

Report of the International Commission on the 2012 Population and Housing Census of Chile

2 September – 22 November 2013

Roberto Bianchini

Griffith Feeney

Rajendra Singh

Summary of Contents

List of acronyms and abbreviations

1. Introduction
2. Methodology
3. Preparatory work
4. Field operations
5. Data processing
6. Census products and services
7. Post-enumeration survey
8. Demographic equation estimates
9. Sampling of non-responding housing units
10. Next population and housing census
11. Conclusions and recommendations

Annex 1. Persons met

Annex 2. List of Comunas for INE Regional Directors

Annex 3. International Commission request to regional offices

Annex 4. Summary of the optical character reading process

Annex 5. Imputation of occupied housing units with non-responding inhabitants

Annex 6. Counts of housing units and persons for Bio Bio field visit EAs

Annex 7. Type of enumeration in censuses of Latin America and the Caribbean

List of Acronyms and Abbreviations

CELADE Latin American & Caribbean Demographic Centre - ECLAC

EA Enumeration area (census sector)

EC External Commission

IC International Commission

INE National Institute of Statistics of Chile

UN United Nations

Contents

Executive summary	5
1. Introduction	8
2. Methodology.....	8
2.1. Introduction	8
2.2. Time in Santiago, Antofagasta, and Biobio	9
2.3. Access to information.....	9
2.4. Meetings	9
2.5. Documents and maps	9
2.6. Field visits.....	9
2.7. Census data files.....	9
2.8. Special tabulations	10
2.9. Request to INE regional offices	10
2.10. Meeting with INE regional office directors.....	10
2.11. International principles, recommendations, and practices.....	10
3. Preparatory Work.....	11
3.1. Introduction	11
3.2. Type of enumeration	11
3.3. Method of enumeration	11
3.4. Timing of enumeration	11
3.5. Census reference time.....	12
3.6. Duration of enumeration.....	12
3.7. Questionnaires	12
4. Field operations.....	14
4.1. Introduction	14
4.2. Methodology.....	14
4.3. Planning	16
4.4. Pre-census.....	16
4.5. Demarcation of enumeration areas	17
4.6. Production of census maps.....	17

4.8. Census tests	21
4.9. Staff recruitment and training	21
4.10. Enumeration	21
5. Data Processing	24
5.1. Introduction	24
5.2. Methodology.....	24
5.3. Stages of data processing	24
5.4. Data capture and coding.....	24
5.5. Incorporation of delayed questionnaires	25
5.6. Rectification of housing unit and household codes	25
5.7. Producing person, household and housing unit records.....	26
5.8. Editing for inconsistencies and missing values	26
5.9. Imputation for non-responding housing units.....	27
5.10. Imputation for unobserved housing units	29
5.11. Calculation of missing value rates	29
5.12 Missing value rates for person variables (unedited data)	31
5.13. Analysis of the main stages of data processing for selected areas	34
6. Census products and services.....	35
6.1. Introduction	35
6.2. Methodological and administrative report.....	36
6.3. Standard tables	37
6.4. On-demand tabulation service	38
6.5. Statistical briefs.....	39
6.6. Evaluation studies	40
6.7. Public use microdata sample	40
7. Post-enumeration survey.....	40
7.1. Introduction	40
7.2. Definitions.....	41
7.3. The 2012 census post-census survey (PCS)	41
7.4. PCS sample design.....	43
7.5. PCS data collection	43
7.6. PCS matching methodology.....	44
7.7. PCS field follow-up operation not undertaken	45

7.8. PCS estimation methodology.....	45
7.9. Conclusion and recommendation	46
8. Demographic equation estimates.....	46
8.1. Introduction	46
8.2. Demographic equation estimation.....	47
8.3. Demographic estimate of the 2012 census population	48
8.4. Demographic estimates of the 2012 subnational population	49
8.5 Demographic equation estimates and projection estimates	50
9. Sampling of non-responding housing units.....	51
9.1 Sub-population sampling of non-responding housing units.....	51
9.2 Sample Design.....	51
9.3 Margin of Error Aspect	51
9.4 Operational Aspects	52
10. Chile’s next population and housing census	53
10.1. Introduction	53
10.2. Planning	53
10.3. Type of enumeration	57
10.4. Reference time.....	57
10.5. Length of enumeration period.....	57
10.6. Census date.....	57
11. Conclusions and recommendations.....	58

Executive summary

The International Commission on the 2012 Population and Housing Census of Chile began its work on 2 September and concluded on 22 November 2013. Commission members Griffith Feeny and Rajendra Singh spent eight weeks in country, Commission member Roberto Bianchini 4 weeks.

The Commission was given full access to INE staff, documents and census data. Members spent over 70 hours in meetings with INE staff, and conducted a limited review of documents and maps, including census questionnaires, planning and enumeration area maps, enumerator and supervisor manuals, data processing specifications, and published census reports.

The Commission reviewed preparatory work leading up the enumeration, field operations during the enumeration and subsequent data processing through meetings with INE staff, documents and field visits. Observations were assessed in relation to international good practice as given, for example, in the United Nations Principles and Recommendations for Population and Housing Censuses.

The 2012 census was a *de jure* census (the 2002 census was *de facto*). It used the interviewer method of enumeration with paper questionnaires. Enumeration was planned for 9 April 2012 through end June, but continued in some areas through end July. No census reference time was specified.

The main census questionnaire sections are (A) identification, (B) housing unit data, (C) household data, (D) household roster and (E) person data. Questionnaires provided space for recording information on up to six persons.

Field visits to selected enumeration areas (EAs) in Santiago Metropolitan area, Antofagasta Region and Biobio Region. Preparation for field visits included meetings with INE cartography staff on the 2011 pre-census and subsequent EA demarcation and with data processing staff on outgoing and incoming logistics.

Commission members visited EAs accompanied by INE Regional Directors and staff. Arriving at an EA, they identified the starting point for enumeration of the EA on the ground and on the EA map, followed the path the enumerator would have followed, and correlated observations on the ground with the EA map.

Field visits were preceded or followed by office meetings with regional office staff to understand the overall process of enumeration, including the role of supervisors. In Biobio region, the only field visit with Roberto Bianchini present, a meeting reviewing the process lasted over seven hours.

Field operations were also assessed by identifying comunas with extreme values for several indicators and asking regional directors questions on measures taken to ensure the quality of enumeration for these comunas. A meeting with all 14 regional directors was then held in Santiago to clarify their responses.

Commission members identified seven stages of data processing: data capture and coding; incorporation of delayed questionnaires; rectification of housing unit and household codes; production of person, household and housing unit records; editing these records for inconsistencies and missing values; imputing person/household records to non-responding housing units; and imputation of unobserved housing units.

Data processing was assessed by the clarity with which the stages and the format of the output files were defined, by data capture error rates, and by missing value rates. An INE sample-based study found data capture error rates of 0.14% for housing units and 0.29% for persons. Missing value rates for the great majority of person variables were very low. High rates were observed for half a dozen variables, including variables for persons born outside Chile.

The data processing unit provided scanned questionnaire images and counts of housing units and persons for the EAs observed in the field visits at each stage of processing, making it possible to see the result of each stage of processing for each EAs.

