"The differences in population characteristics between an administrative system and census data, and their impact on population estimates: a comparison of methods of estimating population in Israel"

Ahmad S. Hleihel[1]

The Component Method (CM) of population estimation uses census counts plus the effect of vital statistics events, internal migration and international migration. In Israel all individuals' events are registered in the administrative system (AS). To apply the CM we translate individuals' records (from the census and the AS) to an aggregate level and calculate population estimates in aggregated components. However, this method disconnects individuals' characteristics from aggregated characteristics, and biased estimates result when individual census characteristics differ from the same individual's characteristics in the AS. The Israel Central Bureau of Statistics (ICBS) conducted a project to implement another method, that makes it possible to conduct follow-ups on the individual level, rather than on the aggregated level. In this paper I will describe the two methods; present the advantages and disadvantages of each method; and finally, analyze the different results of the two methods, focusing on their effect on internal migration

---

[1] *Central Bureau of Statistics, Israel*

Introduction

Population estimates for small areas are essential for planning and budgeting. Therefore, the estimates need to be as accurate as possible. The censuses of population are an exclusive source for estimating the population of small areas. In the years following the census (and until the subsequent census is conducted) the Israel Central Bureau of Statistics prepares population estimates for small areas, which are based on the results of the most recent census conducted and adjusted each year to reflect the changes in the population registry. Although registration of demographic changes in Israel is thorough, there are delays and inaccuracies in registration. Because of these delays, there is sometimes inconsistency between the estimates obtained in the census and the updated statistics following the census. Therefore, the longer the period that elapses after the census, the less accurate the population estimates for small areas will be. In other parts of the world, and especially in European countries, where administrative data from population censuses are used extensively to prepare estimates of populations in small areas, accuracy and timeliness of the data are the main problems encountered (see: Laihonen et. al. 1998, Scheuren 1999).

In this paper, I will examine the level of accuracy of the population registry in Israel by comparing two methods of preparing population estimates for small areas. The two methods can be distinguished in the way they apply changes in demographic data to update the estimates. Even though the paper does not aim to determine which of the two systems is preferable, the question inevitably arises from the analysis and discussion of the data. Ostensibly, updating of population estimates at the level of individuals would appear to be preferable to aggregated updating. However, the results of our investigation show that three conditions are necessary for individual updating to be more effective than aggregated updating: (a) maintaining a registry of residents of institutions; (b) maintaining a registry of emigrants from Israel; and (c) ensuring that future population censuses allow for construction of a base population that is suitable for application of that method.

The Administrative System in Israel and Population Estimates – Background

The population registry of Israel, which identifies people according to their identity card numbers, was established several months after the declaration of Israel's independence in 1948. Since then, residents are required by law to register any demographic changes at the various authorities, which in turn are required to report and update the population registry about those changes. The Ministry of Health is responsible for registering births and deaths, the Ministry of the Interior is responsible for registering changes of address in the country, and the Border Police is responsible for registering arrivals and departures from Israel. Files that contain a list of the people who registered changes are sent to the Central Bureau of Statistics, which calculates the population estimates.

There are two major problems with the population registry and the administrative system in Israel that affect the population estimates. One problem is the lack of an emigrants register, and other is the lack of correspondence between the addresses listed in the population registry and actual addresses of residence (see: Kamen, 2004). The Central Bureau of Statistics has had difficulty compiling an emigrants register for two main reasons (see Hleihel and Ben-Moshe 2002): (a) inaccurate registration of departures and arrivals of Israelis up to the 1980s; and (b) many Israelis have dual citizenship, so that their arrivals and departures can be registered under different passports and it is difficult to follow how long Israelis stay abroad. In the 1995 census, 25% of the enumerated residents lived at an address that was different than the address listed in the population registry. Although there is no documentation to confirm the reasons for these discrepancies, my experience working with the files and with the census of population and housing in 1995 suggests that there are three main types of people whose addresses in the registry are different from the actual addresses of residence:

1. People who just moved to a new address and did not have a chance to update their file. In these cases, it can be assumed that the population estimates in areas that were established in recent years tend to be larger than actual number of people listed as residents of those in areas in the population registry.

2. People who reside at a temporary address ("floating people"). This group includes young singles, university students, and people living in long-term residential institutions. In these cases, it can be assumed that in areas with long-term residential institutions or in areas that attract young singles and university students, population estimates tend to be larger than the actual number of people listed as residents of those areas.

3. People who list an address at the registry that is different from their actual address of residence in order to receive discounts and services from certain authorities. These people usually seek to benefit from services such as parking, school registration, tax discounts, etc. In those areas, it can be assumed that population estimates tend to be smaller than the actual number of people listed as residents of those localities.

Therefore, population estimates in Israel are based on the results of the most recent census and updated according to the latest population movements. To date, Israel has been using the aggregated method (AGRM) to construct population estimates. According to this method, the estimates of the last census and the administrative records are converted from the level of individuals to the level of cells (a cell is equivalent to the total population in groups with a particular combination of demographic and geographic characteristics). Even though both of the sources for constructing population estimates derive from individual records, the method

does not allow the identification of specific individuals and does not allow for examination of the discrepancies between the old and new characteristics for each individual. This generates excess or insufficient updates of people, which leads to biased estimates of populations in small areas, and particularly in areas with a large concentration of biased registration as described above.

In light of that problem, the Israel Central Bureau of Statistics conducted an experiment in which population estimates were constructed based on the population census files and the population registry on the level of individuals – the Individual method (INDM). This system enables us, for example, to examine whether address changes reported to the administrative system are new or late. To conduct individual population estimates by linking records through the PIN, all of the records in the population base that was constructed from the 1995 census must be defined in terms of the PIN. Because this was not possible, the base population had to be reconstructed in the way described below.

