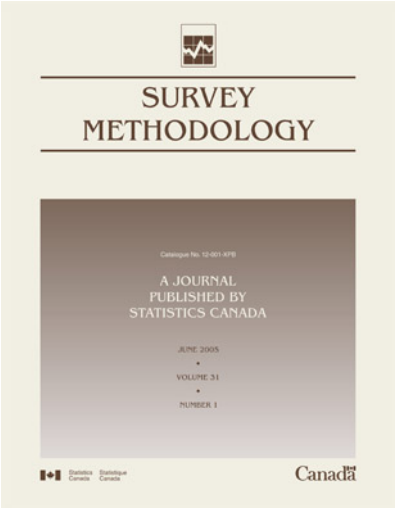
Catalogue no. 12-001-XIE

Survey Methodology

December 2003



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.

Statistics Canada

Business Survey Methods Division

Survey Methodology

December 2003

Published by authority of the Minister responsible for Statistics Canada

February 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Vol. 29, N° 2, pp. 155-165
Statistics Canada, Catalogue No 12-001-XPB

Coverage Error in Population Censuses: The Case of Turkey

H. Öztaş Ayhan and Sühendan Ekni [1]

Abstract

Coverage errors and other coverage issues related to the population censuses are examined in the light of the recent literature. Especially, when the actual population census count of persons are matched with their corresponding post enumeration survey counts, the aggregated results in a dual record system setting can provide some coverage error statistics.

In this paper, the coverage error issues are evaluated and alternative solutions are discussed in the light of the results from the latest Population Census of Turkey. By using the Census and post enumeration survey data, regional comparison of census coverage was also made and has shown greater variability among regions. Some methodological remarks are also made on the possible improvements on the current enumeration procedures.

Keywords : Census coverage error; Coverage error measures; Coverage error estimation; Dual record system estimate; Population census; Post enumeration survey.

1. Introduction

Coverage has been an important issue in censuses as well as in sample surveys. The difference between the census count and the target population count is the coverage error. When the census count is less than the target population count, it creates an undercount as is common in many countries.

Several techniques are available to understand the problem of coverage errors in censuses. Dual record system (DRS) estimator (Chandra Sekar and Deming 1949) was also extended by many researchers (Ayhan 2000; Casady, Nathan and Sirken 1985; Hogan 1990, 1993a and 1993b; Isaki 1992; Marks, Seltzer and Krotki 1974).

Dual record system estimates based on the census enumeration and a post enumeration survey (PES) are used by the U.S. Census Bureau to measure census coverage error (Hogan 1993a and 1993b; Mulry and Spencer 1988, 1990 and 1993). Post enumeration surveys can be used to improve the population estimate (Ayhan and Ekni 1991; Diffendal 1988; Hogan 1990; Hogan and Wolter 1988).

For the United States, the 1980 Census Post Enumeration Program attempted to measure census coverage through direct measurement using sample survey models (Fay, Passel, Robinson and Cowan 1988). Several methods are also proposed for the adjustment of census count for under enumeration (Choi, Steel and Skinner 1988; Cressie 1988 and 1990).

Recently, models for population coverage error have been studied extensively (Isaki 1992; Wolter 1986). A method of overlapping data systems or multiple frame methodology was used to improve the population estimates (Goodman 1949; Hartley 1962 and 1974; Bankier 1986).

This study highlights the methodological problems related to the population census coverage and proposes remedies to some of the issues covered. In addition, it proposes and discusses alternative estimates for the population census coverage errors. To achieve the above goals, coverage evaluation issues are included in the design of the PES.

In this paper, section 2 discusses methods of census enumeration and section 3 covers post enumeration survey procedures. Methods of coverage error estimation is presented in section 4 of the paper. Estimators of the population total is given in section 5 and comparison of the coverage error statistics are presented in section 6. Important findings are summarized in the conclusion.

2. Methods of Census Enumeration

Population censuses have many common features in most countries. The method of enumeration can either be based on *de facto* or *de jure* system. In *de jure* system people are enumerated at their normal residence, while *de facto* system enumerates people actually there. *De facto* system is widely used in developing countries, and the *de jure* system is generally used in developed countries. Countries that are using *de facto* system of enumeration seem to have more problems related to coverage, than the countries which are using *de jure* system of enumeration. These problems stem mainly from their existing imperfect frames for their target population.

De facto based population censuses are generally conducted on a single day, as a complete count, to determine the total population within the country on the day of enumeration. The citizens of the country who are living outside the country were excluded from the census, whereas alien population who are present within the country were included in the census.

1. H. Öztaş Ayhan, Department of Statistics, Middle East Technical University, 06531 Ankara, Turkey; Sühendan Ekni, State Institute of Statistics, 06100 Ankara, Turkey.

Design of the Census Operations

The Population Census of Turkey was conducted, on the basis of *de facto* system of enumeration, by the State Institute of Statistic (SIS) to determine the quantitative, social and economic characteristics of the population.

For the Census enumeration, the *list of buildings* are created by the local authorities and send to the local Census Committees. Based on the list of buildings (Forms 1 or 2), *enumeration districts* (EDs) *list of buildings* (Form C) are formed (see Appendix 1 for details). Due to the lack of timely availability of the complete list of buildings in the SIS Central Office, the number of EDs are estimated by projection techniques for advance fieldwork planning of the Census, as well as PES. EDs are obtained by assigning 100 persons per enumerator in province and district centers, and 200 persons per enumerator in sub-districts and villages, based on average daily workloads. They are then numbered sequentially. In the Census, the listed addresses were taken as the base for identifying the "dwelling units (DUs)", while the "individual persons" within the household(s) of the dwelling unit is considered as the unit of enumeration.