Based on these observations the Commission concluded that field operations and data processing were generally conducted to a high standard and in accordance with international principles, recommendations and guidelines. Exceptions were noted, such as the non-specification of a census reference time, but these exceptions do not substantially impair the usability of census data.

Evaluation of census results, as distinct from the processes that produced these results, is problematic. The Post-Census Survey as implemented cannot be used to estimate census omission for several reasons: lack of a census reference time; the long delay between the beginning of the census enumeration and the beginning of the PCS; matching procedure difficulties due to first name only being collected; and the absence of field follow up for non-matched persons.

Demographic equation estimates as a method for assessing completeness of census coverage at the national level due to the lack of evaluation studies providing estimates of completeness of birth and death registration. At the sub-national level, in particular for comunas, they are impossible because no data on inter-comuna migration flows for the period between the 2002 and the 2012 census exists.

It is obviously desirable to have direct evidence of the quality of census results, in particular of completeness of enumeration, as well as the indirect evidence on quality of census processes. Based on the available evidence, however, the Commission finds no basis for doubting the usability of the census data for most, if perhaps not all, of the customary uses.

The Commission accordingly recommends that the results of the 2012 census of population and housing be made available to users as soon as possible, subject to two essential conditions.

First, the Commission recommends that the imputation of housing units not observed on the ground during the enumeration be eliminated.

Second, the Commission recommends no census results be released until they can be accompanied by a methodological and administrative report documenting census preparatory work, field operations and data processing.

The commission further recommends that, when these two conditions have been met, INE prepare the following census products and services: the methodological and administrative report; standard census tables; an on-demand tabulation product/service that enables users to produce and table they specify, subject to conditions necessary to ensure confidentiality; statistical briefs to promote awareness and encourage use of the census results; evaluation studies of the quality of census data; and if use demand so indicates, a public use microdata sample.

The above does not have to be produced at the same time, but planned and disseminated with a calendar. Census tables and the methodological and administrative report are the first products we recommend to release.

The Commission recommends, in response to a specific request from INE, that the proposed sample of non-responding housing units not be carried out because it is unlikely to significantly improve the census results.

The Commission recommends that INE take the following considerations into account in deciding the date of the next population and housing census.

First, given the problems, real and perceived, of the 2012 census, the Commission recommends that a full pilot census be conducted on year before the next census.

The date of the next census should take into consideration also the length of time required for census preparatory work, all of which must be carried out regardless of the length of the questionnaire.

In addition to the usual preparatory work, the Commission recommends that INE research emerging census methodologies and consider whether or not they may be appropriate for the next census of Chile.

1. Introduction

The 2012 population and housing census of Chile was carried out by the National Institute of Statistics (INE) from April 9 to end July 2012. Census results were announced on 4 April 2013.

On 26 April 2013, following public allegations of inappropriate processing of the census data, the chairman of INE submitted his resignation. On 29 April, Juan E. Coeymans became chairman of INE. On 2 May, publications of census results that had been placed on the INE website were removed from site.

On 9 May INE appointed an External Commission to investigate the issues raised in the press. The External Commission delivered its report, *Informe final Comisión Externa Revisora del CENSO 2012*, on 7 August 2013. The report is available on the INE website (www.censo.cl/informe_final.pdf, visited 11-Nov-2013).

On 23 August 2013 INE appointed an International Commission to further investigate census issues. The Commission was composed of three independent international experts on population and housing censuses. The experts carried out the work without any interference from INE and from any other institution or person. This is the report of the International Commission.

International Commission members visited Santiago three times during September-November 2013 and made field visits in Antofagasta and Biobio regions, and in Santiago Metropolitana.

The following section describes the methodology we followed to arrive at our conclusions and recommendations.

Sections 3-7 discuss phases of the 2012 census activity in chronological order.

Section 8 discusses demographic analysis for census evaluation.

Section 9 discusses matters pertinent to planning the next population and housing census.

Section 10 summarizes and extends the conclusions and recommendations made in preceding sections.

2. Methodology

2.1. Introduction

This section describes the methodology used by the International Commission. Additional detail on methodology is provided in several subsequent sections.

The United Nations *Principles and Recommendations for Population and Housing Censuses* (full reference in section 2.11 below) states that census evaluation should include assessment “of census operations” as well as evaluation of census results (paragraph 1.381).

This report focuses mainly on evaluation of census planning, processes and operations as detailed in Chapters II, III and V of the *Principles and Recommendations*.

2.2. Time in Santiago, Antofagasta, and Biobio

Members of the International Commission visited Santiago during 2-13 September, from 7 October to 1 November, and during 11-22 November 2013.

Griffith Feeney and Rajendra Singh were present through these periods. Due to prior commitments, Roberto Bianchini was in Santiago from 15 October to 1 November and during 14-22 November 2013.

Feeney and Singh made field visits in Santiago on 6 September and in Antofagasta region on 8-9 September. Bianchini, Feeney and Singh made a field visit to Biobio region on 24-25 October 2013.

2.3. Access to information

INE provided the International Commission full access to INE staff, documents, and census data sets. All substantive requests for meetings with staff, documents and data sets were met on a timely basis. Commission members signed a confidentiality agreement, “Confidentiality Agreement and Access to Data Protection by Statistical Secrecy”.

2.4. Meetings

The Commission spent over 70 hours in meetings with current and former INE staff, members of the External Commission, CELADE staff, and other persons interested in the census.

2.5. Documents and maps

The Commission conducted a limited review of available documents and maps, including census questionnaire forms, planning and enumeration area maps, enumerator and supervisor manuals, field operations forms, data processing specifications, and published census reports.

INE provided full or partial English translations of documents as requested by the International Commission. Because of the volume of documentary material, International Commission members frequently reviewed source documents with bilingual INE staff.

2.6. Field visits

The Commission made field visits to INE regional offices and census enumeration areas (census sectors) in Antofagasta, Biobio, and Santiago. Enumeration areas visited were selected by the regional office head based on criteria we specified. During the field visits we collected enumeration materials available in the field but not in INE headquarters in Santiago.

2.7. Census data files

Following the visits we requested and received copies of computer data files for the enumeration areas (EAs) visited. The files received for each EA included the output files of each stage of data processing described in section 5 below. The files were provided by the

INE Logistics and Information Systems Unit (Unidad de Gestión, Logística e Informática Censo).

2.8. Special tabulations

The Commission requested and received from INE various special tabulations of census data sets at various stages of processing.

2.9. Request to INE regional offices

The Commission identified four groups of comunas (administrative units) having relatively extreme values for four indicators and asked INE regional office directors to provide explanations for the extreme value for each comuna; to indicate what steps were taken during field operations to reduce extreme values; whether a document describing these steps could be provided; and whether values of indicators before and after these steps were taken could be provided.

2.10. Meeting with INE regional office directors

After receiving responses from the regional directors we met with all 14 regional directors and 6 zone chiefs of Santiago metropolitana for one day to seek clarification on their responses and to learn more of their experience during field operations.

2.11. International principles, recommendations, and practices

Information on the 2012 census was reviewed primarily in relation to United Nations publications on principles, recommendations and practices pertaining to population and housing censuses and vital statistics systems.