Methodology

Description of Methods

Both of the methods rely on the same sources. The Census of Population and Housing used to construct the base population, and administrative files are used to update the population. To obtain an initial description, both methods can be applied to calculate population estimates in area "i" during the year "t", using the classic component method (Formula 1), which takes into account all of the components of population size changes. Because all of the change components are obtained from the administrative system, both methods use the same components and the same records (total births - $N_{B,t}$, total deaths – $N_{D,t}$, internal migration – $N_{E,t}$, total immigrants to Israel – $N_{II,t}$ and total emigrants from Israel – $N_{IO,t}$).

$$P_i^t = P_i^{t-1} + \Delta_i^t \tag{1}$$

Where

$$\Delta_i^t = \sum_{j=1}^{N_{B,t}} (B_{ij}) - \sum_{j=1}^{N_{D,t}} (D_{ij}) + \sum_{j=1}^{N_{E,t}} (EI_{ij}) - \sum_{j=1}^{N_{E,t}} (EO_{ij}) + \sum_{j=1}^{N_{II,t}} (II_{ij}) - \sum_{j=1}^{N_{IO,t}} (IO_{ij}),$$

and where the record of an update in each of the components is given the values "0" or "1". The value "0" is given when it is decided not to update the population in area "i" of that person, and the value "1" is given when it is decided to update the population in area "i" of that person.

Even though we use the same sources, three main differences between the two systems can be identified.

Differences in the base population

Population estimates at the level of small areas are based on the results of the last census, which was conducted in 1995. The census results were assessed in a "Post Enumeration Survey" (PES), according to existing estimates, which were based on the 1983 census, and according to demographic analyses that took into account vital statistics on births, deaths, and migration between 1983 and 1995. PES revealed that for the entire country there was an undercoverage of 0.7%. The other sources indicated the groups for which current population estimates had to be adjusted (State of Israel, 1998). After identifying the target groups for adjustment, it was necessary to impute records with characteristics similar to those of the target population group. One of the problems with this adjustment is allocating our imputation among small areas. In AGRM, other records were selected from the existing records in the census ("hot deck imputation method"), and random allocation was conducted among the various areas, where different weights were allocated according to the size of the area. Similar allocation cannot be conducted in INDM, because it requires real records where identification is carried out on the basis of identity card numbers. Because identification is necessary, the base population had to be reconstructed, with emphasis on two components: (1) we completed the identification of about 4% of the Israelis who were missing identification numbers. PIN was completed by replacing these records with records from the "registry excess"[2], having similar geographic and demographic characteristics. This procedure included imputation of demographic characteristics from the population registry, but left the address as it was obtained from the population census. (2) To adjust the shortage discovered in the census target groups, we looked for people whose characteristics fit those of the target population in the excess registry ("cold deck imputation method"). The total number of records found in the excess registry, for the various groups, was similar but not identical to the adjustment made using the aggregated method. The addresses of the people who were added to the population following the adjustment was defined as the address listed in the population registry, in contrast to the random dispersion which was conducted using the aggregated method. This led to slight differences between the two basic methods used on the national level, and larger differences between the estimates in small areas.

At the end of 1995 (two months after census day), a difference of 3,247[3] residents was found in the areas selected for analysis. Although this is a very small difference, the difference between the imputation methods causes differences in the size of populations in small areas.

---

[2] "registry excess" includes the list of people in the registry who are list as being alive but were not found in the census (by linking through PIN) and were not defined specifically as emigrants.
[3] At the national level, a discrepancy of 6,081 records was found between the two methods, which amounts to 0.1% of the aggregated estimate.

On the average, the difference between estimates of the base of the population was negative and approached zero, but the standard deviation was about 1.5 (see Figure 1a).

Population Living in Institutions

Because there is usually a difference between the addresses listed in the population registry for residents of institutions and the addresses found in the census, and because there is no source of data that allows for follow-up of those residents, each method deals with this problem differently. In AGRM, the population living in institutions is frozen as it was obtained in the census, on the assumption that the size, distribution, and demographic characteristics of the population will not change until the subsequent census. Even though this assumption is incorrect, it is essential in order not to reduce the population size in small areas with a high proportion of persons living in institutions. Because this solution cannot be applied with the INDM, it does not take residents of institutions into consideration. As a result, it is expected that over time areas with high concentrations of people in institutions will lose population. These differences may be expressed as follows:

In the AGRM, persons in regular households are separated from persons in institutions (for details on distinguishing between household populations and institutional populations, see for example Rees et al., 2004, pp. 11-12). Therefore, the population in region "i" at time "t" equals:

$$P(AGR)_i^t = P(AGR\text{-}IN)_i^{t-1} + \Delta(AGR)_i^t + IN_i(0) \qquad \textbf{(2)}$$

Where

$P(AGR\text{-}IN)_i^{t-1}$ is the population at time "t", excluding the population of institutions;

$IN_i(0)$ is the institutional population., which is constant and equal to the institutional population on the census date.

According to the INDM, the population in area "i" at time "t" equals:

$$P(IND)_i^t = P(IND)_i^{t-1} + \Delta(IND)_i^t \qquad \textbf{(3)}$$

Where $P(IND)_i^{t-1}$ is the total population, including the population of institutions.

In sum, the AGRM method maintains the size of the permanent population in the places where they were enumerated, while updating the demographic changes in the place where the residents are listed in the population registry. The INDM method, by contrast, relates to residents of institutions as a regular population, and changes in that population are updated at the actual place of residence. If that address is different from the one listed in the population registry, then it will be changed to the address of residence regardless of the address listed in the registry.

As a result of this difference, areas with a large concentration of institutions lose considerably more in INDM estimates than in AGRM estimates. Therefore, in these areas AGRM estimates tend to be much higher than the INDM.

Differences in Methods of Updating Population Estimates

The above-mentioned difference is the result of the discrepancy between the results of the population census and the figures listed in the population registry. The problem is when the observed value in the census is "x", whereas the value of the population registry is "y". If the same person changed his listing in the registry from "y" to "x" (or to any other value), and if these differences are not controlled as in the AGRM, then characteristic "y" will be excessively updated, and characteristic "x" will not be updated enough. This will cause bias in population estimates for value "y" as well as for value "x". In the INDM, the differences between the census and the administrative system can be controlled. Therefore we update the characteristic only if the new value obtained from the administrative system is different from the value observed in the census. Therefore, when we calculate the change in the population each year we will obtain the difference (Δ_i^t). This difference derives from the difference between the decision about the value of each record change, which is coded as "0" or "1". Table 1 presents a summary of the differences for each of the components, regarding the decision about when the components receive a value of "1".