The workload of each enumerator is taken as an ED, which contains a list of addresses to be covered within a specified close interval. Instructions are given to the enumerator to treat this interval as a compact segment. If an enumerator encountered an address not on the list, it is included in the enumeration by work definition. For vacant and nonexistent units the related information is also recorded. There was no special procedure for dealing with reluctant respondents or in general any non-interviewed units, due to the compulsory nature of response by the related Act. The enumerator's workload is set in such a way that, they will complete all the interviews in a given day. For very special cases, the instruction is given to complete the enumeration of the segment during the extended hours in the same census day.

Additional enumerators were assigned to enumerate the "*special enumeration district*" such as the places of the mobile populations (travelers, persons on duty, nomadic tribes, *etc.*) and institutional populations (hospitals, prisons, factories, military establishments, *etc*).

Institutional population are covered by additional enumerators, who are assigned for these special EDs. The mobile populations travelling by vehicles are stopped and were enumerated as a group when they first appeared within the borders of the provinces. The passengers continues their journey after enumeration and duplicate enumerations are avoided by placing an "*Enumerated*" sign on the vehicle after the census operation, and later their individual identification is also checked by manual and computerized algorithms, against the other records of the relevant settlement.

The Census was conducted on a Sunday, and the enumeration was completed on the same day. On the Census day, a national curfew was declared. The enumerators visited each household (HH) within the dwelling units listed in their enumeration district building lists and completed the census questionnaire (see Appendix 2 for details). For the *Household Module*, the information is collected from an adult household member for the general household characteristics, while for the *Individual Person's Module* the information is obtained from self respondents on their personal characteristics.

The following type of errors occurred during the different stages of the census operations;

(1) Omission errors and erroneous inclusions has occurred during the construction of the List of Buildings. However, due to the use of compact segment approach in the enumeration process of the census operation, these errors are mostly eliminated.

(2) Response errors based on memory recall error, cheating, and inadequate answer for coding has occurred during the census enumeration. These are measured as the response inconsistency in the *Response Reliability Study* (Ayhan and Ekni 1991; SIS 1994) of the Population Census, which was based on the PES.

(3) Some enumerator errors (failure to probe, inadequate perception of response, and recording errors) are also observed during the census operation. These are also covered by the *Response Reliability Study*.

(4) Processing errors such as, coder and verifier errors also occur during the data processing and these are eliminated later during the data handling in the office.

3. Post Enumeration Survey Procedures

The objectives of the PES are to determine coverage error in the population census as well as obtain measures of response reliability of the questions in the census. In this paper, the first objective is discussed for the Population Census of Turkey, and the preliminary findings for both objectives are summarized by Ayhan and Ekni (1991).

3.1 Sample Selection Procedures

The sample design for the PES is initiated 3 months before the Census operation. At this stage, creation of the Population Census EDs was not complete yet.

Stratification and estimation of population EDs. The previous Population Census enumeration district lists of the State Institute of Statistics is used as the base for sampling frame for PES operation. The population of people is first stratified into 5 *geographical-socio economical regions* of Turkey. A second explicit stratification variable is also used, which is based on the 8 *size groups for the place of the settlements*, in a nested structure within regions. Here,

urban-rural boundary corresponds to a population size of 10,000. The number of census enumeration districts were estimated for 40 *design strata* for the Census day by using forward population projection method, which were based on the person counts of the two previous population censuses. For the census enumeration, EDs were created by using Form C within the Central Office. A total of 479 251 EDs are established for the Census. Sampling frame information is given in Table 1.

Table 1
Estimated Number of Population and Sample EDs by Regions and Urban-Rural Strata

Region	Urban Popu. ED	Urban Samp. ED	Rural Popu. ED	Rural Samp. ED	Total Popu. ED	Total Samp. ED
h	$M_h^{(U)}$	$m_h^{(U)}$	$M_h^{(R)}$	$m_h^{(R)}$	M_h	$m_h^{(1)}$
1	125,726	125	40,333	32	166,059	157
2	42,442	42	24,992	20	67,434	62
3	65,466	76	45,925	36	111,391	112
4	15,790	16	30,459	22	46,249	38
5	39,358	40	48,760	34	88,118	74
Total	288,782	299	190,469	144	479,251	443

The expansion factors: $F_h^{(1)} = N_n^{(1)} / n_h^{(1)} \neq M_h / m_h^{(1)} = F_h^{(2)}$.

The coverage of the number of dwelling units in the Census and PES were achieved by the following procedures. The number of population EDs for each province was determined and numbered sequentially. Then, the number of population EDs in each population strata was estimated by, dividing the projected strata population ($N_{h.}$) to the fixed daily workload of enumerators ($B_{h.}$). Population EDs were estimated for urban areas as $M_{hi} = N_{hi} / B_{hi}$ and for rural areas as $M_{hj} = N_{hj} / B_{hj}$, where the ED sizes are taken as fixed daily workload, $B_{hi} = 100$ persons in urban strata and $B_{hj} = 200$ persons in rural strata. The results of the population projections for each strata by urban-rural aggregation are also obtained. The estimated number of population EDs and expansion factors for regions and urban-rural strata are also computed.