The following publications are most frequently cited in subsequent sections, which may also contain references to further publications.

[1] *Fundamental Principles of Official Statistics*. 1994-2013. United Nations Statistical Commission. Available in Arabic, Chinese, English, French, Russian, and Spanish at unstats.un.org/unsd/dnss/gp/gpintro.aspx, visited 13-Nov-2013.

[2] *Handbook of Statistical Organization: The Operation and Organization of a Statistical Agency*. Third edition, 2003. United Nations Department of Economic and Social Affairs, Statistics Division. Studies in Method, Series F No. 88. ST/ESA/STAT/SER.F/88. Available in Arabic, Chinese, English, and Spanish at unstats.un.org/unsd/dnss/HB/, visited 13-Nov-2013.

[3] *Principles and Recommendations for a Vital Statistics System Revision 2*. United Nations Department of Economic and Social Affairs. Statistics Division. ST/ESA/STAT/SER.M/Rev. 2. Sales No. E.01.XVII.10. ISBN 92-1-161439-2. Available in Arabic, Chinese, English, French, Russian and Spanish at unstats.un.org/unsd/pubs/gesgrid.asp?id=264, visited 17-Nov-2013.

[4] *Principles and Recommendations for Population and Housing Censuses*. Rev. 2, 2008. United Nations Department of Economic and Social Affairs. Statistics Division. Statistical papers Series M No. 67/Rev.2. ST/ESA/STAT/SER.M/67/Rev.2. New York, 2008. Available in

Arabic, Chinese, English, French, Russian, and Spanish at unstats.un.org/unsd/demographic/sources/census/census3.htm, visited 13-Nov-2013.

[5] *Handbook on Census Management for Population and Housing Censuses*. 2001. Department of Economic and Social Affairs. Statistics Division. Studies in Methods, Series F No. 83/Rev.1. ST/ESA/STAT/SER.F/83/Rev.1. New York: United Nations. Available in Arabic and English at the following URL, visited 15-Nov-2013. unstats.un.org/unsd/censuskb20/KnowledgebaseArticle10066.aspx

[6] *Handbook on Geospatial Infrastructure in Support of Census Activities*, 2009. United Nations Department of Economic and Social Affairs, Statistics Division, Studies in Methods, Series F No. 103. ST/ESA/STAT/SER.F/103. Sales No. E.09.XVIII.8. ISBN: 978-92-1-161527-2. New York: United Nations.

[7] *Handbook on Population and Housing Census Editing*. Rev. 1, 2010. United Nations Department of Economic and Social Affairs. Statistics Division. Studies in Methods Series F No. 82/Rev.1. ST/ESA/STAT/SER.F/82/Rev.1. New York: United Nations.

[8] *Post Enumeration Surveys – Operational guidelines – Technical Report*. 2010. Department of Economic and Social Affairs, Statistics Division. 2010 World Population and Housing Census Programme. New York: United Nations. Available at unstats.un.org/unsd/censuskb20/KnowledgebaseArticle10685.aspx, visited 17-Nov-2013.

3. Preparatory Work

3.1. Introduction

Most preparatory work for the 2012 census lies outside the scope of this report. Several key features are noted in this section to provide background for subsequent sections.

3.2. Type of enumeration

The Chile 2012 population and housing census enumerated persons at their usual place of residence. It was thus what is often called a *de jure* census. Usual residence, defined in the enumerator manuals, generally follows the United Nations recommendation.

The Chile 2002 population and housing census enumerated persons at their place of enumeration, that is, it was a *de facto* census. Populations of sub-national areas from the 2012 census are accordingly not comparable with the populations of these areas from the 2002 census. There are no statistical studies of the extent of the incomparability so far as we have been able to determine.

3.3. Method of enumeration

The 2012 census used the interviewer method of enumeration with paper questionnaires.

3.4. Timing of enumeration

The 2012 census enumeration began 9 April 2012. It was planned to be completed by June 30, but some canvassing continued through end July.

3.5. Census reference time

The United Nations *Principles and Recommendations for Population and Housing Censuses* (Rev. 1, 2008) defines a population census as the “process of collecting, compiling, evaluating, analysing and publishing or otherwise disseminating demographic, economic and social data pertaining, *at a specified time*, to all persons in a country ...” (emphasis added).

The *Principles and Recommendations* likewise identifies four essential features of a population and housing census: individual enumeration, universality within a defined territory, simultaneity, and defined periodicity. Of “simultaneity” it is stated that “Each person and each set of living quarters should be enumerated as of *the same well-defined point in time ...*” (emphasis added).

The United Nations *Handbook on Census Management for Population and Housing Censuses* likewise states that “an essential feature of a census is that each person, or each set of living quarters, is enumerated with reference to *the same predetermined point in time*. This census reference time is usually midnight at the beginning of the designated census day” (emphasis added).

No reference time was specified for the 2012 population and housing census. This may be due in part to the change to a usual residence census from the 2002 census one day *de facto* census, for which something close to a reference time is implicit.

A reference time should be specified for the next population and housing census in Chile.

3.6. Duration of enumeration

The United Nations *Principles and Recommendations for Population and Housing Censuses* states that “It is desirable to keep the enumeration period short in order to avoid double counting and omissions, which can occur in spite of a single reference date” (paragraph 1.287).

In most developing countries, it is stated, “the enumeration period has generally varied from a few days to two weeks” (paragraph 1.28). No reference to typical durations of enumeration in developed countries. We are aware of no global summary statement of country practices.

3.7. Questionnaires

The United Nations *Principles and Recommendations* states that “two general frameworks within which individuals are identified: (a) households and (b) institutions” and that “The household is a general framework within which most individuals are identified” (paragraph 1.442).

Most households occupy housing units (*viviendas particulares*), but households may be homeless (paragraphs 1.452 and 2.109). The 2012 census did not attempt to enumerate homeless persons.

Individuals may also be enumerated in collective living quarters (viviendas colectivas), which include, for example, hotels, institutions, camps and worker's quarters.

The *Principles and Recommendations* classifies living quarters in paragraph 2.415, which includes Figure 7.

Three questionnaires were used for the 2012 census enumeration, a questionnaire for households living in housing units (Cuestionario Censal – Viviendas Particulares), a questionnaire for persons living in collective quarters (Cuestionario Censal – Viviendas Colectivas), and a questionnaire for persons not usually resident in the country. Questionnaires of the last type were processed but not tabulated.

The questionnaire for households living in housing units consists of five sections.

- A. Identification
- B. Housing unit data
- C. Household data
- D. Household roster
- E. Usual residents (persons)

Section A contains information to identify the household and housing unit for which information is recorded. This information includes the enumeration area (EA) barcode (numero de portafolio), from which the code for region, province, comuna, district, area, district/locality, zone/entity, block and enumeration area may be determined.

The housing unit is uniquely identified by the portafolio code and the housing unit code. The household is uniquely identified by the portafolio code, the housing unit code and the household code. The household code is "00" for unoccupied housing units, "01" for the first household in a housing unit, "02" for the second household in a housing unit, and so on.