Table 1: The Difference in Methods for Deciding on Value "1" Components

Components		Aggregated Method decision	Individual Method decision
B_{ij}	Births	The administrative address at date of birth is area "i"	The newborn is not in the population at t-1, and the administrative address at date of birth is area "i"
D_{ij}	Deaths	The administrative address at date of death is area "i"	The person is in the population of area "i" at "t-1"
EI_{ij}	Internal Migration (in area "i" flows)	The destination administrative address is area i	The person is out of area "i" at t-1, and the destination administrative address is area "i"
EO_{ij}	Internal Migration (out of area "i" flows)	The source administrative address is area "i"	The person is in area "i" at t-1, and the administrative target address is not area "i"
II_{ij}	International Migration (in area "i" flows)	The administrative address is in area "i"	The person is not in the population at "t-1", and the administrative address is area "i"
IO_{ij}	International Migration (out of area "i" flows)	The administrative address is in area "i"	The person is in the population of area "i" at t-1

As shown in the table, the differences in the decision about components are influenced by two main factors: the difference in address listings in the population registry (about one-fourth of the difference derives from the problem of institutional residents) and the listings in the

census, and obtaining demographic changes in the registry for people who do not belong to the population according to the INDM.

Measuring the Difference between Methods

We seek to calculate the difference between the two estimates and follow the development of this difference over time. Because the small areas are different sizes, we will calculate the percentage of the difference between the two estimates in relation to the estimate that was obtained by the AGRM.

$$\frac{S_i(t+n) + S_i(0)}{P_i^{t+n}(\text{AGRM})} \ X \ 100 \quad = \quad \frac{P_i^{t+n}(\text{AGRM}) - P_i^{t+n}(\text{INDM})}{P_i^{t+n}(\text{AGRM})} \ X \ 100 \qquad (4)$$

Where

"t-1" is the base time (census day).

$S_i(t+n)$ the (t, t+n) cumulative differences that are affected by the updating difference.

$$S_i(t+n) = \sum_{s=t}^{t+n} (\Delta(\text{AGR})_i^s - \Delta(\text{IND})_i^s) \qquad (5)$$

$S_i(0)$ the difference at t-1.

Formula 4, above, does not express the difference between the two systems in their treatment of the population of institutions. This is because residents of institutions do not cause the difference in the base population, and the difference derives from the listing in the population registry and the listing in the census among residents of institutions. Therefore, the difference in the population of institution residents is expressed mainly as $S_i(t+n)$, and its behavior over time.

The Convergence of $S_i(t+n)$

Two main differences impact on the cumulative difference $S_i(t+n)$ – differences in listings of the base address, and differences in the way each method treats updates of records that are outside of the population. If we assume that $S_i(t+n)$ is affected only by the difference in the listing of the base address, then we can conclude that any person whose address in the registry is different from the one enumerated in the census (enumerated in area "i" and found in area "j", where i≠j), only has an effect once, when one of two changes is first listed: (a) departure from the population (died or emigrated), which reduces $S_j(t+n)$ by ((-1)-(0)) 1, and

8

increases $S_i(t+n)$ by (0)-(-1) 1. (b) Listing of address changes from area "j". The effect of this change depends on the new target listing as summarized in Table 2.

Table 2: Impact on the Difference between the Two Methods, by Source and Target Addresses Obtained from the Administrative System

Internal migration registration from area j to:	The impact on the cumulative difference		
	$S_i(t+n)$	$S_j(t+n)$	$S_k(t+n)$
Area j → Area i	(+1) – (0) = +1	(-1) – (0) = -1	(0) – (0)=0
Area j → Area k ($\neq$i)	(0) – (+1) = -1	(+1) – (0) = +1	(+1) – (+1) = 0

In the initial years after the census, we can expect numerous address changes that will impact on the cumulative difference. After a certain period of time, the marginal addition to the cumulative difference diminishes.

If M equals the number of cases in which the base address was different from the administrative address

M^t equals the number of cases in which the base address was different from the administrative address, and in which the address was changed for the first time or in which the person left the population during year "t"

Then as of year "m" we will obtain

$$\sum_{s=t}^{t+m} (M^s) \text{-------> } M \qquad \Longleftrightarrow \qquad S_i(t+m) \approx S_i(t+m+1)$$

"m" is dependent on the cycle of people's address changes and the reasons that people were listed in one place and enumerated in another. For example, if the discrepancy is because the person moved to a new address and didn't have time to register it, then the change can be expected to occur quickly. If the discrepancy is because a person sought to receive services that are provided in the area where they are registered, then the change may take more time to be listed and depends on the person's need for those services.

In reality, "m" is not dependent only on the differences in registration of the address. It also depends on changes registered in the population registry for people who "left" the list. Even though it can be expected that there won't be many such cases, this can cause the convergence to be incomplete. Moreover, it can cause a permanent discrepancy between the cumulative difference that results from registration of movements of emigrants.

Data and General Findings

For the purpose of this paper, I select 1,468 small areas numbering at least 500 inhabitants, in which the majority of the population is Jewish[4]. Localities that have less than 10,000 inhabitants are considered to be one area, and localities that have 10,000 or more inhabitants are divided into statistical areas (each statistical area includes 4-6 voting precincts or 2000-5000 inhabitants). Each statistical area is considered one area.

From 1995 to 2003, the "components of addition" and the "components of reduction" were calculated each year for every area according to the rules in Table 1. Following those calculations, the population estimates for the end of December each year were calculated according to both methods. The difference between the two estimates in the areas selected for this paper was 3,247 at the end of 1995 (4,324,848 and 4,346,095 according to the AGRM and INDM, respectively). This difference can be explained by two factors – the difference in construction of the population base on census day (4 November 1995), and the difference in updating the population every year until the end of the year. The difference between the two estimates at the end of 2003 (eight years after the census) was 42,629 (4,859,119 and 4,901,748 according to the AGRM and INDM, respectively). The change in the difference between 1995 and 2003 can be attributed mainly to the difference in updating the "components of addition" (births and immigration to Israel), and the "components of reduction" (deaths and emigration from Israel) for the total population. As can be seen in Figure 2, the difference in the "components of addition" is smaller than the difference in the "components of reduction". The impact of the births component was negligible, and was found only in 1995. By contrast, the impact of the deaths component was greater - beginning with a difference of about 9% in 1995 and stabilizing afterwards at around 5% (with minor changes). The percentage of change in the component of immigration to Israel[5] was around 5% throughout the entire period. At the end of the period, the percentage of change resulting from immigration to Israel was very high (about 60% in 1995 and 30% in 1996). Afterwards, it declined to about 10% Because the impact of the "reduction components" was greater than that of the "addition components", the estimate that was obtained through INDM eventually became larger than that obtained through the AGRM. Therefore the difference was greater. Consequently, the findings indicate that the impact of the differences in the variables was unidirectional, where the INDM estimate became greater than the AGRM estimate. This unidirectional is not maintained when we consider the differences between estimates at the

---

[4] The Non-Jews areas (Arabs areas) are excluded because there are no addresses in these areas. Lack of address disables us to estimate population size in small areas.