Selection of sample EDs. A stratified multistage sample of localities and blocks are selected systematically for PES from the available *master sampling frame* of the State Institute of Statistics at the Central Office. The blocks of the master sampling frame is periodically updated for the multi purpose selection of other samples on routine basis. The interviewers of the PES is recruited and trained in the Central Office, and then interviewers are send as a team to the local sample settlements for the independent enumeration of the selected PES sample. For the identification purpose, the selected sample blocks are linked to their corresponding Population Census EDs of the settlement in the field by previously given instructions to the PES interviewers.

For the use of Dual Record System estimation, the sample enumeration districts for the PES should be determined independently from the census frame. This is an absolutely crucial assumption of the DRS model, which was emphasised by many researchers during the past 50 years (Ayhan 2000; Chandra *et al.* 1949).

Due to the use of unwanted old ED lists in some areas, the range of the planned workload per ED per enumerator may have changed and consequently the selected sizes of the EDs may be different from the actually enumerated sizes. This will effect the achieved sampling fractions, which will naturally be different from the selected.

A total of $m = 443$ sample enumeration districts are selected in 16 province centers, 23 districts, 16 sub-districts and 43 villages within the 40 strata. For the PES, a sample of 443 EDs are selected from the created ED list by systematic sampling.

The sampling fractions and sample allocation was achieved in the following way. Equal probability selection method was used to select the sample enumeration districts in all strata. Sampling fractions were planned to be $f_h \cong 0.001$ for all strata. However, the sampling fractions are also varied among strata. Technical details of the sampling fractions and the sample allocation are given below. The sampling fractions [$f_h^{(.)}$] and sample allocation [$n_h^{(.)}$] can be achieved as,

$$f_h^{(1)} = n_h^{(1)} / N_h^{(1)} = 1/F_h^{(1)} \text{ and } f_h^{(2)} = m_h^{(1)} / M_h = 1/F_h^{(2)} . \quad (1)$$

The total population sizes of urban (U) and rural (R) EDs are,

$$N_n^{(\mathrm{U})} = M_h^{(\mathrm{U})} B_{hi} = \left[\sum_i^I M_{hi} \right] B_{hi} \ \forall\, h \,\&\, i \text{ and} \quad (2)$$

$$N_h^{(\mathrm{R})} = M_h^{(\mathrm{R})} B_{hj} = \left[\sum_j^J M_{hj} \right] B_{hj} \ \forall\, h \,\&\, j \quad (3)$$

where the components are defined earlier.

Then the population size of each stratum was determined as

$$N_h^{(1)} = \left[N_h^{(\mathrm{U})} + N_h^{(\mathrm{R})} \right]. \quad (4)$$

Similarly, the corresponding sample sizes of each stratum are

$$n_h^{(1)} = \left[n_h^{(\mathrm{U})} + n_h^{(\mathrm{R})} \right] \quad (5)$$

where

$$n_h^{(\mathrm{U})} = m_h^{(\mathrm{U})} B_{hi} \text{ and } n_h^{(\mathrm{R})} = m_h^{(\mathrm{R})} B_{hj}. \quad (6)$$

3.2 Design of the PES Operations

The fieldwork operation for PES was identical to the census, where the details are given in section 2.2. For operational purposes, each ED was defined as a close

interval of dwelling unit numbers within the streets. In terms of special enumeration districts (*i.e.*, institution) the total number of enumeration districts are checked with prior information which was obtained at province level.

Due to previously given instructions to the enumerators, PES starts in the sample enumeration districts an hour after the starting time of the census enumeration on the same day. PES enumerators visit the same households in the same (ascending) order as the census enumerators, so that PES enumerators did not visit the same household before the census enumerators. Results of the PES was used as a basis for evaluation, after matching the individual cases with the census records for the corresponding EDs.

4. Methods of Coverage Error Estimation

This section addressed coverage error estimation, by stating data matching procedures and dual system estimation procedures and related findings. The evaluation and estimation of population coverage is obtained using the list of EDs from two independent sources. In this section, the data matching procedures, dual record system estimators, alternative population total estimators are proposed and the estimates are evaluated. A comparison of the computed coverage error statistics are also presented here.

4.1 Data Matching Procedures

Several models (Deming and Glasser 1959; Nathan 1967 and Tepping 1968) have been proposed for determining the optimum matching procedures. These are based on establishing procedures that minimise the '*estimated net matching error*' subject to given costs and other constraints (Marks *et al.* 1974). These models provided valuable concepts to the theory and practice of matching, but none of are completely satisfactory for all purposes.

The work of Tepping (1968), extended by Srinivasan and Muthiah (1968), required a minimum set of characteristics for the '*exact agreement*' in matching. Also, Ayhan and Ekni (1991) and SIS (1994) have used similar methods based on the following specifications;

(1) *Matching of the population of the EDs*. The total population of the ED was taken as the sum of the household population within the total DUs of the ED.

(2) *Matching of the households within the EDs*. Several sets of information (address of the dwelling unit, names of household head and number of persons in the household) was considered for matching of households.

(3) *Matching of the individual persons within the matched households*. A total of 4 Census/PES variables (names, age, sex and education level) were all used for exact agreement in matching of individuals.

(4) *Matching of non-matched individuals of the households*. This was achieved by matching with the other individuals in the neighboring households (from the other data source) by searching. The same criteria was used for exact agreement in matching of individuals.

The preliminary work of matching operation is done clerically, while matched households and persons are evaluated by automation. For the matching procedure of persons the frequencies $n(r, c)$ are shown in Table 2.

