the 'left county' definition of migration, for several reasons. First, this is generally the standard used in the literature to define internal migration in China (Chan et al. 1999; Fan 2005). Second, since townships are generally small geographic areas, movements across a township can often simply be the result of normal housing choices rather than migration in the traditional sense. We exclude intra-city moves among urban residents and those who have been away from their origin for fewer than 6 months. We also exclude tourists and short-term visitors, who are also excluded by the NBS. Using our preferred definition, the origin-based number of migrants is 151 million and the destination-based number 96 million, a difference of over a third. We again note that this gap does *not* imply that 151 million is the 'correct' number of migrants in China, since it is predicated on a variety of decisions about how to interpret the data. However, we do argue that origin-based counting is generally more accurate and for this reason we prefer this estimate to a destination-based total.

Calculating migration rates

In Table 3, we show migration rates in China for those who currently hold urban and rural *hukou* separately by origin-based and destination-based definitions. These calculations are based on the reported *hukou* at either the origin or destination. Note that if migrants are able to obtain urban *hukou* following their migration, this will generate a mismatch between those included in each sample. However, since conversions are relatively rare for rural migrants (except in limited circumstances such as when joining the military), and transferred *hukou* should be removed from household registration upon conversion to urban *hukou*, this should not greatly affect the results. Also note that this calculation is performed excluding within-city and intra-township migration, corresponding to row 2 and columns (3) and (4) of Table 2. As shown in the table, the 150.6 million origin-based migrants comprise 14.0 million urban and 136.6 million rural migrants, and the 95.8 million destination-based migrants comprise 19.9 million urban and 75.9 million rural migrants. This implies that while there are 60.7 million missing migrants among those with rural *hukou*, among those with urban *hukou* there are 5.9 million more destination-based migrants than origin-based migrants. This is also reflected in migration rates, which show that numbers of

Table 3 Origin vs. destination-based migration rates using alternative definition of migration

		All		Urban		Rural	
		Origin-based (1)	Destination-based (2)	Origin-based (3)	Destination-based (4)	Origin-based (5)	Destination-based (6)
(A)	Number of migrants (millions)	150.6	95.8	14.0	19.9	136.6	75.9
(B)	Total population (millions)	1,307.6	1,307.6	356.7	356.7	950.9	950.9
(C)	Migration rate	0.115	0.073	0.039	0.056	0.144	0.080
(D)	Implied 'missing' migrants (millions)		54.8		−5.9		60.7
(E)	Sex ratio of migrants	1.34	1.08				
(F)	Sex ratio of missing migrants	1.99					

Notes: Sample comprises persons defined as migrants either interviewed at their destination or reported away from home by family members. Totals correspond to our preferred choice of migrant definition from columns (3) and (4) in Table 2, where we exclude intra-city district movements and migration within a county and include migration duration of less than 6 months. Total population figures reported in row (B) are based on the total number of observations in our sample minus migrants reported at their destination. In row (C) we report the migration rates under the two definitions, using (B) as our estimate of China's population. In row (D) we calculate missing migrants assuming that the origin-based calculation is accurate and destination-based surveying fails to account for all actual migrants. In row (E) we report the sex ratio of migrants using either the origin-based or destination-based definition. Results are adjusted for sampling weights.
Source: As for Table 1.

migrants reported by origin-based and destination-based methods are in the ratio 0.039 to 0.056 for those with urban *hukou* but in the ratio of 0.144 to 0.080 for those with rural *hukou*. This suggests that most internal migration in China is among those with rural *hukou*, and that the challenge of enumerating migrants at their destination is for those with rural *hukou*. For urban migrants, the challenge is enumerating them fully at their origin.

There are two main reasons for the undercount of urban migrants at their origin. Older urban migrants are usually economically established when they move

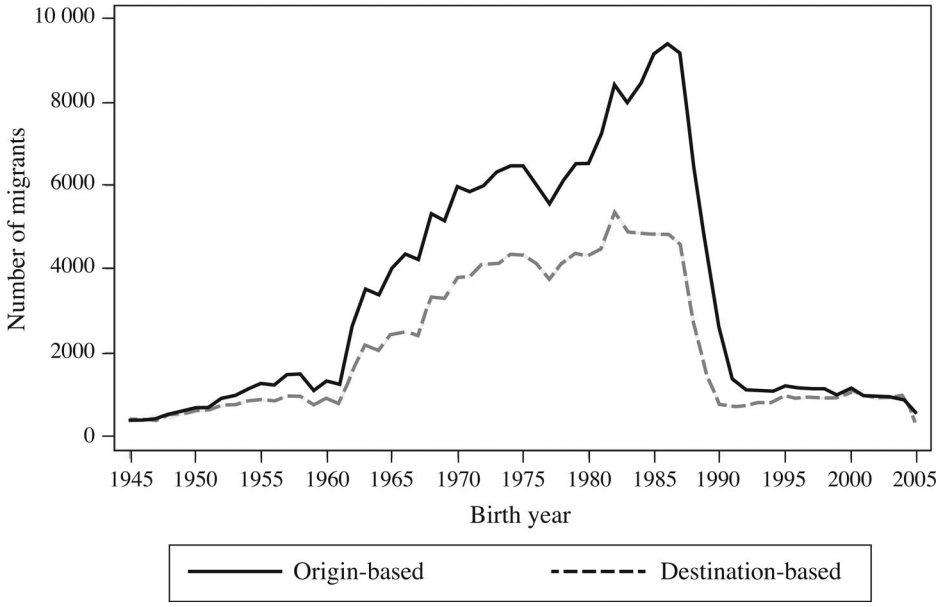

Figure 1 Migration estimates by origin-based vs. destination-based definitions, China 2005
Notes: The sample defines an 'origin-based' migrant as one reported by the household as a member who was temporarily away. The sample defines a 'destination-based' migrant as an individual who reports a different *hukou* of usual residence.
Source: As for Table 1.