[5] The component of immigration to Israel consists of two main groups: immigrants who did not previously have the status of residents of Israel, and immigrants who had lived in Israel in the past and returned to live in the country. The difference in updating of the first group is very similar to that which was found for the component of births. Therefore, the main difference derives from the Israelis who returned to live in the country. As of 2000, about 50% of that group was not updated by the INDM.

level of small areas, because the internal migration factor enters, and its effect on one area is the opposite of its effect on another area.

Results of Differences in Small Areas

The development of differences at the level of small areas is influenced by two main factors: the first is the difference in updating of "addition and reduction components". Unlike the national level, these components can have an influence in two directions, depending on the characteristics of the population in those areas. Thus, in an area with a difference in "addition components", there is not necessarily a difference in "reduction components", i.e., the ratio of reduction differences to addition differences in one area is not equivalent to "addition/reduction" ratio in all of the areas. The second factor that has a major influence is the difference in updating internal migration. Even though the percentage of difference in updating of internal migration is not much greater than the other components, its impact is considerable because of the volume of address changes that occur each year. (For example, in 2003 there were about 420 thousand changes of address, about 38 thousand deaths, about 145 thousand births, about 29 thousand new immigrants to Israel, about 9.5 thousand returning residents, and about 28 thousand emigrants from Israel).

Comparison of the distribution of the percentage of difference between the two estimates during a given year with the percentage from the preceding year (Figure 1) reveals several results:

- Over the years, there has been a distribution similar to the normal distribution of parameters that change with time.
- On the average, the INDM estimate is larger than the AGRM estimate. Over time, the difference between the estimates has increased. In 1995, the average percentage of the difference was –0.023, compared with -1.316 in 2003.
- There is a process of convergence in the distribution of the percentages of differences. At the beginning of the survey period, the distribution changed significantly from one year to the next. In recent years, the yearly changes in the distribution have not been drastic. This finding is illustrated in the diagram, as well as in the changes in distribution parameters.

To enhance the validity of these conclusions, a distinction should be made between different types of small areas, based on the factors that cause the record in the population registry to be different from the population census. In our description of the administrative system of Israel, we mentioned three main factors. However, we do not have all of the data on the characteristics of areas in the three groups. Therefore, we cannot conduct any analyses by different types of areas. The only variable that can be examined is the number of residents of institutions who were enumerated in a given area. Figure 3 portrays two distributions for two

11

types of areas. The first category includes areas in which at least 5% of the population resided in institutions in the 1995 census (N=185). The second includes the rest of the areas. Analysis of the differences between the distributions of the difference between the two groups of areas leads to the following conclusions:

- Even though the average is positive in areas with institutions and negative in other areas, it can be concluded that in the base year (1995) there were no significant discrepancies in the distribution of the differences between the two types of areas.

- The differences between the distributions in the two types of areas increase rapidly, as shown. The percentage of the difference between the two methods in areas with institutions increases in a positive direction, and the percentage of difference in other areas increases in a negative direction. There are two explanations for this difference. First, the population of institutions has usually been enumerated outside of the location in which it was registered in the population registry. The second cause is the special treatment of the population of institutions in the AGRM. Notably, the aggregated method keeps the framework of institution residents frozen for a length of time, whereas the individual method subtracts that population from the locality in which was enumerated without replacing residents of institutions. Therefore, in contrast to the AGRM, the INDM eventually empties out the areas with residents of institutions, which causes the average difference between the estimates to continue increasing in a positive direction.

- The change in the average percentage of difference and in the standard deviation is more rapid in areas with institutions than in other areas.

- Here, too, we see a process of convergence in the distributions, from two perspectives: (a) the change in the difference between the distribution of the percentage of the difference by type of area diminishes over time. This is graphically displayed in Figure 3, and is also shown in Table 3, which presents the relative change in the difference in the parameters of the distribution. (b) The convergence of the entire distribution of the percentage of the difference is presented separately for each type of distribution. This is graphically displayed in Figure 4, and also shown in Table 3, which presents the relative change in the difference in each of the parameters of the distribution separately.

- Even though convergence is evident in both types of areas, the process is slower in areas with institutions.

- Even if we ignore the problem of institutions, which is one of the main differences between the two methods, we will still find that the distribution of the

percentage of the difference changes over time. The trend of this change is similar to the process of convergence in all areas with the difference in the parameters of the distribution. This similarity can be attributed to two main causes. First, residents of institutions also influence the difference in the areas where they are registered, which are not necessarily classified as areas with institutions. Moreover, the other areas evidently include other types of populations, which lead to differences that were not necessarily revealed by our data.

Table 3: The Annual mean and standard deviation of the difference between the AGRM and INDM distributions, by type of area 1995-2003

.

	Total areas		Including areas with institutions (at least 5% of the people residing in institutions in 1995)		Excluding areas with institutions (less than 5% of the people residing in institutions in 1995)	
	N=1496		N=185		N=1283	
	Mean	SD	Mean	SD	Mean	SD
1995	-0.023	1.466	0.078	1.525	-0.038	1.457
1996	-0.235	3.049	1.051	3.352	-0.421	2.958
1997	-0.434	3.935	1.811	4.458	-0.757	3.747
1998	-0.634	4.727	2.535	5.355	-1.091	4.449
1999	-0.859	5.344	3.057	6.136	-1.424	4.974
2000	-1.035	5.762	3.508	6.707	-1.687	5.306
2001	-1.171	6.155	3.824	7.320	-1.892	5.616
2002	-1.316	6.468	4.073	7.824	-2.093	5.856
2003	-1.298	6.745	4.386	8.213	-2.119	6.085

Discussion

The above analysis of the development of differences between the two methods (INDM and AGRM) aimed to examine the accuracy of the Israeli population registry and the extent of its impact on population estimates in small areas. Both methods reveal inaccuracies

in the population registry. In particular, inaccuracies have accumulated following the 1995 population census which cannot be observed without further investigation in the field (i.e., in the forthcoming population census). The discrepancies between records in the population registry and the results of the 1995 census can be partially solved by using the INDM. This solution improves population estimates in small areas by reducing "over-updating" due to discrepancies between the data in the census and the data in the population registry on census day. However, the INDM does not solve the problem of residents of institutions, where the population changes constantly over time. The AGRM solution for institutions is preferable, even though we do not manage the population of institutions. This solution assumes that the size and characteristics of the population of institutions remain constant over time. Even though this assumption is not accurate, it allows us to maintain the size and characteristics of the population in areas with institutions. Because the INDM does not include residents of institutions, small areas with institutions are "emptied" of the population in those institutions because the "reduction components" for residents of institutions are conducted from the place where they are enumerated and not from the place in which they are registered – in contrast to the AGRM which "reduces" that population in the place where they are registered.