Table 2
The Layout of the Matching Procedure

Matching Procedure		Data Source 2: (PES) In	Not in	Total
Data	In	$n(1, 1)$	$n(1, 2)$	$n(1, *)$
Source 1:				
(Census)	Not in	$n(2, 1)$	$\hat{n}(2, 2)$	$n(2, *)$
Total		$n(*, 1)$	$n(*, 2)$	$\hat{n}$

On the basis of the above specifications, the households are matched at the first stage, and within the matched households the persons are matched at the second stage. The results are presented in the following tables by regions. Enumeration districts are located in sample settlements within 19 provinces in 5 regions of the sample design. Out of 443 selected EDs, 437 were matched with their corresponding population census counterparts and other 6 EDs could not be matched due to differences in independently given instructions for their creation by the local offices. The information on the regional breakdown of the 6 non-matched EDs are provided in Table 3, while the information on the urban-rural breakdown was not obtained.

The matching procedure of households can be illustrated by $k(r, c)$ in the same way as presented for persons in Table 3. In the stratified case, the number of households in each strata can also be denoted by $k_h(r, c)$. The total number of households in the Census which are not matched with PES households can be estimated for each strata as

$$k_h(1, 2) = [k_h(1, *) - k_h 1, 1)] \qquad (7)$$

and the total number of households in the PES which are not matched the Census households can also be estimated for each strata as

$$k_h(2, 1) = [k_h(*, 1) - k_h(1, 1)]. \qquad (8)$$

Information on matched and non-matched households are given in Table 3.

Table 3
Matched and Non-matched Households in the Post Enumeration Survey and Census Enumeration Districts by Regions

Regions	Selected no.of EDs	Enumerated no. of EDs	Matched HHs	Non-matched households	
				Census	*PES*
h	$m_h^{(1)}$	$m_h^{(2)}$	$k_h(1, 1)$	$k_h(1, 2)$	$k_h(2, 1)$
1	157	154	3,320	168	144
2	62	62	1,262	27	30
3	112	112	2,636	262	259
4	38	38	645	204	80
5	74	71	995	170	175
Total	443	437	8,858	831	688

In these 437 EDs, a total of 8,858 households were matched. In the Census 831 (9.38%) and in PES 688 (7.77%) households could not be matched. The Census based household match rate was 90.62%, while PES based match rate was 92.23%, which is presented in Table 3.

Coverage rates for the Census and PES households are given by regions in Table 4. Comparison of Census and post enumeration results for the coverage rate of households (C_h) were higher for the Census in most regions (except Regions 2 and 5) and the total. Here all the coverage rates were greater than it was expected. In terms of persons within the covered households, the coverage rates (C_h^*) were higher for the Census for all regions and for the total. Total of matched persons were $n(1, 1) = 41{,}020$ in the Census and PES.

Differences in the coverage of EDs in the Turkish Census and PES comes due to the following reasons;

(1) Additional Forms of C and D are established by the Census Committee of the provinces through list of buildings (Forms 1 and 2). List of buildings are created by the local authorities and they are not reliable enough for some settlements.

(2) Numbering of EDs are also done at the local level, they are also effected by the insufficient numbering operation.

(3) Forms C and D may or may not contain 100 persons in urban and 200 persons in rural areas due to outdated listings.

(4) Due to different starting points by the Census and PES enumerators, the number of dwelling units covered were different.

(5) Application of the PES questionnaire was started at least 2 hours after the actual Census operation within the selected EDs. Coverage differences may be due to the mobility of the members of census completed households within the same ED.

(6) During the one day enumeration period, some of the planned Census and PES questionnaires could not be completed, resulting inconsistency during matching. Of course, this is a source of under-count, which happened rarely during the field enumeration.

(7) Because of the *de facto* enumeration base, the local visitors (from other dwellings of the apartment) for either data source were subject to change.

(8) Again, due to *de facto* enumeration, there will be counting errors for the mobile population for the Census. The PES only planned to cover the household population.

(9) The PES was not planned to cover the special EDs and mobile populations (*i.e.*, travelers, persons on duty, *etc*). By definition, international and domestic travelers are permitted to continue their travel after being counted, if their journey had started before the official census starting time. During this research, the mobile population was excluded from the analysis.

(10) Nomadic tribes (Special enumeration techniques are required for the census of nomadic tribes. *De jure* rather than *de facto* enumeration base, as well as mobile interviewers may be recommended for the enumeration of nomadic tribes in place of interviewers who are stationary.) will not be covered in the PES due to non-listings.

Table 4
Number of Households and Persons in the Census and Post Enumeration Survey by Regions

Regions	Number of Household			Number of Persons in Households			
	Census	PES	Coverage	Census	PES	Matched	Coverage
h	$k_h(1, *)$	$k_h(*, 1)$	C_h	$n_h(1, *)$	$n_h(*, 1)$	$n_h(1, 1)$	C_h^*
1	3,488	3,464	1.0069	14,035	13,926	13,393	1.0078
2	1,289	1,292	0.9977	6,587	6,582	6,400	1.0008
3	2,898	2,895	1.0010	13,058	12,984	11,644	1.0057
4	849	725	1.1710	4,233	3,580	3,134	1.1824
5	1,165	1,170	0.9957	7,898	7,888	6,449	1.0013
Total	9,689	9,546	1.0150	45,811	44,960	41,020	1.0189

Coverage rates: $C_h = k_h(1, *)/k_h(*, 1)$ and $C_h^* = n(1, *)/n(*, 1)$.

(11) Both Census and PES EDs are enumerated with the same instruction for the previously defined close interval. However, due to the use of different quality of the frames (updated 1990 or outdated 1985 or even outdated 1980) the amount of workload of each interviewer was varying. Consequently, the amount of coverage in each ED may be different from both sources.