to another city and take their whole family. They do not maintain a physical residence at the origin, which means that they will not be enumerated there. Before being granted *hukou* at their destination, they will be counted there as migrants and will not be recorded at their origin. More commonly, however, urban migrants are young, and often recent college graduates from rural areas. When they enrol in colleges (which are all located in cities), their *hukou* may or may not be transferred out of their parents' household and kept collectively at the university while they study. At graduation, if a student has not obtained an urban job that grants local *hukou*, he or she can keep urban *hukou* through their former university for a short period of time (6 months–2 years, depending on the specific university rules). Once this period has expired, the individual can choose to move the *hukou* from the university to the Labour Market Service Agency of the same city, and, for an annual fee, keep the collective urban *hukou* status of that city. In either circumstance, many of these graduates go on to work in other cities but cannot move their *hukou* with them. Thus, when the population census is carried out, the parents of these migrants do not report them to census enumerators because the children are no longer listed as household members on their *hukou* roster. Nor is either the university or the Labour Market Service Agency at the origin of the *hukou* responsible for reporting these graduates' whereabouts. The consequence is that more urban migrants are found at the destination than at the origin.

The undercount of rural migrants at their destination is considered at length in the next section, in which we examine the spatial distribution and demographic composition of the undercounted migrants.

4. Explaining the difference in results between origin-based and destination-based methods

Demographic characteristics of migrants using origin-based and destination-based methods

We begin our consideration of the differences in results between origin-based and destination-based methods by examining the demographic composition of the migrants captured by each method. As shown in Figure 1, the two methods yield very different numbers of migrants, with the differences most pronounced among the migrant population born 20–25 years before the 2005 population survey. The figure indicates that there are large numbers of migrants in cohorts born during the early 1990s that are not found by census officials at their destination. These 'missing migrants' are also disproportionately male.

As shown in Figure 2, the number of 'missing migrants' for each cohort is higher among male migrants, with the peak differences again occurring among young cohorts born 20–25 years before the 2005 survey. Note, however, that we observe missing migrants among cohorts who are in their working years, aged 20–60. Since our data also include information on county of destination *and* origin, we were also able to examine the distance travelled by migrants separately for the destination-based and origin-based migrants.

In Figure 3, we report the average distance travelled using both methods, by birth cohort. The results show that origin-based migrants have travelled shorter distances than destination-based migrants for all cohorts, and that the difference is particularly pronounced for younger migrants (aged 20–30). The results suggest that destination-based surveys are failing to track young migrants accurately, with young male migrants travelling short distances being those most likely to be missed. Table 4 presents the results of a more rigorous statistical analysis of the differences between migrants captured at their origin and those captured at their destination.

Table 4 shows detailed summary statistics for our sample of individual migrants, stratified by whether they were observed at the origin or the destination. We observe that origin-based migrants are on average younger than destination-based migrants, with the average age of the two groups being 28.61 and 30.99. Origin-based methods are also more likely to capture male migrants: while 61 per cent of origin-based migrants are male, only 51 per cent of the destination-based migrants are male. In addition we find that the origin-based migrants travelled shorter distances than destination-based migrants (534 kilometres vs. 586 kilometres). While 68 per cent of destination-based migrants have left their province, only 57 per cent of origin-based migrants have done so. Similarly, 94 per cent of migrants found at their destination have left their prefecture but only 88 per cent of origin-based migration is across prefecture boundaries. Perhaps not surprisingly, origin-based migrants are also spending less time at their destination: 2.22 years vs. 3.67 years. This suggests that destination-based definitions are more likely to capture migrants who migrate longer distances for longer periods of time, while origin-based definitions are more likely to