The ideal solution for updating the population of institution residents based on the INDM would be to conduct a census of residents of institutions[6], which would be updated once a year. The establishment of a census like this makes it possible for the individual method to include residents of institutions, in order to improve the quality of INDM estimates compared with AGRM estimates.

A solution to the problem of institutions alone is not enough to prefer the INDM over the AGRM. Another important condition for preferring the INDM is to conduct population censuses in the future that will enable us to use this system. For the first time in the history of population censuses in Israel, we plan to move from a traditional census to an integrated census in 2008, which will combine a sample of 20% in the field with the listing in the population registry (for more details of the new census approach see: Blum 2003, Kamen 2004, Nirel et. al. 2003). Although the final census product includes records of 100% of the population census, it has two main limitations that may limit the use of the INDM. The first problem is that the final product includes records that do not belong to the population (emigrants). The second problem is that we obtain population registry records in which each record is given a weight. The total weight of all records in every area yields the estimate of the total population in that area. Therefore, construction of the base according to both

---

[6] This solution would require a yearly administrative census of residents of institutions. A list of residents can be obtained by the administration of the institution and/or from the administration of the umbrella organization that finances those institutions. In so doing, we obtain almost the same result as we would by examining only the changes and not the entire list.

methods is possible. The problem is, what about the updates of the population in the years that follow the census? This problem is easily solved by the AGRM, because the component is converted into an aggregate, which makes it possible to manage the population estimates. Regarding the INDM, it is still not clear whether the new census method will allow for this. If it is possible to manage the population estimates with the INDM, then in order to use the method there must at least be a clear distinction between those who belong to the population versus the emigrants. This distinction will be possible only if a stock of emigrants from Israel is conducted.

Besides the two main problems mentioned above, the integrated census attempts to improve data on addresses by using other administrative sources such as the electric company and municipal tax (*Arnona*) files. If this investigation reveals that the method is successful, and if the administrative files are used to adjust address listings regularly, then the estimates obtained on the basis of the INDM will be improved.

One of the important findings of this study relates to the convergence of the differences between the two estimates. Convergence does not mean that we have reached a point where all of the records in which there was a difference between the area of residence enumerated on census day and the area of residence listed in the population registry (about 20% of the records) have led to a change that will generate differences between the two methods (change of address, deaths, and emigration). Between 1995 and 2003, only 56.1% of the records led to at least one of the changes that affect the difference (Figure 5). Most of the changes focus on the component of address changes. There are no data that can be used to examine why the remaining 44% still did not change their address in the population registry. This category can be divided into three main groups: (1) people who were enumerated at a temporary address and who returned to their permanent address after the census, so that they did not have to change their address in the registry; (2) people who intend to change their address but take more time to do so; (3) people who have no intention to change their address. We will only to able to distinguish between those groups (and mainly between groups 1 and 3) in the 2008 census. Groups 1 and 3 are not the source of the difference between the two methods, but they influence the difference between each of the methods and the actual status of the population. Group 3 does not influence the difference in the results obtained on the basis of the two systems and the actual status of the population, but will continue to influence the difference between the listings in the population registry and the actual status of the population. Group 1 biases the estimates, because there is a population that changed addresses without registering the move in the population registry, and this affects the differences between the results of the estimates and the actual status of the population. However, it is important to understand that Group 1 is correctly registered in the population registry.

If we assume that most of the records are in Group 1, then we can conclude that after a period of several years, the main difference between the file obtained through the INDM and the population registry file will focus on population category (a distinction between emigrants and residents of Israel), rather than on addresses of residence. This finding makes a very important contribution to understanding the quality of the estimates that are obtained from the administrative records. If the conditions mentioned above are met, and the INDM is preferable to the AGRM (i.e., a registry of emigrants, a census of institutions, and use of additional administrative files), then we will be able to obtain better population estimates in small areas, which will maintain their value for a long period after the population census. If some of those conditions do not exist, then after a certain period the records we obtain from the population registry will not be as good. In sum, the INDM will be preferable only on the condition that the frequency of censuses in Israel is increased to a point where a census is conducted every five years (instead of the current system of conducting a census every 11-13 years). If the censuses are conducted more frequently, then it will be possible to produce good population estimates even if all of the conditions specified above are not fulfilled. This discussion is theoretical of course, because reality is more complex and in order to fulfill the conditions above and/or conduct more frequent censuses, there is a need for increased budgets, which are not always available.

Even though the main goal of these procedures is to improve the quality of population estimates in small areas, and even though it is less important whether the aggregated or individual methods are used to update population statistics, it is important to update and manage population statistics for other purposes as well. Individual management make it possible to distinguish between the people who belong to a given population and those who do not belong to it. In this way, it provides a solution to some of the problems involved in updating the population registry. This will be very helpful in constructing samples on an individual basis in surveys conducted at the Central Bureau of Statistics. For example, it will enable us to avoid sending enumerators to people who do not belong to the population or whose real address is different from the one listed in the population registry.

Conclusion and Next Steps

In this paper, we aimed to present the problems involved in calculation population estimates in small areas in Jewish localities in Israel. In particularly, we aimed to present problems deriving from the delay in updating the population registry. For that purpose, we compared the results of estimates obtained by two methods. One method (the AGRM) does not take into account updates in the population registry by actual results obtained in the 1995 census. The other method, INDM, does take into account the discrepancies between the two sources. As

shown in this paper, we cannot determine which of the two methods is more accurate or timely, because each method has advantages and disadvantages.