4.2 Dual Record System Estimation

Dual record system is used as a method for determining the estimated number of households and persons through a matching procedure. The results are used to estimate the total number of persons in each region and the total population. The model assumes independence of data collection from two sources, where the Census and the PES are used. In theory, all cells [n (r, c)] are observable except for n (2, 2) and any of the totals that include n(2, 2). Chandra *et al.* (1949) assumes that, there is no correlation bias with the estimate for cell n(2, 2). For practical purposes, this paper also considers this assumption as valid. On the other hand, further discussion on the validity of such assumption is recently reported by Ayhan (2000).

The methodology and the estimation procedures are presented below. Estimation of the number of persons not in the Census or in PES

$$n(2,2) = [n(1,2)\, n(2,1)] / n(1,1). \tag{9}$$

Total number of persons is estimated as

$$n = n(1,1) + n(1,2) + n(2,1) + n(2,2) \tag{10}$$

or alternatively,

$$n = [n(*,1)\, n(1,*)] / n(1,1). \tag{11}$$

Table 2 earlier illustrated the matching procedure used for the dual record system method. The computational procedure presented here was repeated for each region separately. The estimates are given in Table 5. For each strata, n_h is computed as n previously.

Table 5
Matched and Non-matched Number of Persons in the Census and Post Enumeration Survey by Regions

Regions	Matched	Census non-match	PES non-match	Estimated omissions in both sources	Dual record system estimate
h	$n_h(1,1)$	$n_h(1,2)$	$n_h(2,1)$	$n_h(2,2)$	$n_h^{(D)}$
1	13,393	642	533	26	14,594
2	6,400	187	182	5	6,774
3	11,644	1,414	1,340	163	14,561
4	3,134	1,099	446	156	4,835
5	6,449	1,449	1,439	323	9,660
Total	41,020	4,791	3,940	673	50,424

4.3 Total Population versus Household Population

The total population was considered as the target population for the population projections, which was used to estimate the total number of EDs in the population. On the other hand, PES sample design only considered the household population as the target population. Because the PES design was based on the selected sample dwelling units only, which excluded the special enumeration districts (the institutional population).

As stated earlier, the PES sample design was taken as the base for the comparison of two different enumeration systems during the matching procedures. This naturally led us to consider the household population as the target population for the appropriate estimation of the population total by the proposed estimators. In order to achieve this, the institutional population was computed later, from the 1990 Census data, for regions and population size groups. The institutional population of regions are presented (by aggregating over the size groups) in Table 6.

Table 6
Determination of Household Population and Sample Sizes by Regions

Regions	Projected population size	Institutional population estimate	Household population size	Household survey sample size	Expansion factors	
h	$N_h^{(1)}$	$N_h^{(2)}$	$N_h^{(3)}$	$n_h^{(1)}$	$F_h^{(1)}$	$F_h^{(3)}$
1	20,639,200	367,184	20,272,016	18,900	1,092.02	1,072.59
2	9,242,600	89,934	9,152,666	8,200	1,127.15	1,116.18
3	15,731,600	176,031	15,555,569	14,800	1,062.95	1,051.05
4	7,670,800	55,104	7,615,696	6,000	1,278.47	1,269.28
5	13,687,800	249,309	13,438,491	10,800	1,267.39	1,244.30
Total	66,972,000	937,562	66,034,438	58,700	1,140.92	1,124.95

$N_h^{(3)} = N_h^{(1)} - N_h^{(2)}$ here $F_h^{(1)} = N_h^{(1)} / n_h^{(1)}$ and $F_h^{(3)} = N_h^{(3)} / n_h^{(1)}$.

For the further use of the information on the institutional population, it was also assumed that, there were no coverage errors associated in measuring the institutional population during the 1990 Census enumeration. The household population of each region, are then computed by subtraction.

There were several reasons for removing the institutional population from the total population;

(1) The PES sample design only reflected the household population.

(2) The correct selection probabilities for the ideal coverage (representation) of each sample strata can only be based on the household population, not on the total population.

(3) The proposed coverage error estimates should only be based on the household population.

(4) The proposed estimators for the population total should also be based on the household population, where the PES results are household based.

(5) It will be wrong and misleading to make comparison of coverage error statistics, when the base populations are different.

(6) The census undercount is artificially inflated if the wrong population (namely, the total population) is taken as the target population.

4.4 Coverage Error Measures

Many coverage error statistics are proposed in the literature. Some of these error statistics are based on simple ratios or proportions, and others are based on more complex adjustment procedures. To simplify the solution to this problem, the following coverage error measures are proposed for the regional and total population. These are census coverage rate, census discrepancy rate and the amount of census discrepancy. The following coverage error measures are proposed which are based on the *household population*.

Census Coverage Rate:

Regional estimators:

$$\lambda_h^{(s)} = N_h^* / \hat{\mathbf{N}}_h^{(s)} \quad \forall \text{h} \;\; h = 1, 2, \ldots, H \tag{12}$$

where $N_h^* =$ Census count of the household population $[N_h^* = N_h - N_h^{(2)}]$ and $\hat{\mathbf{N}}_h^{(s)} =$ Estimate from source (or method) *s*.