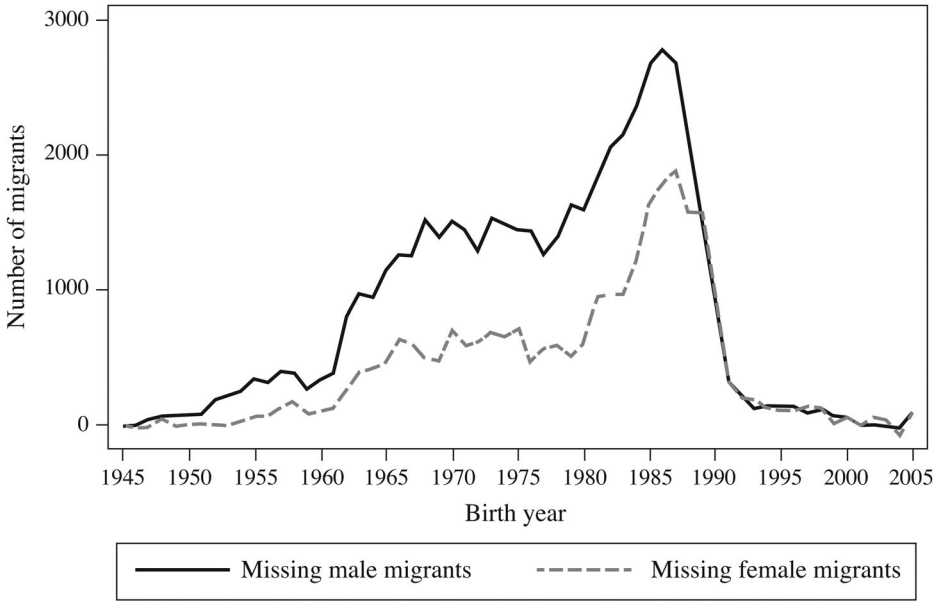

Figure 2 Missing migrants by sex, China 2005
Notes: Migrants are defined as 'missing' if the number of origin-based migrants by age group and sex exceeds the number of destination-based migrants by age group and sex. This includes both urban and rural migrants and uses our preferred definition described in Table 3.
Source: As for Table 1.

capture migrants who migrate shorter distances and for shorter periods of time. The duration of migrant stays is investigated further in the next section.

We were also able to observe differences in dwellings and household characteristics when migrants were surveyed at their origin rather than their

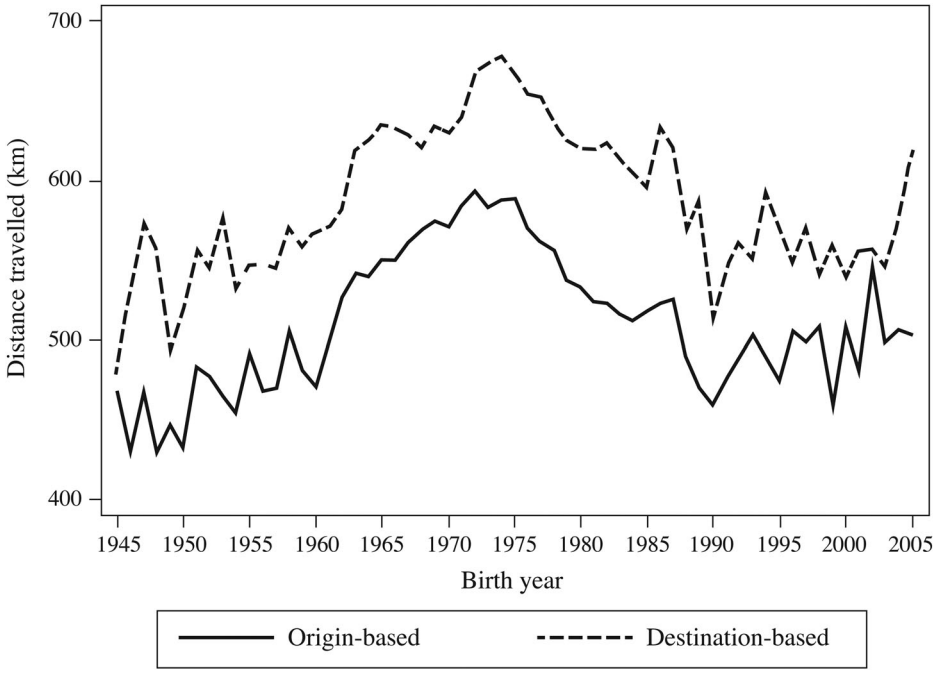

Figure 3 Distance between home and destination, origin vs. destination-based migration, China 2005
Notes: Migrants are defined as 'missing' if the number of origin-based migrants by age group and sex exceeds the number of destination-based migrants by age group and sex. This includes both urban and rural migrants and uses our preferred definition described in Table 3.
Source: As for Table 1.

Table 4 Sample averages among origin-based and destination-based migrants, China 2005

	Origin-based migration (1)	Destination-based migration (2)	Difference (3)
Individual information			
Age	28.61 (0.04)	30.99 (0.08)	−2.38*** (0.063)
Fraction male	0.61 (0.00)	0.51 (0.00)	0.10*** (0.002)
Time out (years)	2.22 (0.01)	3.67 (0.01)	−1.45*** (0.013)
Distance travelled (km)	534.41 (1.82)	586.49 (3.59)	−52.08*** (2.791)
Left prefecture (1 = yes)	0.88 (0.00)	0.94 (0.00)	−0.06*** (0.001)
Left province (1 = yes)	0.57 (0.00)	0.68 (0.01)	−0.10*** (0.002)
Household information			
Whole family move	0.02 (0.00)	0.16 (0.00)	−0.14*** (0.001)
Fraction renting	0.01 (0.00)	0.58 (0.00)	−0.56*** (0.002)
House price (000s yuan)	26.92 (0.28)	109.01 (2.52)	−82.09*** (0.779)
Monthly rent (yuan)	92.00 (4.86)	374.07 (3.57)	−282.07*** (12.973)

*Significant at 10 per cent; **significant at 5 per cent; ***significant at 1 per cent.
Notes: Sample comprises all persons defined as migrants as either being surveyed at their destination or reported away from home by family members, using our preferred definition of migration. Housing price is available for owner-occupied housing and monthly rent is available for renters. Data on housing characteristics are taken from the migrant's origin in column (1) and at the destination in column (2). Results are adjusted for sampling weights.
Source: As for Table 1.