The next step, therefore, is to continue calculating population estimates by both methods until the next population census in 2008. After that census, the results of the census can be compared with the results obtained under each of the two systems, using a method similar to that of Simpson et al. (1996) in England and Wales. This kind of comparison would enable us to compare the quality of the estimates obtained under each of the systems with the actual estimates. Then it will be possible to define the characteristics of the areas in 2008, by comparison with the characteristics revealed in the 1995 census. This kind of comparison would answer the question about the method that should be used to prepare population estimates in small areas in the 2008 census.

References

Blum Olivia, 2003 "Wholistic Approach to Multifaceted Integrated Census". Administrative Records I, FCSM 2003 Research Conference, November 17-19. (http://www.fcsm.gov/03papers/Blum.pdf).

Hleihel Ahmad and Ben-Moshe Eliahu, 2002 "Measuring Emigration from Israel". Presented on CESD-Rome MEDSTAT Programme, MED-Migr project on Migration Statistics
Workshop on International Migration Statistics, Rome 30-September- 3-October 2002

Kamen Charles S. 2004 "The 2008 Israel Integrated Census of Population and Housing". United Nation Expert Group Meeting to Review Critical Issues Relevant to the Planning of 2010 Round of Population and Housing Censuses. 15-17 September 2004, New York (http://66.102.9.104/search?q=cache:bILMbABFISIJ:unstats.un.org/unsd/demographic/meetings/egm/CensusEGM04/docs/AC98_9.pdf+Israel+Integrated+Census+&hl=iw)

Laihonen Aarno ,Ib Thomsen, Elisabetta Vassenden and Britt Laberg. 1998. *Final report from development project in the EEA: Reducing Costs of Censuses through use of Administrative Registers*. Statistics Norway, Department of Social Statistics, January 1998.

Nirel Ronit, Hagit Glickman and Dan Ben-Hur, 2003 "A Stratigy For A System Of Coverage Samples For An Integrated Census". Proceeding of Statistics Canada Symposium 2003, Challenges in Survey Taking for Next Decade.
(http://www.statcan.ca:8096/bsolc/english/bsolc%3Fcatno%3D11-522-X20030017729&e=9838).

Redfern Philip, 2003, "Estimating Census Undercount by Demographic Analysis: New Approaches to the Emigrant Component" Journal of Official Statistics, Vol. 19, No. 4, pp. 421-448.

Rees Philip, Paul Norman and Dominic Brown, 2004. "A framework for progressively improving small area population estimates", *Journal of the Royal Statistics Society*, Series A (Statistics in Society), Volume 167, Part 1, PP. 5-36.

Scheuren Fritz, 1999. "Administrative Records and Census Taking" *survey Methodology*, Vol. 25, No. 2, PP. 151-160. Statistics Canada
Simpson Stephen, Ian Diamond, Pete Tonkin and Rachel Tye, 1996. "Updating Small Area Population Estimates in England and Wales" *Journal of the Royal Statistics Society*, Series A (Statistical in Society), Vol. 159, No. 2. pp. 235-247.

STATE OF ISRAEL – CENTRAL BUREAU OF STATISTICS 1998, DEMOGRAPHIC CHARACTERISTICS OF THE POPULATION IN LOCALITIES AND STATISTICAL AREAS: Population by Sex, Age, and Marital Status. 1995 Census of Population and Housing Publication. Publication No. 7 Vol. B.