Households of more than 6 persons, the maximum number provided for by the questionnaire, require additional questionnaires. These are identified by the portafolio code, the household code, and a continuation code at far right.

Section A also includes information on the street address of the housing unit.

Section B contains housing unit information.

Section C contains household information.

Section D provides space to list nine persons. Relation to head of household is recorded in this section.

Section E contains person information. Though section D provides space to list nine persons, section E allows recording information for only six persons. Households of 7-12 persons require two questionnaires, households of 13-18 persons three questionnaires, and so on.

The questionnaire for persons living in collective living quarters contains the same questions as section E of the questionnaire for households living in housing units, though the item numbers are different.

4. Field operations

4.1. Introduction

The United Nations *Principles and Recommendations for Population and Housing Censuses* (Rev. 1, 2008) describes field operations for a population and housing census as a number of consecutive and parallel activities, which are largely connected each other. The main activities are: 1) planning; 2) mapping; 3) realization of census tests; 4) definition and functioning of field organization; 5) logistics; 6) implementation of a quality assurance programme; 7) staff recruitment and training; 8) enumeration; 9) census evaluation, 10) communication and publicity.

4.2. Methodology

This section describes field operations based on our observations during field visits, on our study of official documents, and on intensive discussions we conducted with INE staff directly involved in their implementation.

Census evaluation is described in chapter 7 of this report. No comprehensive quality assurance programme was planned and implemented in the 2012 census.

As mentioned above, the members of the International Commission conducted three separate field visits, one in the northern region of Antofagasta, one in the center of the country in Santiago metropolitan area and one in the southern region of Biobío, and met with the heads of the fifteen INE regional offices.

A technical meeting was also organized with the Unit *Tecnología Cartográfica* of the Department *Infraestructura Estadística y Tecnológica* of INE.

Several census enumeration areas were visited to learn how 2012 census field operations were organized and carried out in the field. We identified the type of enumeration areas we wanted to visit with the intention of gaining some understanding what may have caused omission in the census. We also considered visiting the areas that may have caused differences in comuna population due to change from *de facto* to *de jure* census.

Considering the possible causes for omissions and the difference in 2012 census type, we decided to visit enumeration areas in low, middle and upper income areas, in areas difficult to enumerate areas for security, transport scarcity or other reasons, areas affected by natural disasters, and mining areas.

INE provided us with enumeration area maps, pre-listing and census listing sheets and the regional office staff to accompany us and answers questions on field operations during the census. In each enumeration area, we followed the path that an enumerator was supposed to follow.

In addition to our field observation, we requested information from all INE regional offices for selected comunas. The comunas were selected for having extreme values for one or more of four indicators. The extreme values were chosen somewhat arbitrarily based on an initial inspection of comuna data. Initially, we considered only the four indicators, but resulted in a large number of comunas including many very small comunas that would have little effect on data quality. We therefore added a further criterion to select only larger comunas. Four groups of comunas were formed based on four indicators – one group for each indicator.

These four indicators for identifying comunas with extreme rates were as follows.

1. Percent of housing units unoccupied greater than or equal to 15%, only comunas with 12,000 or more housing units.
2. Percent of non-responding housing units greater than or equal to 5%, only comunas with 5,000 or more housing units.
3. Percent difference between the 2012 census number of housing units and the 2011 pre-census number of housing units less than or equal to -5%, only comunas with 4,000 or more housing units.
4. Percent difference between the 2012 census number of housing units and the number of housing units projected from the 2002 census more than 20% or less than -20%, only for comunas for with 40,000 or more housing units as of the 2012 census.

INE provided the comuna data required to compute these indicators. The comunas so identified are listed in Annex 2.

These lists of comunas were sent to INE regional office directors. The directors were asked to answer the following four questions for each comuna.

Questions for Regional Offices to Respond:

1. Provide an explanation (or causes) for a rate to be so extreme in your comuna.
2. What steps, if any, did you take to reduce such an extreme rate?
3. If needed, would you be able to provide a document that describes your procedure(s) that you used during the census operations to reduce these extreme rates?
4. Do you have any data to show the rates before and after you applied a procedure to reduce the high rate in your comuna?

The regional office responses are summarized in Annex 3.

4.3. Planning

Planning for 2012 census field operations began in 2008 with a decision to carry out a *de facto* census in 2012. It was also decided to introduce new technologies, including Geographic Information System (GIS) tools for data dissemination and Global Positioning System (GPS) receivers, and to carry out a pre-census project in preparation for 2012 census field operations. Plans were summarized in a Gantt chart, which we observed.

We were informed by INE staff that plans for 2012 census field operations were monitored and revised as appropriate from time to time, between 2008 and beginning of 2012.

The first phase of 2012 census planning ended in 2010 with a decision on the topics to be covered, preparation of base maps, organization of the pre-census, and establishment of the National Census Commission.

A second planning phase during 2010-2011 covered implementation of the pre-census, demarcation of enumeration areas, preparation of census questionnaires and manuals, organization of staff recruitment, and development of a census communication plan, including a publicity campaign.

Between 2010 and April 2012, regional, provincial and municipal census commissions were established, and final census maps prepared. Enumerator portfolios containing blank questionnaires and other enumeration materials were prepared for each enumeration area and distributed to local census offices throughout the country. The publicity campaign was implemented.

A major revision to the plan occurred when it was decided in August-September 2011 to carry out a *de jure* rather than a *de facto* census.

4.4. Pre-census

Pre-census activities began in January 2011 in the extreme regions of the country (south and north), and ended in November 2011.

The objective of the pre-census was to acquire in the field updated information (mainly location and number of housing units) across the country for the demarcation of the enumeration areas. This information was subsequently used for planning field operations (determining the number of census staff and census questionnaires in each area, for example). Pre-census activities were managed by INE regional offices following guidelines provided by INE headquarters.

Map updating and housing unit listing carried out during the pre-census phase was based on the cartographic materials prepared by INE headquarters. The *Tecnología Cartográfica* of the Department *Infraestructura Estadística y Tecnológica*, worked for about two years for the preparation of the base maps, from April 2009 to March 2011. The work involved about 50 INE staff and followed a timetable of activities developed in 2008.

Digital cartographic data used by INE as baseline for the implementation of pre-census activities was acquired from the *Instituto Geográfico Militar*. Topographic maps at scale

1:50,000 were used for rural areas City plans at scale 1:5,000 from the *Servicio Aerofotogramétrico* of the Air Force, were used for urban areas.

Satellite images or aerial photographs were used in some parts of the country to update base maps, to locate populated areas, and to identify new construction, especially in rural areas. Base cartographic materials prepared by INE covered all 346 comunas of the country. Populated areas were provisionally delineated in each map of comunas, and for those areas larger scale maps were also prepared until the level of EAs.

4.5. Demarcation of enumeration areas

Continental Chile covers an area of approximately 757,000 sq. km. The 2012 census was carried out also in the Insular and Antarctic territory. The continental country has a large variety of landscapes, desert areas in the north, the Andes Mountains in the east, and forest areas in the south.