destination. Table 4 presents evidence that migrants are more likely to be reported at their destination if the *entire* household has migrated. This is reasonable because it is less likely in this circumstance that census personnel would effectively identify the members of the *hukou* who were absent. The finding also provides insight into why origin-based migrants are travelling shorter distances for shorter periods: it is because these migrants may be married and wish to visit their families. A comparison of the information provided by households in the two groups also shows that migrants are generally leaving rural areas that have lower-cost owner-occupied housing for more expensive areas where people generally rent. The average home prices quoted at the origin is 26.92 thousand yuan, but at their destination it averages 109.01 thousand yuan. The fraction of renters is much higher at migrant destinations than at their origin, 58 per cent vs. 1 per cent. Since migrants are generally moving from relatively sparsely populated rural areas to denser urban areas, they may be forced to live in suburban or lower-quality housing, which is less likely to be surveyed by census officials.

Duration of migration among origin-based and destination-based migrants

As reported in Table 5, the missing migrants are primarily short-term migrants who have recently moved to their destination city. We find that 74 per cent of male and 95 per cent of female missing migrants report their duration of migration as between 6 and 12 months. The degree of undercounting generally declines as the duration of migration increases, with longer-term migrants more likely to be found at their destination. In fact we found that when the duration was longer than 5 years, the number of migrants found at the destination exceeded those reported at the origin. This may be because long-term migrants tend to move with their entire family, leaving no household member to act as a survey respondent at the origin. The table also shows that for migration spells lasting less than 6 months, the destination-based measure exceeds the origin-based measure. One possible explanation for this finding is that in order to reduce their workload, census enumerators wrongly classify destination-based migrants as having been at the destination for less than 6 months. Another possible reason is that migrants, fearing that their answers to a survey could lead to fines or expulsion, provide incorrect information. A third possibility applies in the case of the large numbers of missing migrants who had migrated less than 6 months earlier than the survey. It is that when recent college graduates move to work in a different city, they relocate without taking with them the *hukou* of their destination city, which leads to more migrants being found at the destination than at the origin. As shown in the table, our total number of missing migrants, which is the total of migrants in all these categories, understates the gross error in migration tracking by duration. The essential problem is that it is very difficult for census personnel to enumerate migrants and to solicit accurate information on the duration of their stays.

Table 5 Missing migrants in China by duration of absence and sex (000s)

	Males				Females			
	Origin-based (1)	Destination-based (2)	Missing migrants (3) (1) − (2)	Per cent of total (4) (3)/SUM (3)	Origin-based (5)	Destination-based (6)	Missing migrants (7) (5) − (6)	Per cent of total (8) (7)/SUM (7)
Living in their *hukou*	1	1	−1	0.00	4	1	3	0.02
Less than 6 months	5,824	9,926	−4,103	−12.68	3,802	7,919	−4,117	−28.92
6–12 months	33,770	9,861	23,908	73.92	22,662	9,163	13,499	94.83
1–2 years	17,934	9,398	8,536	26.39	14,088	9,296	4,792	33.67
2–3 years	12,710	6,981	5,729	17.71	10,359	7,046	3,313	23.27
3–4 years	7,517	4,653	2,864	8.86	5,963	4,573	1,389	9.76
4–5 years	3,599	3,198	400	1.24	2,957	3,045	−88	−0.62
5–6 years	2,224	3,115	−890	−2.75	1,815	2,776	−961	−6.75
More than 6 years	8,402	12,504	−4,102	−12.68	6,600	10,194	−3,595	−25.25
Totals	91,980	59,638	32,342	100.00	68,249	54,014	14,235	100.00

Notes: Migration calculations based on census question for all origin-based and destination-based migrants requesting 'time spent living away from their *hukou*'.
Source: As for Table 1.

A more fundamental challenge in assessing the duration of migration is posed by migrants who relocate frequently, and who may have difficulty interpreting the questions appropriately. For example, when migrants temporarily reside at multiple locations for a total of more than 6 months but spend less than 6 months at any given location, they may not report that they have been away from the origin for more than 6 months, and would therefore not be counted at their destination. Enumerators will face great challenges in properly accounting for these migrants, who are partly responsible for the existence of 'missing migrants' and contribute to the gap between origin-based and destination-based migrant tabulations. The existence of these 'circulators' emphasizes the challenge of defining migration appropriately in China.

Spatial distribution of origin-based and destination-based migrants

Table 6 compares the number of migrants yielded by the origin-based and destination-based measures across China's provinces, using our preferred definition of migration. As expected, the table shows large numbers of missing migrants in provinces specializing in manufacturing, such as Guangdong or Zhejiang. It is also worth noting that some provinces, such as Tianjin and Shanghai, have *negative* numbers of missing migrants. We propose several explanations for this. First, it may be that for these areas, family members erroneously reported the destination province. For example, a migrant plans on settling in Beijing and this is what the family report, but in fact the migrant finds work in nearby Tianjin. As a result, the origin-based measure of migration will understate the true numbers of migrants in Tianjin. A second possibility is that these cities have more full-family moves, which results in no one living at the origin *hukou* to report the migration. A third possible reason is that census personnel in these places are more effective in enumerating migrants. For example, Beijing and Tianjin reported a similar number of migrants (found at the destination), which is highly unlikely to be an accurate account of the relative size of the migrant population in each city. A fourth reason is that cities which attract a larger number of recent college graduates (such as Tianjin) may have larger numbers of migrants found at the destination than at the origin. Since there is a great deal of variation in how different cities treat the 'local *hukou*' of new college graduates, this may generate variation across cities in the recorded population of migrants.