Figure 1: Annual change in the distribution (in Percents) of the difference percent between AGRM and INDM on small areas: 1995-2003 (N=1496).
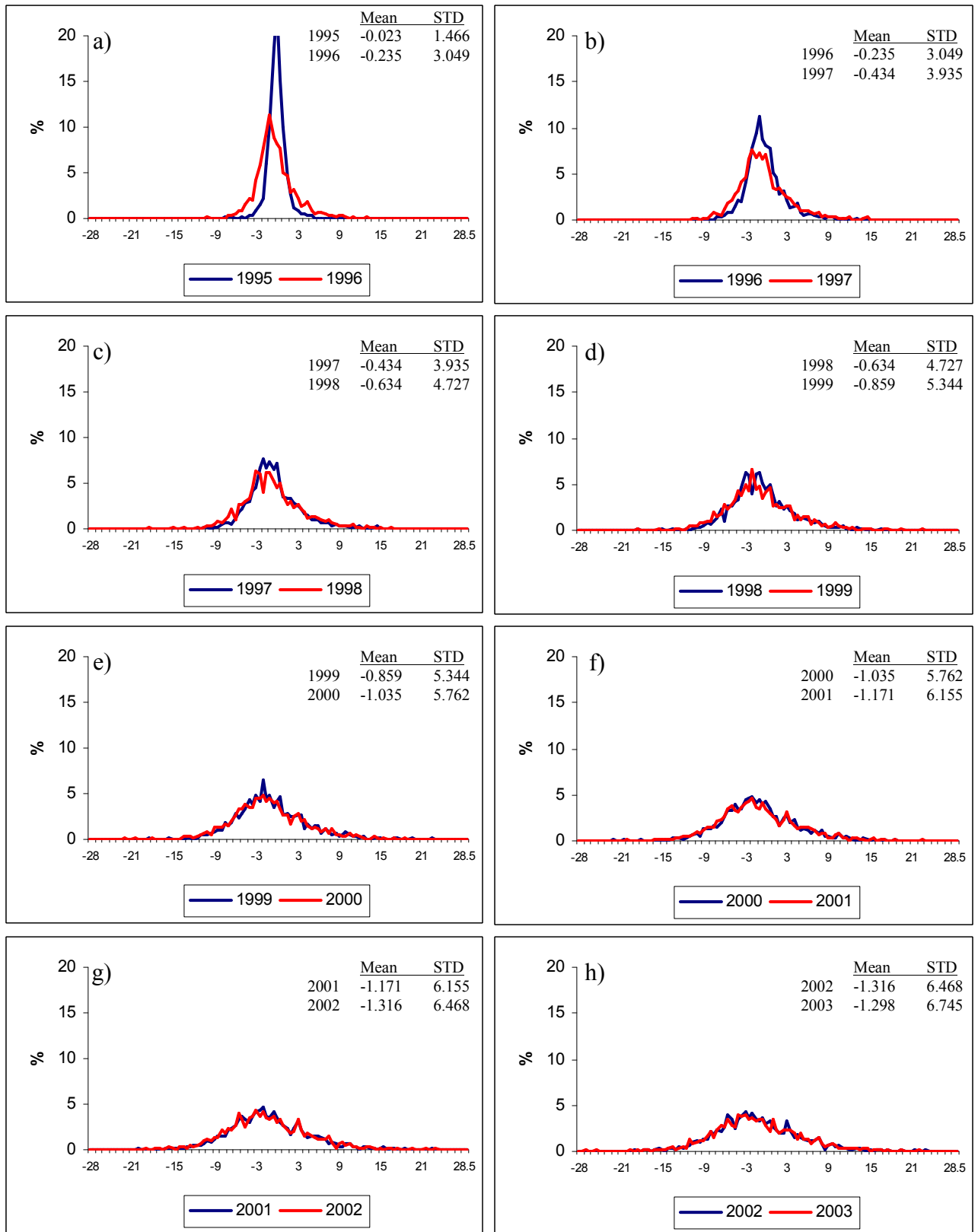
Figure 2: Annual percentage of records not updated on the INDM (relative to AGRM) by components (vital statistics, internal migration, and international migration) 1995-2003.
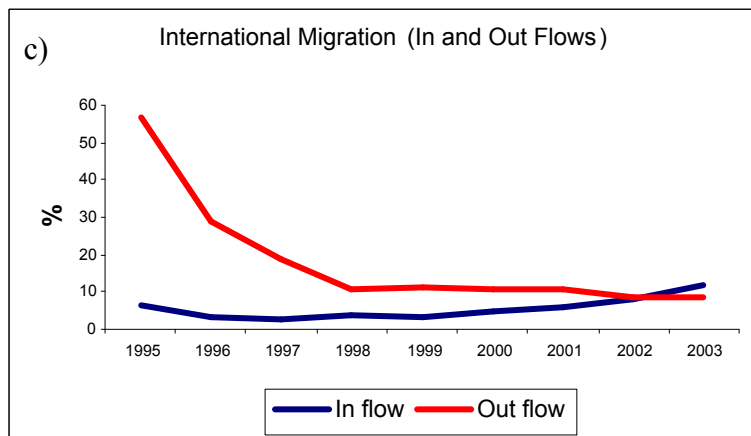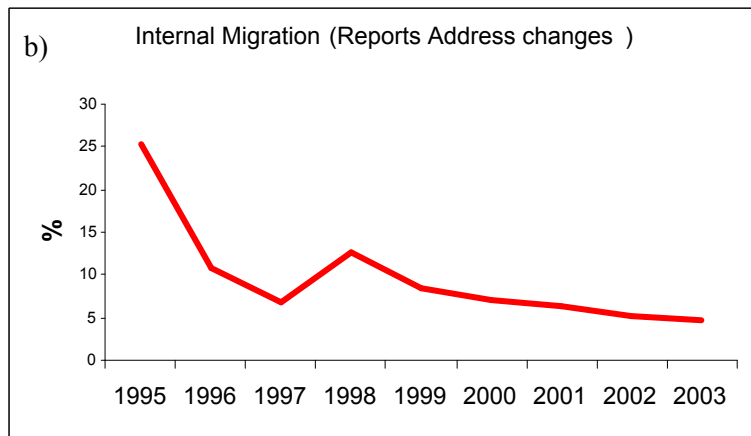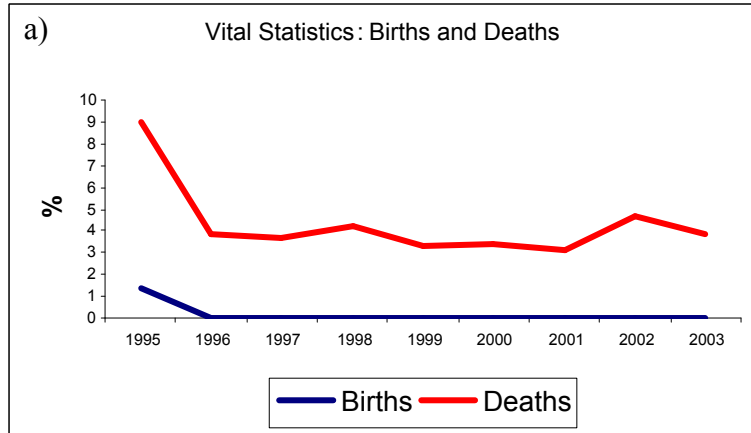
Figure 3: Annual change in the difference between the distribution (in Percents) of the difference percent