Administratively, the country is divided into 15 Regions, 54 Provinces and 346 comunas. For census purposes, comunas were divided into census *districts*, classified as urban or rural. Urban districts were subdivided into *zones*, zones into *blocks* (*manzana*), and blocks into enumeration areas (sectors). Rural districts were divided into *localities*, localities into *entities*, and entities into enumeration areas. Urban districts were delimited by “census urban boundaries”.

A different mapping approach was used for urban and rural areas. For rural areas, housing units were located, coded and represented as points with the support of Personal Digital Assistants (PDAs) with built-in a GPS receiver.

In urban areas administrative address system was used. An ID code was assigned to each housing unit, from 1 to *n* at entity level in rural areas, and from 1 to *n* at “Manzana” level in urban areas.

On the basis of the results of the map updating and housing listing, enumeration areas (EAs) were demarcated to contain approximately 12-15 housing units. EAs were demarcated according to the original plans for a *de facto* census, following the criteria that one enumerator could have interviewed one EA in one day.

4.6. Production of census maps

INE Regional Offices organized and supervised map updating activities and the collection of data on housing units during the pre-census phase. The updated maps and data on housing units were then digitized and census maps prepared for each enumeration area, identified by the unique portafolio number. Figures 4.1 and Fig. 4.2 show illustrative EA maps for urban and rural areas, respectively.

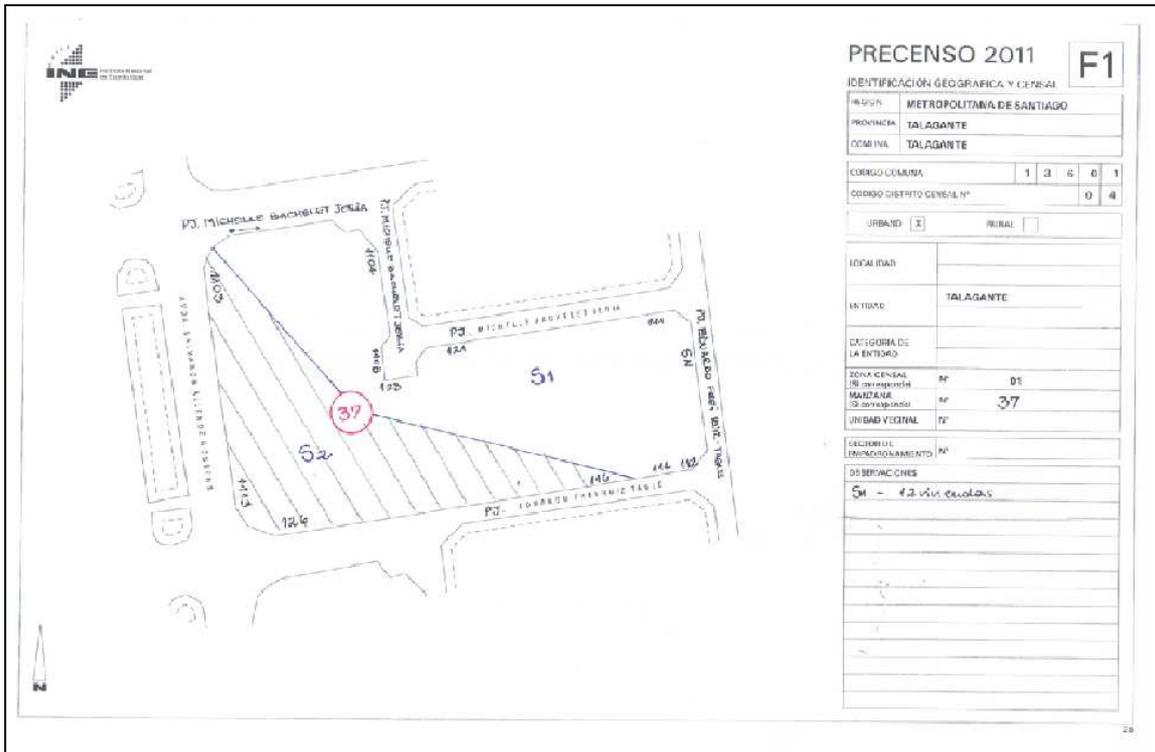


Fig. 4.1. Example of EA map for urban areas.

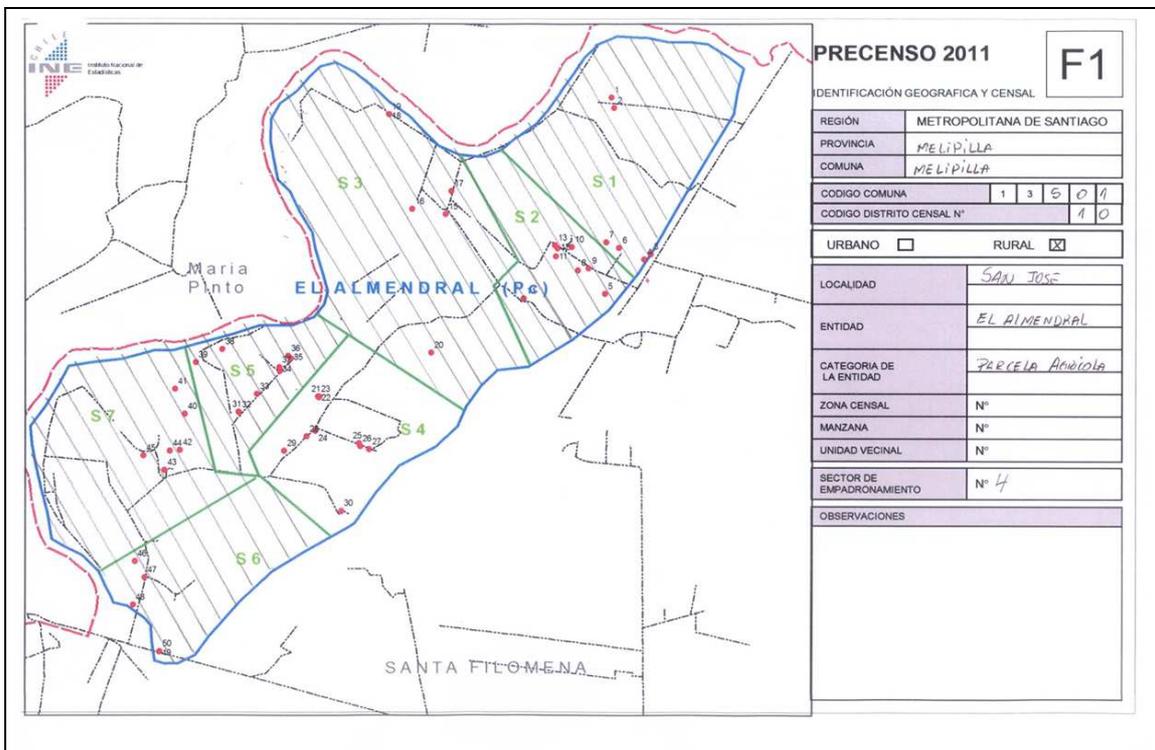


Fig. 4.2. Example of EA map for rural areas.