Standard error of regional estimators: Making the following scale transformation $\lambda_h^{(s)}(0.5) = \tilde{\lambda}_h^{(s)}$ which is taken as a proportion, realizing that within each strata $\tilde{\lambda}_h^{(s)} + (1 - \tilde{\lambda}_h^{(s)}) = 1$, the standard error estimators of the census coverage rates of each region is computed as

$$\text{se}\left[\tilde{\lambda}_h^{(s)}\right] = \left[\frac{\tilde{\lambda}_h^{(s)} - \left(1 - \tilde{\lambda}_h^{(s)}\right)}{n_h^{(D)} - 1}\right]^{1/2}. \tag{13}$$

Total population estimator: $\lambda = N^* / \hat{\mathbf{N}}^{(s)}$. (14)

Census discrepancy rate:

Regional estimators: $\varphi_h^{(s)} = 1 - \lambda_h^{(s)} = \left[\delta_h^{(s)} / \hat{\mathbf{N}}_h^{(s)}\right]$. (15)

Total population estimator: $\varphi = 1 - [N^* / \hat{\mathbf{N}}^{(s)}] = 1 - \lambda$. (16)

Census discrepancy:

Regional estimators: $\delta_h^{(s)} = \mathbf{N}_h^{(s)} - N_h^* \quad \forall h$ (17)

Due to the limitations of the one day enumeration by the *de facto* system, other additional local coverage measures could not be considered for this study. Such additional coverage measures for the local areas could provide useful additional information for more complex coverage error models in countries who are employing *de jure* system of enumeration in their census taking.

Even with the limitations of the *de facto* census, one could compute coverage estimates for large domains (such as provinces), where the population would not likely to shift very much between Census and PES interview. This was not possible, due to the limited sample size of PES which did not provide independent provincial estimates to be made within the regions. In addition, the sample sizes might not be large enough to give sufficient precision.

5. Estimators of Population Total

The *estimated population total* is taken as the weighted sum of the all regional estimates.

$$\hat{\mathbf{N}}^{(s)} = \sum_h^H \hat{\mathbf{N}}_h^{(s)}. \tag{18}$$

The *standard error estimators* for the total household population of each region is computed as

$$\text{se}\left[\hat{\mathbf{N}}_h^{(s)}\right] = \hat{\mathbf{N}}_h^{(s)} \left[\frac{p_h(1 - p_h)}{n_h^{(D)} - 1}\right]^{1/2} \tag{19}$$

while the proportion of each strata is computed as $p_h = n_h^{(D)} / \sum_h^H n_h^{(D)}$.

The determination of the coverage error of a given Census is not an easy task, especially when a perfect list of a target population is not available to compare the results. This is always the case in most countries of the world, except the ones with population registers.

Table 7
Estimates of the Regional and Total Household Population for 1990 by the Expanded Dual Record System Estimate and Their Standard Errors

h	$\hat{\mathbf{N}}_h^{(1)}$	se$[\hat{\mathbf{N}}_h^{(1)}]$	$\hat{\mathbf{N}}_h^{(2)}$	se$[\hat{\mathbf{N}}_h^{(2)}]$	$\hat{\mathbf{N}}_h^{(3)}$	se$[\hat{\mathbf{N}}_h^{(3)}]$
1	15,936,939	58,967*	15,436,073	57,113	15,653,378	57,917
2	7,635,314	31,305	7,367,741	30,208	7,561,003	31,000
3	15,573,280	57,621	14,571,298	53,914	15,398,933	56,976
4	6,181,402	38,943	5,884,582	37,073	6,136,969	38,663
5	12,242,987	48,972	11,502,934	46,012	12,019,938	48,080
Total	57,569,922	241,794	54,762,628	230,003	56,770,221	238,435

*: *Standard error estimates are rounded to the nearest integer.*

Comparison of the results of a population census with projection figures also creates some kind of comparison problems, due to the validity of the several assumptions relating to projection models. In order to avoid a single base of comparison, the following *expanded dual record system regional estimators* are proposed for the determination of the census coverage errors.

Estimator 1. $$\hat{\mathbf{N}}_h^{(1)} = F_h^{(1)} n_h^{(\mathrm{D})} \qquad (20)$$

where $F_h^{(1)} = N_h^{(1)} / n_h^{(1)}$ and $n_h^{(\mathrm{D})} = \sum_r \sum_c n_h(r, c)$.

Here $n_h^{(\mathrm{D})}$ refers to the unweighted DRS estimate and $n_h^{(1)}$ corresponds to the selected sample size.

Estimator 2. $\hat{\mathbf{N}}_h^{(2)} = F_h^{(2)} n_h^{(\mathrm{D})}$ where $F_h^{(2)} = M_h / m_h^{(1)}$. (21)

Estimator 3. $\hat{\mathbf{N}}_h^{(3)} = F_h^{(3)} n_h^{(\mathrm{D})}$ where $F_h^{(3)} = N_h^{(3)} / n_h^{(1)}$. (22)

The dual record system estimators are expected to yield higher estimated counts than a single round survey (*i.e.*, PES), by definition. Therefore, all the proposed estimators for the household population totals are DRS based. DRS estimates of the total household populations are given in Table 7.

Difference between the three proposed estimates, are only based on the type of expansion factors used. When we examine the expansion factors, $F_h^{(1)}$ is based on the projected population sizes over household survey sample sizes. On the other hand, $F_h^{(2)}$ is based on total population EDs over total sample EDs of the original PES design.

Finally, $F_h^{(3)}$ is based on the household population size over household survey sample size. The first two estimators include institutional population components $[N_h^{(2)}]$ in the numerator of their expansion factors $[N_h^{(1)}$ or $M_h]$, while only the third estimator uses household population information $[N_h^{(3)}]$ in its expansion factor. It is clear that, the expansion factor for the third estimator is derived from the ideal selection probabilities $[f_h^{(3)} = n_h^{(1)}/N_h^{(3)} = 1/F_h^{(3)}]$ for the PES sample, which is based on household information. Therefore, *Estimator 3* can be considered as more representative of the household population.