between AGRM and INDM on areas include Institute population (N=185) with other areas (N=1283): 1995-2003
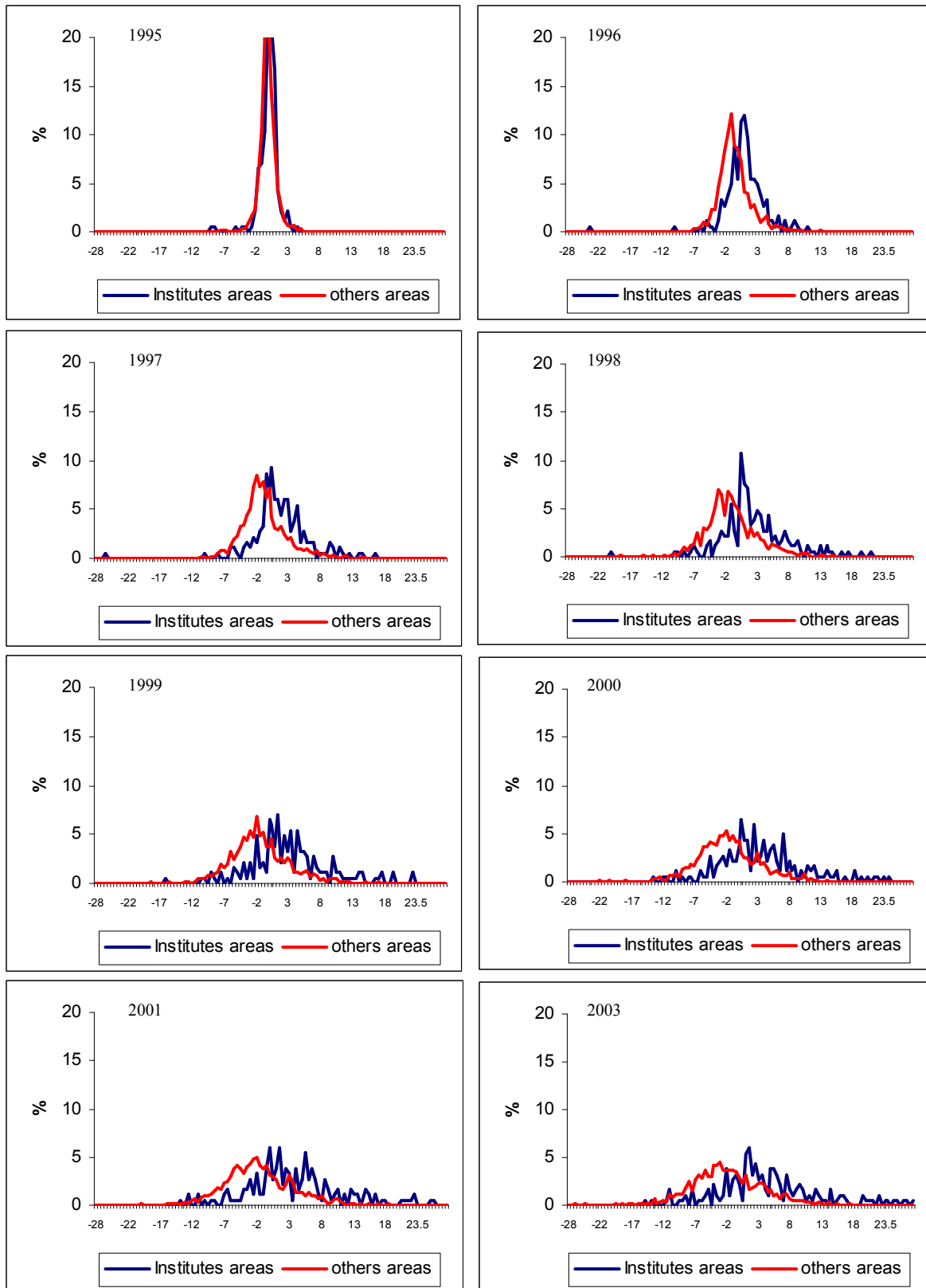
Figure 4: Annual change in the distribution (in Percents) of the difference percent between AGRM and INDM on small areas – excluding areas include at least 5% institute population: 1995-2003 (N=1283).
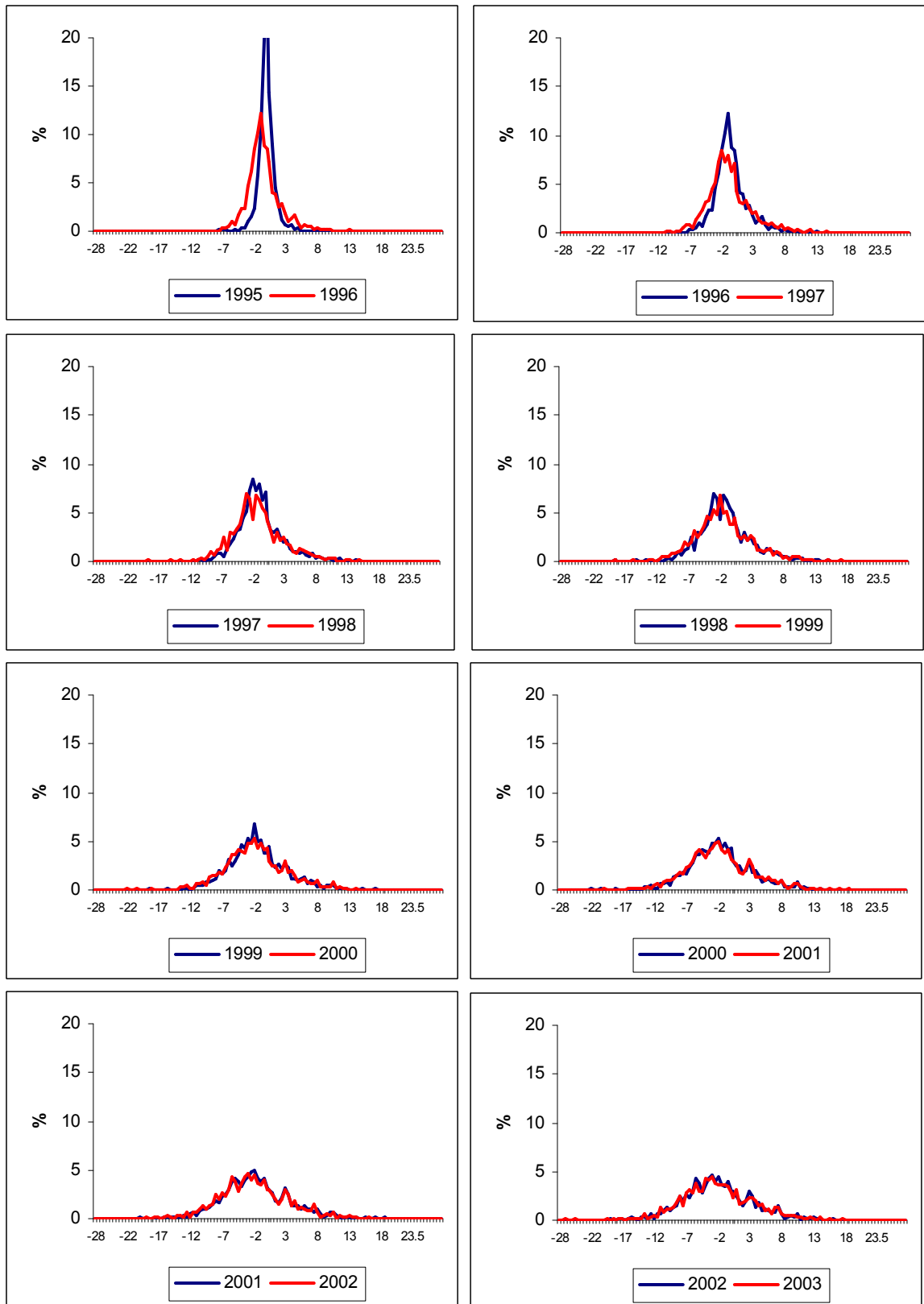
Figure 5: The cumulative percent of first change (include only changes on address, deaths and emigration) among people that their area at the census day was deferent from the area at population register.