The tabulation highlights the difficulty for census personnel of enumerating migrants and obtaining accurate information on current location from

Table 6 Origin vs. destination-based migration totals (in 000s)

Province (1)	Origin-based (2)	Destination-based (3)	Missing migrants (4) (2) – (3)	Missing migrant share (per cent) (5) (4)/SUM(4)
Beijing	5,447	4,643	804	1.51
Tianjin	1,300	4,299	–2,998	–5.61
Hebei	2,469	916	1,554	2.91
Shanxi	3,133	1,406	1,727	3.23
Neimenggu	2,103	1,671	432	0.81
Liaoning	2,588	1,546	1,043	1.95
Jilin	934	978	–44	–0.08
Heilongjiang	1,101	1,199	–98	–0.18
Shanghai	6,673	9,937	–3,264	–6.11
Jiangsu	7,465	3,591	3,874	7.25
Zhejiang	13,113	4,687	8,426	15.77
Anhui	1,532	685	847	1.59
Fujian	6,411	3,138	3,273	6.12
Jiangxi	1,121	526	595	1.11
Shandong	4,283	1,664	2,619	4.90
Henan	2,296	427	1,869	3.50
Hubei	2,625	1,175	1,450	2.71
Hunan	1,748	957	791	1.48
Guangdong	54,726	39,703	15,023	28.12
Guangxi	2,370	903	1,467	2.75
Hainan	1,447	1,131	317	0.59
Chongqing	1,620	641	978	1.83
Sichuan	4,203	1,336	2,867	5.37
Guizhou	1,777	825	951	1.78
Yunnan	4,828	2,896	1,932	3.62
Xicang	562	241	321	0.60
Shanxi	3,977	1,192	2,784	5.21
Gansu	1,895	856	1,039	1.94
Qinghai	858	717	141	0.26
Ningxia	1,097	491	606	1.13
Xinjiang	3,086	1,432	1,655	3.10
Abroad or n/a	452	0	452	0.85
Total	149,241	95,810	53,431	100.0

Notes: Sample comprises all persons defined as migrants as either being surveyed at their destination or reported away from home by family members, using our preferred definition. These numbers do not exactly match the origin totals in Table 3 owing to a small number of international migrants who were enumerated at their *hukou* or individuals for whom destination province was unavailable.
Source: As for Table 1.

family members. On balance, however, the number of origin-based migrants exceeds the number of destination-based migrants in the vast majority of provinces, and the results suggest that migration is understated in China owing to the failure of census personnel to locate the migrants in their destination city.

In Figure 4, we complement the duration analysis with a snapshot of China, highlighting the areas with the largest number of 'missing migrants'. Large coastal manufacturing cities, like Shenzhen and Guangzhou, appear to attract large numbers of migrants who are not found in the destination cities themselves. This is consistent with our demographic analysis, since many manufacturing workers may be short-term residents of their destination cities, and return home after a short period of work.

In Table 7, we consider net migration across provinces using the two measures. For each province, the table shows the origin-based and destination-based measure of both the destination of the migrant and the location of his *hukou*. Note that for this exercise, international migration, which has become increasingly common in China over the last decade, is only available at the origin and not at the destination (Liang and Chunyu 2013). The table reflects large discrepancies between the spatial distribution of migrants and the exodus from certain provinces. For example, our data show that Zhejiang's origin-based migrant population is a third of the province's total population, but a much smaller share are migrants according to those found by local enumerators. In Shanghai, the reverse is found—the