In addition to the EA census maps for each enumeration area, maps were prepared for comunas, districts, villages and other populated areas within the rural district for rural areas, and for cities, zones and “Manzana” for urban areas. Each of the above listed type of maps were prepared at different scale, ensuring a full coverage of the mapped areas. In urban areas, “Manzana” were usually identified with blocks. Each Zone was containing from 200 to 300 blocks. Each block was containing one or more EA.

The International Commission assessed the demarcation of EAs and the preparation of census maps as in general appropriate and in line with international recommendations (*UN Principles and Recommendations paragraph 1.132; UN Handbook on Geographical Infrastructure paragraphs 3.19, 3.54*). A general practice in census mapping programmes, however, is to use of remote sensing data to support the location of housing units and the orientation of enumerators in the field, especially when census maps do not show buildings.

Quality assurance procedures and field tests should also be seen as important aspects of census mapping operations to improve the coverage of housing units and persons. In addition, no systematic map updating was carried out between the end of pre-census activities and the commencement of the general enumeration. This may have resulted in some omission of housing units and persons in areas with large numbers of newly constructed housing units.

4.7. Census organization and logistics

The organization of the 2012 census was similar to other countries where a “traditional” census was conducted, with direct interviews to respondents made by enumerators.

A Census National Commission, chaired by the Minister of Economy, was established in 2010 to coordinate and supervise all census operations. Census Regional, Provincial and Municipal Commissions were also established to facilitate the preparatory work, coordination and census operations, on the basis of guidelines from INE.

INE was responsible for census operations at the central and regional level. At the comuna level, a census manager was responsible for the coordination in the comuna, while local census managers were responsible for the coordination in the local census field offices.

Based on what we heard from INE staff, census materials were prepared in line with standards, packed and delivered to local census offices, and returned to INE Santiago according to the following flows (Fig. 4.3. and 4.4.)

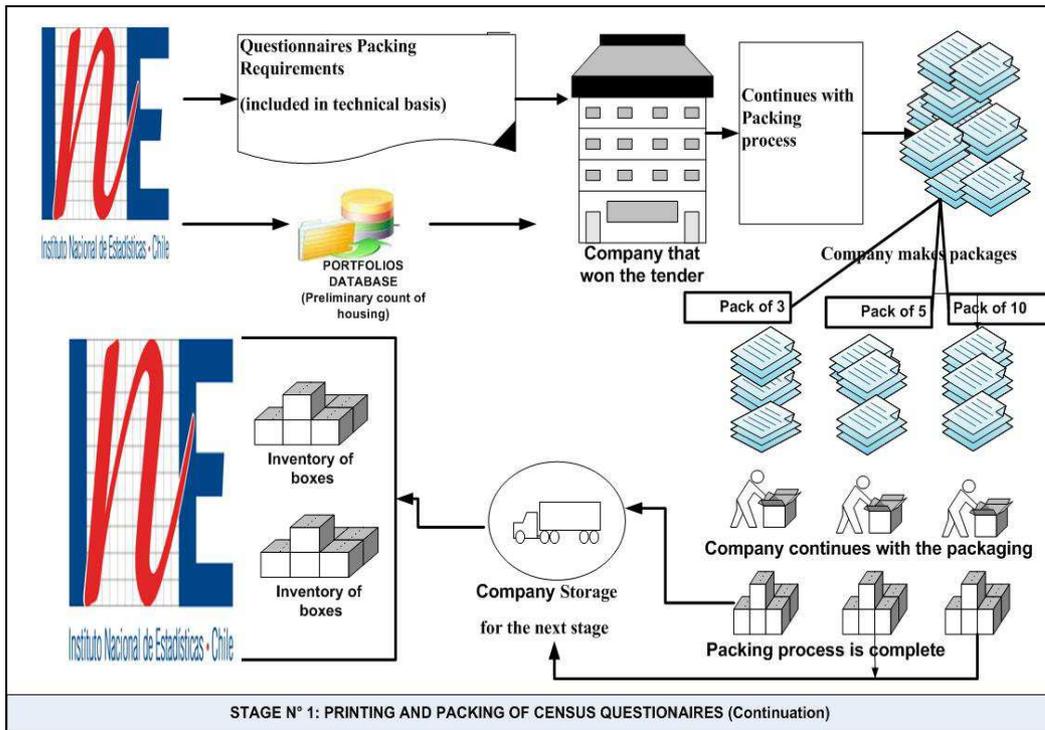


Fig. 4.3. Printing and packing flow

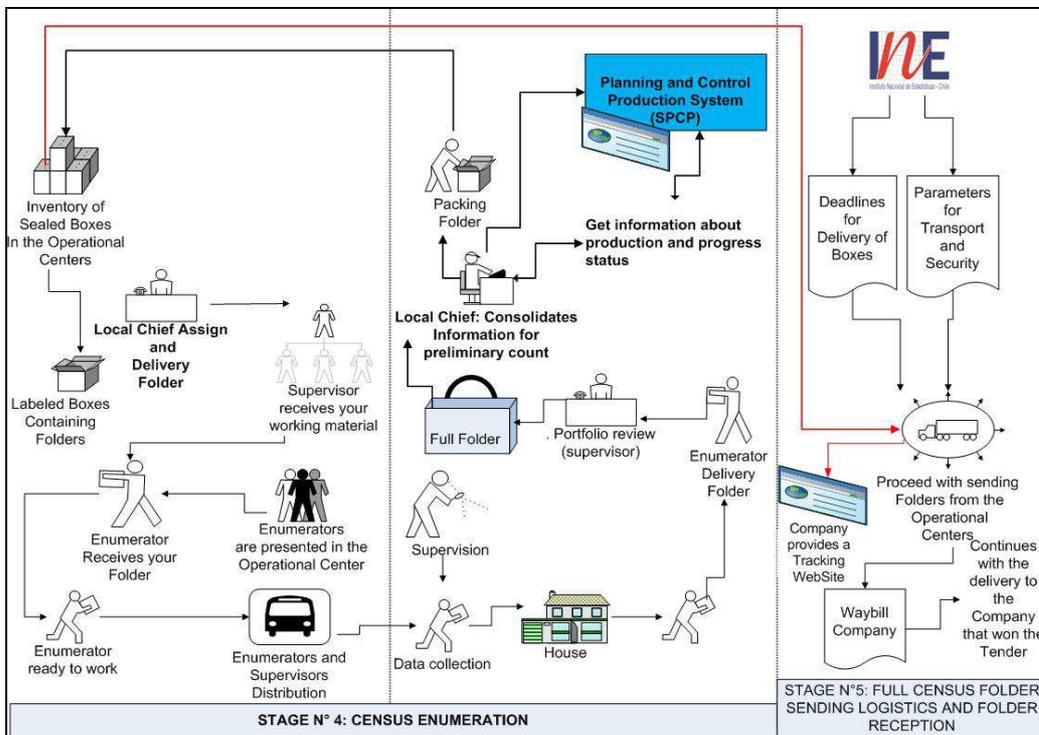


Fig. 4.4. Returning of census materials from field offices to INE Santiago

4.8. Census tests

A proposal to change from a *de facto* to a *de jure* made by INE in June 2011 was accepted by the Census National Commission and formally adopted in September 2011. Chile had not previously carried out a *de jure* census, and INE did not carried out a pilot census. Testing was limited to testing of the revised questionnaires.