6. Comparison of Coverage Error Statistics

For the comparison of error statistics, the population counts should be of the same standard base. It will be recommended to use a household population count which matches the corresponding population estimate. The regional and total population counts are given in Table 8. As mentioned earlier, the institutional population counts are determined from the 1990 Census counts.

Table 8
Regional and Total Population Counts for Turkey, 1990

h	Census counts N_h	Institutional population counts $N_h^{(2)}$	Household population counts N_h^*
1	18,544,967	367,184	18,177,783
2	7,836,940	89,934	7,747,006
3	12,824,347	176,031	12,648,316
4	5,964,565	55,104	5,909,461
5	11,302,216	249,309	11,052,907
Total	56,473,035	937,562	55,535,473

where $N_h^* = N_h - N_h^{(2)}$

For the purpose of population coverage error evaluation, the *census coverage rate* and the amount of *census discrepancy* was used. The computed population coverage error rates are given by regions and the total in Table 9.

Table 9
Estimates of the Census Coverage Rates for Regional and Total Household Population in Turkey 1990 and Their Standard Errors

h	$\lambda_h^{(1)}$	se[$\tilde{\lambda}_h^{(1)}$]	$\lambda_h^{(2)}$	se[$\tilde{\lambda}_h^{(2)}$]	$\lambda_h^{(3)}$	se[$\tilde{\lambda}_h^{(3)}$]
1	1.14061	0.00410	1.17762	0.00407	1.16127	0.00408
2	1.01463	0.00607	1.05148	0.00607	1.02460	0.00607
3	0.81218	0.00407	0.86803	0.00411	0.82138	0.00408
4	0.95601	0.00718	1.00423	0.00719	0.96293	0.00719
5	0.90272	0.00506	0.96088	0.00508	0.91955	0.00507
Total	0.96466	0.00223	1.01411	0.00223	0.97825	0.00223

There is a clear pattern for certain regions, for all estimates. The census coverage rates can also be expressed as the amount of census discrepancy. A similar pattern is expected for the three estimators, since the estimators are highly correlated.

For the total population, estimates based on methods (1) and (3) has resulted in census undercount when compared with the corresponding actual population counts. Due to the computational procedures, *Estimate 3* can be recommended among others because *Estimate 3* is based on the projected household population, where the comparison base is the same as the selection.

There is also a pattern for regional estimates, regardless of the method of estimation. For regions 1 and 2, all estimates indicated census overcount, while census undercount was observed for all other regions by all estimates, except for *Estimate 2* in region 4.

7. Conclusions

The coverage error study of the population census had provided some useful information in evaluating the methodological issues which is summarised below.

Comparison of the three proposed population total estimates indicates that, the first estimate provided the highest value of the total count, while *Estimate 3* provided more representative result for the total household population.

The evaluation of the census coverage error rates and the amount of census discrepancy had shown that, for the total population, *Estimates 1 and 3* has resulted in census undercount. There is also a distinct pattern for regional estimates, regardless of the method of estimation. There seems to be a census overcount in the first two regions, while census undercount was observed for the other three regions by all estimates (except for *Estimate 2* in Region 4).

For the developing countries, the main problem of census taking is based on the undercount. In Turkey, the overcount issues in census taking only occur in very limited local areas and they are re-evaluated later and removed from the census data before release of the census results.

On the basis of these findings it is clear that, the comparison of several sample based estimates with the population census count indicated the existence of some methodological problems which are present in the enumeration procedures of the Turkish Population Census. The most important of these issues are the following;

(1) Improving and updating the list of possible EDs in rapidly growing peripheries of the large cities by the use of area methods.

(2) Obtaining a perfect list of all dwelling units within the EDs. This can be better achieved through a continuous screening operation by the local authorities, where they are responsible for this by law. Alternatively, a Census of Housing can be taken just before the population census by the SIS which will also provide a useful frame for the population census enumeration.

(3) There are many laws in the country which refers to the latest "population counts". This suggests that, major changes might be necessary on legal issues as well as in enumeration techniques.

(4) Enumeration of the mobile populations also requires special attention, methods and qualified personnel.

One would like to hope that, measuring the characteristics of the population through the Censuses may be considered important, by the responsible officials in time and the necessary developments will take place along these directions.

Acknowledgements

We would like to thank Professors Orhan Güvenen, Yalçin Tuncer, Vijay K. Verma, Moti Lal Tiku, M. Qamar Islam and Mr. Ömer Güceliogl for their valuable comments. The contributions of Ms. Hasibe Dedes and Ms. Canan Bakici are also gratefully acknowledged. Finally, the comments and suggestions of the Editor, Associate Editor and the referee's of the Journal is very much appreciated. The views expressed are attributed to the authors and do not necessarily reflect those of the State Institute of Statistics, Turkey.

Appendices: Tools of Enumeration

The following listing forms and questionnaires are used before and during the Census and PES operations.

Appendix 1. Listing Forms Used

Form 1: *List of Buildings* (*for localities with municipal organization*).

This list is created by the local municipality personnel and later produced in triplicate. Used for sequential numbering of DUs in urban areas.

Form 2: ***List of Buildings*** (*for localities without municipal organization*).

This list is created by the village head person and later produced in triplicate. Used for sequential numbering of dwelling units in rural areas.

Form C: ***Enumeration District List of Buildings***.