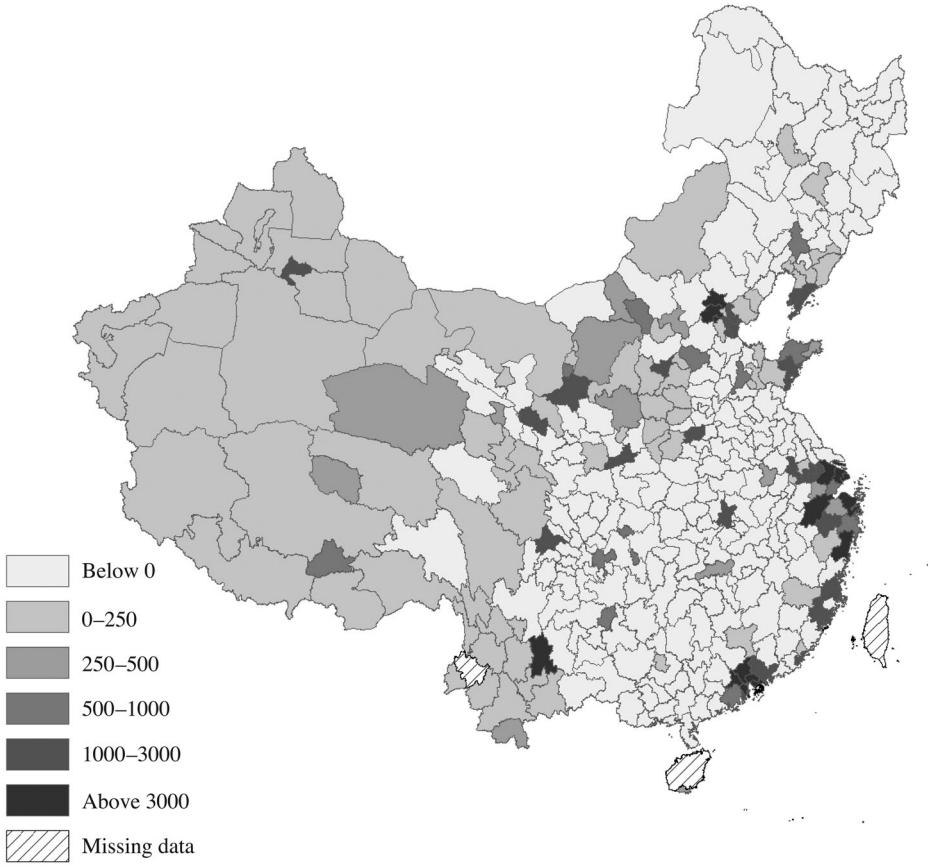

Figure 4 Geographic distribution of missing migrants by prefecture, China 2005
Notes: Migrants are defined as 'missing' if the number of origin-based migrants by age group and sex exceeds the number of destination-based migrants by age group and sex. This includes both urban and rural migrants and uses our preferred definition described in Table 3. The figure reports the number of observations of missing migrants by prefecture.
Source: As for Table 1.

migrants surveyed within the province itself are more than a third of the population (10 million migrants and 23 million locals), but when one relies on the reports at the *hukou*, this shrinks to a smaller share, with only 6.7 million migrants reported at the origin. The actual size of Shanghai's population is affected as well. If census personnel rely on migrant counts at their destination and the 2005 survey totals are taken at face value, Shanghai has a population of over 33 million. However, if migrants are excluded, Shanghai's population is reported to be only 23 million. Note that both these estimates are higher than the actual population of Shanghai, as a result of our use of a single sample weight for our survey, and relate to the idiosyncrasies of our sample. However, the results reflect the importance of accurate tracking of migrants both for understanding the dynamics of the migrant population and for proper urban planning for the needs of the whole population.

5. Policy recommendations and concluding remarks

China has begun to re-evaluate the *hukou* system and to consider revising its policy towards it. At the 2013 National People's Congress, public officials called for deep reforms to the system, including the complete elimination of the urban–rural distinction for granting *hukou*. Nevertheless, for the foreseeable future, many migrants are very likely to continue living in this limbo status, and this makes it crucially important that they should be effectively tracked. In this paper we have examined both origin-based and destination-based tabulations, and found limitations in both measures. We have identified large gaps in the coverage of migrants using destination-based enumeration, with nearly a third of migrants missed relative to the number counted by origin-based methods. While some of this difference may be attributable to sampling error, the results suggest that destination-based

Table 7 Net inflow and outflow of migrants by province, using origin-based and destination-based methods (in 000s)

	Origin method					Destination method				
Province (1)	Arriving (2)	Departing (3)	Deficit (4) (2) − (3)	Resident population (5)	Net migration share (6) (4)/SUM(5)	Arriving (7)	Departing (8)	Deficit (9) (7) − (8)	Resident population (10)	Net migration share (11) (9)/SUM(10)
Beijing	5,447	282	5,165	13,408	38.52	4,643	215	4,427	13,408	33.02
Tianjin	1,300	493	807	29,167	2.77	4,299	180	4,118	29,167	14.12
Hebei	2,469	2,212	257	36,938	0.70	916	2,679	−1,764	36,938	−4.77
Shanxi	3,133	2,381	751	50,924	1.48	1,406	1,294	112	50,924	0.22
Neimenggu	2,103	2,274	−170	18,830	−0.90	1,671	1,676	−5	18,830	−0.02
Liaoning	2,588	1,956	632	27,433	2.30	1,546	1,363	183	27,433	0.67
Jilin	934	1,421	−487	30,730	−1.59	978	1,386	−409	30,730	−1.33
Heilongjiang	1,101	1,712	−610	28,404	−2.15	1,199	2,679	−1,480	28,404	−5.21
Shanghai	6,673	742	5,931	23,226	25.54	9,937	330	9,607	23,226	41.37
Shanghai	6,673	742	5,931	23,226	25.54	9,937	330	−669	36,151	−1.85
Zhejiang	13,113	4,614	8,499	25,993	32.70	4,687	2,220	2,466	25,993	9.49
Anhui	1,532	9,943	−8,411	33,513	−25.10	685	7,008	−6,323	33,513	−18.87
Fujian	6,411	5,003	1,407	20,351	6.92	3,138	2,398	740	20,351	3.64
Jiangxi	1,121	7,712	−6,592	29,443	−22.39	526	5,190	−4,664	29,443	−15.84
Shandong	4,283	4,464	−181	54,660	−0.33	1,664	3,191	−1,527	54,660	−2.79
Henan	2,296	6,117	−3,821	42,274	−9.04	427	6,285	−5,858	42,274	−13.86
Hubei	2,625	8,924	−6,299	40,383	−15.60	1,175	5,727	−4,551	40,383	−11.27
Hunan	1,748	6,871	−5,123	40,291	−12.71	957	8,318	−7,361	40,291	−18.27
Guangdong	54,726	22,147	32,578	120,987	26.93	39,703	10,232	29,471	120,987	24.36
Guangxi	2,370	8,106	−5,735	32,249	−17.79	903	5,598	−4,694	32,249	−14.56
Hainan	1,447	1,489	−41	14,896	−0.28	1,131	816	315	14,896	2.11
Chongqing	1,620	6,969	−5,350	26,225	−20.40	641	3,239	−2,597	26,225	−9.90
Sichuan	4,203	13,539	−9,336	42,740	−21.84	1,336	9,046	−7,710	42,740	−18.04
Guizhou	1,777	6,565	−4,788	29,978	−15.97	825	3,082	−2,256	29,978	−7.53
Yunnan	4,828	5,975	−1,146	66,444	−1.73	2,896	2,161	736	66,444	1.11
Xicang	562	291	271	7,540	3.60	241	119	122	7,540	1.62
Shanxi	3,977	7,034	−3,057	50,236	−6.09	1,192	2,194	−1,001	50,236	−1.99
Gansu	1,895	4,075	−2,180	41,416	−5.26	856	1,419	−564	41,416	−1.36
Qinghai	858	1,013	−155	14,579	−1.06	717	541	176	14,579	1.21
Ningxia	1,097	942	156	10,495	1.48	491	367	124	10,495	1.18
Xinjiang	3,086	937	2,149	18,441	11.65	1,432	597	835	18,441	4.53
Abroad or n/a	452	0	452	0	0.00	0	1	−1	0	0.00
Total	149,241	150,599	−1,357	1,058,344	0.0	95,810	95,810	0	1,058,344	0.0