Some aspects of field operations, such as the packing, delivery and return of census materials, are similar for *de facto* and *de jure* censuses. Other aspects, such as recruitment and training of the staff, field organization, monitoring and reporting during the enumeration are very different.

Implementation of a pilot census is a widely accepted practice and is internationally recommended (*UN Principles and Recommendations paragraph 1.195*). It is a main milestone to evaluate the quality of census operations (*UN Principles and Recommendations paragraph 1.430*).

A pilot census would have also supported to measure the level of participation and the response rate of the population, and the appropriate use of the concept of “usual resident” which characterize a *de jure* census.

4.9. Staff recruitment and training

Enumerators were recruited with the assistance of an external private company that provided lists of candidates for different positions based on criteria and procedures specified by INE.

Contracts with field staff were made by INE regional offices, who supervised the recruitment of the staff. In total, about 15,000 staff were recruited, about 11,000 full time enumerators, 2,500 part-time enumerators, 1,700 supervisors and 370 local census managers.

Even with 20% of staff planned as reserve, in some areas of the country many enumerators resigned during the enumeration period and some local census offices faced difficulties to replace them.

Training for INE staff was carried out in January 2012. Training for supervisors and enumerators was carried out in March and early April 2012. The training followed a “cascade” approach. Supervisors and full time enumerators received 5 days training. Part-time enumerators received 3 days only of training.

4.10. Enumeration

The enumeration started on 9 April 2012. In some areas of the country it continued through end July 2012, a period of almost 4 months (section 3).

The census enumerators canvassed the housing units in their assigned enumeration areas, making changes to the pre-census housing unit list when needed, as explained in the instruction manuals.

During our field visits we walked with the INE staff to observe how the census was carried out. We took the materials that an enumerator would carry to an enumeration area. We used maps to identify the selected enumeration area and found the point at which the enumerator was supposed to begin the enumeration.

We compared the housing units on the map with the housing units on the ground to see that the two were consistent with the map. We compared the pre-census list of housing units with the census list. In the areas we visited, we did not find inconsistencies or omitted units. In enumeration areas we observed that a number of housing units were added to the pre-listing. We noticed also that census enumerators added correctly number housing units that were not listed during the pre-list operations.

We were told by INE regional staff, however, that in some areas the maps had some problems, such as some neighboring comunas having overlapping boundaries, missing or confusing street names, duplication of zones, enumeration maps did not correspond to the expected areas because of some errors in the geographic codes reported in the form F1.

We asked what quality checks were used during the census operations, checking the classification of housing units as occupied and unoccupied, for example.

Because the census was taken more than a year ago, it was not possible to check the accuracy of occupied-unoccupied classification. We did however investigate what steps were taken if enumerator identified a housing unit either unoccupied or non-responding. We were told that the supervisors went to such housing units to verify the accuracy of the classification. The same response was received from all regional offices.

The regional offices also implemented procedures to check the quality of enumerators. If an enumerator failed the quality check, they were dismissed. In one regional office, 7 out of 120 enumerators were dismissed due to errors in classification of housing units.

Based on our field visits, on discussions with the regional directors, and on responses from the regional offices to our questions on outlier comuna, we note the following factors as potentially explaining high frequencies of unoccupied and non-responding housing units.

1. Many housing units were built as second houses for vacation or holidays
2. Population growth in some comunas did not keep up with the housing unit growth
3. Some comunas suffered drought for a several years that caused lack of job opportunities in the area and people left for other areas with better job opportunities

4. A large number of new housing units were built in comunas that were badly damaged by earthquake and tsunami but they were not occupied yet.
5. Bedroom communities where occupants come late to sleep and leave early to go to work in another community
6. Difficult access to condominium buildings
7. Interviewing time was limited in dangerous areas even with police protection

Actions taken by INE regional offices to reduce the frequency of occupied housing units misclassified as unoccupied and of non-responding housing units included the following:

1. Implemented the quality process to verify the accuracy of classification
2. Analyzed the data of growth of housing unit and population since 1992 census to check the validity of the unoccupied housing unit rates
3. Reviewed the records of the ministry of housing, and regional governments to check the validity of the unoccupied housing unit rates
4. Followed the regular census procedures regarding unoccupied housing units including many visits at different days and times of the day, getting information from neighbors and/or police
5. Supervisor verify each unoccupied housing unit for its correct classification
6. Enumerators went back to housing units on different days of the week and at different times of the day
7. Left note for occupants under the door with INE telephone number to encourage residents to coordinate interview time
8. More media publicity to encourage residents to participate in census
9. At least two regional directors requested INE Santiago if they could conduct proxy interviews with knowledgeable persons
10. Both supervisor and a special team consisting of enumerator and local leader worked in high risk and urban areas
11. Used police protection in dangerous and high risk areas
12. Followed the instructions in supervisors' and interviewers' manuals.

5. Data Processing

5.1. Introduction

Traditional, interviewer-based censuses using paper questionnaires have been carried out in countries throughout the world for many decades and remain the most common type of census enumeration. This section describes the various stages of data processing for the 2012 Chile population and housing census.

5.2. Methodology

Information on processing of the 2012 census was obtained through an extended series of meetings with INE data processing staff, through examination of census data sets for enumeration areas observed during field visits, and from special tabulations produced at our request by data processing staff.

5.3. Stages of data processing

Census data processing involves a more or less standard sequence of stages, beginning with data capture and ending with the production of census products and services. This report presents information on the following stages of 2012 population and housing census data processing.

1. Data capture and coding
2. Incorporation of delayed questionnaires
3. Rectification of housing unit and household codes
4. Producing person, household, and housing unit records
5. Editing of person, household and housing unit records
6. Imputing persons/household records to non-responding housing units
7. Imputing housing units not observed in field operations

5.4. Data capture and coding

Data capture, the process of converting information recorded on questionnaires to a format that can be processed by computer (*Principles and Recommendations* ¶1.303), was effected by guillotining questionnaires, scanning the resulting single pages, and applying optical mark recognition (OMR) and intelligent character recognition (ICR) software to the scanned images.

The output of this initial stage of processing is a computer database of census questionnaires containing (a) images of the scanned questionnaires and (b) files containing a record for each questionnaire showing the characters produced by the OMR/ICR software.

Computer assisted coding of questionnaire text matter (such as comuna and country of birth) was incorporated into data capture, so the output data base includes codes (comuna and country codes, for example) as well as the source text matter.

Data capture and coding were out-sourced to a private company experienced in census data processing. INE staff were present at the operations center where questionnaires were processed to exercise quality control checks and to operate computer assisted coding.

INE staff designed and executed a quality assurance study of data capture. The results are reported in the document *Verificación Lectura Óptica Información Censal, XVIII Censo Nacional de Población y VII de Vivienda 2012*, Instituto Nacional de Estadísticas, Junio / 2013 (64 pages). INE provided an English synopsis of this document titled “Summary of the Optical Character Reading Process” (provided by INE), attached as Annex 4. The last two paragraphs of this document are as follows. The stated error rates indicate good quality of data capture.