This list is based on Forms 1 or 2. The EDs are formed on the basis of this list in urban and rural areas, separately.

Form D: ***Census Control List***.

This is an update of Form C which was completed by the enumerator after the census field operation and returned to the Local Census Committees with the completed census questionnaires. This form and the completed census questionnaires are forwarded to the SIS after the census field operation.

Appendix 2. Questionnaires Used

Form A: ***Population Census Questionnaire***.

The population census questionnaire consisted of four main parts. The information is collected through a personal interview by a paper and pencil approach.

Part 1. *Address details*.

Part 2. *Type of place of the residence*.

Part 3. *Household module* [contains 7 precoded household questions].

Information is collected to identify the household head, presence of head, total number of persons in HH, number of guests, number of HH members absent, ownership of present DU, and ownership of any other DU.

Part 4. *Individual person's module* [contains 26 precoded individual questions].

For each person present, information is obtained on sex, age, relation to HH head, place of birth, citizenship, permanent residence, educational background, marital status, fertility information, employment status, and main occupation.

Form B: ***Post Enumeration Survey Questionnaire***.

PES questionnaires are generally based on a subset of questions of the main study. However, for this study it was decided by the Census Advisory Committee to use the complete census questionnaire for the PES. The questionnaire for PES is completed in the same way as the Census.

References

Ayhan, H.Ö., and Ekni, S. (1991). Coverage and response errors in 1990 Turkish Census of Population. *Bulletin of the International Statistical Institute*, 54, 45-46.

Ayhan, H.Ö. (2000). Estimators of vital events in dual-record systems. *Journal of Applied Statistics*, 27, 157-169.

Bankier, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. *Journal of the American Statistical Association*, 81, 1074-79.

Casady, R.J., Nathan, G. and Sirken, M.G. (1985). Alternative dual system network estimators. *International Statistical Review*, 53, 183-197.

Chandra Sekar, C., and Deming, W.E. (1949). On a method of estimating birth and death rates and the extend of registration. *Journal of the American Statistical Association*, 44, 101-115.

Choi, C.Y., Steel, D.G. and Skinner, T.J. (1988). Adjusting the 1986 Australian census count for under enumeration. *Survey Methodology*, 14, 173-189.

Cressie, N. (1988). When are census counts improved by adjustment? *Survey Methodology*,14, 191-208.

Cressie, N. (1990). Weighted smoothing of estimated undercount. U.S. Bureau of the Census, *1990 Annual Research Conference Proceedings*, 301-325.

Deming, W.E., and Glasser, G.J. (1959). On the problem of matching lists by samples. *Journal of the American Statistical Association*, 54, 403-415.

Diffendal, G. (1988). The 1986 test of adjustment related operations in Los Angeles County. *Survey Methodology*, 14, 71-86.

Fay, R.E., Passel, J.S., Robinson, J.G. and Cowan, C.D. (1988). The Coverage of Population in the 1980 Census. *Evaluation and Research Reports*, PHC 80-E4, U.S. Bureau of the Census. 123.

Goodman, L.A. (1949). On the estimation of the number of classes in a population. *Annals of Mathematical Statistics*, 20, 572-579.

Hartley, H.O. (1962). Multiple frame surveys. In *Proceedings of the Social Statistics Section*, American Statistical Association, 203-206

Hartley, H.O. (1974). Multiple frame methodology and selected applications. *Sankhyā*, Series C, 36, 99-118.

Hogan, H. (1990). The 1990 Post Enumeration Survey: An Overview. *U. S. Bureau of the Census Paper*, Washington DC. 6.

Hogan, H. (1993a). The 1990 post enumeration survey: Operations and results. *Journal of the American Statistical Association*, 88, 1047-1060.

Hogan, H. (1993b). Planning for census correction: the 1990 United States experience. Invited Paper, 49th Session of the International Statistical Institute, Florence, Italy. *International Association of Survey Statisticians Booklet*, 133-150.

Hogan, H., and Wolter, K. (1988). Measuring accuracy in a post enumeration survey. *Survey Methodology*, 14, 99-116.

Isaki, C.T. (1992). Model bias effects in small area coverage error estimation. *Communication in Statistics*, Serie A, 21, 1213-1231.

Marks, E.S., Seltzer, W. and Krotki, K.J. (1974). Population Growth Estimation: *A Handbook of Vital Statistics Measurement*. New York: The Population Council.

Mulry, M.H., and Spencer, B.D. (1988). Total error in the dual system estimator: the 1986 census of central Los Angeles county. *Survey Methodology*, 14, 241-263.

Mulry, M.H., and Spencer, B.D. (1990). Total error in post enumeration survey estimates of population: the dress rehearsal census of 1988. U.S. Bureau of the Census, *1990 Annual Research Conference Proceedings*, 326-361 .

Mulry, M.H., and Spencer, B.D. (1993). Accuracy of the 1990 census and undercount adjustments. *Journal of the American Statistical Association*, 88, 1080-1091.

Nathan, G. (1967). Outcome probabilities for a record matching process with complete invariant information. *Journal of the American Statistical Association*, 62, 454-469.

S.I.S. (1994). 1990 Census of Population Response Reliability Survey. *State Institute of Statistics Publication*, No. 1688, Ankara, 65.

Srinivasan, S.K., and Muthiah, S.A. (1968). Problems of matching births identified from two independent sources. *The Journal of Family Welfare*, 14, 13-22.

Tepping, B.J. (1968). A model for optimal linkage of records. *Journal of the American Statistical Association*, 63, 1321-1332.

Wolter, K.M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, 81, 338-346.