Note: The resident population was calculated using those living in their *hukou*. Migrants were calculated using the preferred definition of migration that is used in Table 3. The national population totals for each province do not exactly match the actual population.
Source: As for Table 1.

enumeration alone is inadequate for the proper accounting of China's migrant population, especially for rural migrants. In order to obtain accurate and complete information on migrants, collecting information at the destination as well as the origin can provide a fuller picture of the dynamics of China's internal migration. Thus, we recommend that future census samples undertake a dual approach, one in which both origin-based and destination-based rosters are completed in some locations. This procedure would certainly create a burden for census officials and it may be unrealistic to consider applying it to the full national sample, but our results from the 2005 population survey indicate that either method in isolation fails to characterize China's migrant population accurately. As shown in the 2005 survey, a dual approach to a limited subset of census districts can provide insight into the dynamics of migration in China, and has the additional advantage of opening up potential avenues of research into the outcome of migrating relative to that of not migrating (Liang and Chen 2004).

China's NBS is aware of the challenges involved in enumerating migrants at their destination only, and that awareness is what led to the survey that entailed interviewing migrants at both their origin and their destination. Our study has highlighted the need to interpret the reported migration figures with proper caution, especially for figures reported in the 2000 census, the 2005 population survey, and the 2010 census, all of which were implemented after a period of robust internal migration in China. For the 2010 census, NBS used both measures, and we anticipate that a release of the micro-data from that census will allow a more in-depth analysis of the issues presented here by enabling scholars to consider the migrant population using data collected at both the origin and the destination.

Effectively tracking large migrant populations remains a critical challenge in other less developed countries that are undergoing urbanization, and they too can learn from the experience in data collection of the Chinese NBS. The results imply that future population surveys should track migration using both origin and destination measures to enable more effective policy design. They also highlight the challenges associated with survey design among migrant populations in less developed countries.

Notes

1. Avraham Ebenstein is at the Department of Economics, Hebrew University of Jerusalem, Mount Scopus Campus, #4223, Jerusalem 91905, Israel. E-mail: ebenstein@mscc.huji.ac.il. Yaohui Zhao is at Peking University, Beijing.
2. The authors thank Revital Bar, Eyal Frank, Michael Freedman, Tian Meng, Susan Schwartz, and Chuanchuan Zhang for excellent research assistance.
3. Yaohui Zhao received financial support from the Natural Science Foundation of China and China's Ministry of Education.

References

Cai, Fang, Dewen Wang, and Yang Du. 2002. Regional disparity and economic growth in China: the impact of labor market distortions, *China Economic Review* 13 (2–3): 197–212.[doi:10.1016/S1043-951X(02)00072-X]

Chan, Kam Wing. 2001. Recent migration in China: patterns, trends, and policies, *Asian Perspective* 25(4): 127–155.

Chan, Kam Wing. 2003. Chinese census 2000: new opportunities and challenges, *The China Review* 3(2): 1–12.

Chan, Kam Wing. 2012. Migration and development in China: trends, geography and current issues, *Migration and Development* 1(2): 187–205.[doi:10.1080/21632324.2012.739316]

Chan, Kam Wing, Ta Liu, and Yunyan Yang. 1999. *Hukou* and non-*hukou* migrations in China: comparisons and contrasts, *International Journal of Population Geography* 5(6): 425–448.[doi:10.1002/(SICI)1099-1220(199911/12)5:6<425::AID-IJPG158>3.0.CO;2-8]

China National Bureau of Statistics. 2012. Available: http://www.stats.gov.cn/eNgliSH/newsandcomingevents/t20060322_402312182.htm (accessed 4 March 2012).

Fan, C. Cindy. 2005. Interprovincial migration, population redistribution, and regional development in China: 1990 and 2000 census comparisons, *The Professional Geographer* 57(2): 295–311.[doi:10.1111/j.0033-0124.2005.00479.x]

Lavely, William. 2001. First impressions from the 2000 census of China, *Population and Development Review* 27(4): 755–769.[doi:10.1111/j.1728-4457.2001.00755.x]

Liang, Zai and Yiu Por Chen. 2004. Mender and migration in China: an origin-destination linked approach, *Economic Development and Cultural Change* 52 (2):423–443.[doi:10.1086/380594]

Liang, Zai and Miao David Chunyu. 2013. Migration within China and from China to the USA: the effects of migration networks, selectivity, and the rural political economy in Fujian Province, *Population Studies* 67(2): 209–223.[doi:10.1080/00324728.2012.756116]

Liang, Zai and Michael J. White. 1996. Internal migration in China, 1950–1958, *Demography* 33(3): 375–384.[doi:10.2307/2061768]

Shen, J. 2013. Increasing internal migration in China from 1985 to 2005: institutional versus economic drivers, *Habitat International* 39(1):1–7.[doi:10.1016/j.habitatint.2012.10.004]

Solinger, Dorothy. 1997. *Contesting Citizenship in Urban China: Peasant Migrants, the State and the Logic of the Market*. Oakland: University of California Press.

Zhang, Weimin, Xiru Li, and Hongyan Cui. 2005. *China's Inter-census Survey in 2005*. National Bureau of Statistics of China. Available: http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN021378.pdf. (accessed 6 August 2015).

Zhao, Yaohui. 2000. Rural-to-urban labor migration in China: the past and the present, in Loraine A. West and Yaohui Zhao (eds), *Rural Labor Flows in China*. Berkeley: Institute of East Asian Studies, University of California, pp. 15–33.

Technical appendix

Origin vs. destination-based surveys and integrated approaches

Destination-based measures may be superior to origin-based methods in instances where an entire family moves, and nobody remains at the *hukou* to complete the census roster or, as in the case of recent college graduates, the *hukou* is placed under the care of the university or one of the Labour Market Service Agencies. The origin-based approach also has limitations in that it places a heavy burden on family members still at the *hukou* to report accurately the location and other demographic characteristics of the migrant, such as their occupation, industry, and income. Furthermore, the limitations in counting migrants at the origin owing to full-family moves suggest that the origin-based total cannot be relied on completely either. Statistics that rely on the collection of information from family members at the origin may depend primarily on responses from the elderly. In this case, memory lapse and mental errors may result in migrants being improperly characterized by destination or length of stay.

A strategy that may prove effective in capturing rural migrant families in future census surveys may be to complement origin-based accounting with questionnaires given to village leaders. Rural migrant families often have to maintain contact with village leaders for necessities of life related to *hukou* registration. For example, newborn children must be registered at the *hukou* location of one of the parents and social insurance programmes (such as the new rural medical insurance plan and old-age pension scheme) are provided at the *hukou* location as well. In addition, village leaders are in charge of distributing agricultural subsidies and maintaining land records. Although village leaders may not have access to detailed information on the current status of the migrant (e.g., employment status), they can provide information useful for producing an accurate migrant population count.

Comparing population surveys with other sources of migration statistics

Migration statistics in China can vary dramatically by their source. As discussed in depth by Chan (2012), government sources can differ significantly in the number of migrants reported. For example, China's NBS conducts annual Rural Migrant Labour (RML) surveys using the households which participate in the Rural Household Survey. This has the benefit of being an origin-based survey, similar to the source we relied on in our calculations. The RML estimate reported for 2005 is 125.8 million, which includes all migrants who had left their *hukou* place township for at least 6 months. Our estimate was 189 million, a difference of nearly 64 million. As noted by Chan (2012), however, it may be that the RML understates the number of migrants since it enumerates households that have someone available for interview and to keep diaries and probably omits households that have fully migrated, or households in which the remaining household members are illiterate and unable to keep expenditure diaries. Also, since migrants are in some cases seasonal workers, they are likely to move multiple times within the year and be harder to enumerate using standard population-counting procedures. Finally, the RML figures refer to the working population only, while the survey we rely on included both the working and non-working population, so that the two sources cannot be directly compared.

De jure *vs.* de facto *migration*

An important distinction exists between examining migrants based on their current location (*de facto*) vs. classifying them as migrants based on their lack of local *hukou* (*de jure*). As demonstrated in the paper, a significant number of the migrants in the legal sense of the term (*de jure*) have lived in their destination for a long period of time, and may even have children who were born at this destination. Individuals who transfer their *hukou* to their destination are not considered migrants by our definition, even if they only recently moved to this location. Since they will respond to the survey as 'living in their *hukou*', they will not be considered migrants, irrespective of their length of stay at this new residence. Our rationale for using this *de jure* '*hukou*-based' definition of migrant is driven by two considerations. First, the vast majority of migrants in China do not transfer their *hukou* to their new residence. The Chinese bureaucracy still treats local urban *hukou* as a privilege, and so it is extremely rare for even those who permanently reside in a new city to be able to access local urban *hukou*. Second, migrants living outside their *hukou* are generally not eligible for social protections or public benefits. Thus, defining migrants as those living outside their *hukou* enables us to focus on this particularly vulnerable population.