“In phase I, from paper format to digitalized database, a frame of 4,684,733 housing units were used, and it was estimated according to the sample that 0.14% presented problems in terms of housing unit data, corresponding to 6,526 housing units, with at least one element of data that was wrong. However, in Phase II, it was estimated that 0.42% of the housing unit suffered from some data error (19,564 housing units).

“In terms of usual residents, where the frame used corresponded to information for 15,681,893 inhabitants, it was estimated that there were errors in the data for persons in Phase I and II of 0.29% and 0.72%, respectively, corresponding to 45,867 persons with at least one data error in Phase I, and 113,431 persons in Phase II.

5.5. Incorporation of delayed questionnaires

The enumeration plan included a provision for “closing” portfolios for EAs when enumeration of the EA was deemed complete. As field operations proceeded, however, a decision was made to “re-open” some EAs, with enumerators returning to the field to renew attempts to carry out interviews for persons in housing unit units that had been classified as non-responding.

This required development of computer programs to incorporate the resulting “delayed questionnaires” into the data capture database, ensuring no duplication of questionnaires. This incorporation was based primarily on the street address information contained in Section A (Identification) of the main census questionnaire. The output of this stage was the database that would, if the original enumeration plan had been followed, resulted from data capture and coding.

5.6. Rectification of housing unit and household codes

The principal units of enumeration in a population and housing census are persons, households and living quarters (*Principles and Recommendations* Chapter V, Section A, “Units of enumeration”, paragraphs 1.442-1.460).

The data capture database consists of questionnaire records. The information on questionnaire records must be transformed into information on records for persons, households, and housing units for further processing.

Most housing units are occupied by a single household of six or fewer persons, and for these housing units and households there is a one-to-one correspondence between housing units, households, and questionnaires.

A housing unit may be occupied by two or more households, however, and a household may have more than six members. These conditions may occur in combination.

Person and household records for households of more than six persons must be created by merging the information from whatever number of questionnaires required to enumerate household members (one questionnaire for every six persons).

Records for housing units must be extracted from the multiple questionnaires used when, for whatever reason, there is more than one questionnaire for the housing unit.

These operations are effected using the housing unit and household codes in Section A (Identification) of the main census questionnaire. Before this can be done, however, missing housing unit codes and household codes must be supplied and incorrect codes corrected. That is the purpose of this “rectification” stage.

The output of this stage is a questionnaire database with the same format as the input database, but with housing unit and household codes corrected.

5.7. Producing person, household and housing unit records

This process takes questionnaire records with rectified housing unit and household codes as input and produces six types of records as output. The six record types are

1. Enumeration area geocodes (via portafolio barcode number)
2. Street address (Section A)
3. Housing variables (Section B)
4. Household variables (Section C)
5. Household roster (Section D)
6. Person variables (Section E)

All records are included in a single file, with record type codes in the first position identifying the record type.

5.8. Editing for inconsistencies and missing values

Editing refers here to (a) identifying and rectifying inconsistencies in person, household, and housing unit records and (b) supplying missing values for selected variables by imputation or other means. Editing was carried out using CSProX from Serpro Information System Engineering (serpro.com).

Editing specifications are contained in the following documents.

Metodología de Validación e Imputación para el Módulo de Población. Versión Final, Actualizada el 28/8/2013.

Metodología de Validación e Imputación para el Módulo de Población Colectiva. Versión Final, Actualizada el 28/8/2013.

Metodología de Validación e Imputación para los Módulos de Vivienda y Hogar. Versión Final, Actualizada el 4/7/2013.

Griffith Feeney and Rajendra Singh met with the INE data processing staff responsible for the editing to review editing for inconsistencies and missing values. A limited review of the first of these documents, *Metodología de Validación e Imputación para el Módulo de Población*, was made during this meeting.

We found the approach to the development of the editing specifications generally in accordance with the United Nations *Handbook on Population and Housing Census Editing*. We did not examine edit specifications for individual variables, nor did we ascertain the level of involvement of INE subject matter staff expertise in the development of the edits.

5.9. Imputation for non-responding housing units

Non-responding housing units (“viviendas particulares personas ausentes”) are housing units that were identified as occupied during the enumeration, but for which, despite repeated visits, it was not possible to complete an interview with a qualified occupant.

Housing unit non-response may occur because occupants refuse to come to the door, come to the door but refuse to be interviewed, or cannot be found at home by enumerators.

The United Nations *Handbook on Population and Housing Census Editing* addresses the issue of non-responding housing units as follows.

“If no population records appear for what is supposed to be an occupied unit, then the editing team must decide whether to count it as a vacant unit or substitute persons from another unit. ... If the unit is occupied, however, then the editing team must decide whether and how to assign persons from another unit with the same number of persons, with similar characteristics, if possible. Since it is impossible to know the characteristics of missing persons, this method should be used, if at all, only when the editing team decides it has no other alternative. (Chapter III, Section D, sub-section 1, “Vacant and occupied housing”, page 54, paragraph 255).

“Procedures for substituting whole households or individual missing persons are described elsewhere in this chapter. These procedures require assuming that the missing persons have the same characteristics as the substituted persons, which is almost certainly not usually the case, and the procedures themselves are very difficult. Still, without these procedures,

the counts of numbers of persons, and persons by characteristic, may decrease. (same, paragraph 258)

“... Several means are available for adding missing households when it is clear that they are in fact missing and need to be supplied. One method is to simply duplicate the previous household. However, if the number of people in the household is known, as is often the case (even if their characteristics are not known), it is possible to work backwards and duplicate the previous unit with the same number of people. Similarly, if you know the age and sex of the household members, that information can be of assistance in obtaining a substitute house. It is not advisable to try to use hot deck imputation to create information about household members, since this method often produces variables inconsistent with each other. (same, paragraph 261).

When households and persons are imputed to non-responding housing unit units, the general approach is to search for a “donor” housing unit whose occupants are considered likely to have characteristics similar to the occupants of the non-responding housing unit. Household and person records for the donor unit are then copied to the non-responding unit, either as they occur in the donor unit or with values of specified variables set to “missing”.

Implementation of this procedure requires various decisions that may be made differently for different censuses, guided by the principle that occupants of donor units should be as similar as possible to occupants in non-responding units. For the Chile 2012 census, for example, it was required that donor housing units be of the same housing unit type as the non-responding unit.

Type of housing unit is item 2 in section B of the main census questionnaire. There are 12 housing unit types: detached house, semi-detached house, apartment in building with elevator, apartment in building without elevator, traditional indigenous housing unit, room in old house or tenement, shack or precarious abode, ranch or hut, precarious housing unit made with reused/recycled materials, mobile home, and other.

For the 2012 census INE decided to impute households and persons to non-responding housing units. The procedure is described in a document provided by INE, in English, titled “Imputation of Occupied Housing Units with Absent Inhabitants”. This document is attached as Annex 5.

We discussed the imputation procedure described in this document in detail during a meeting with INE data processing staff. We concluded that if INE decides to impute households and persons to non-responding households, the method described in this document is acceptable.

5.10. Imputation for unobserved housing units

Following editing for inconsistencies and missing values and imputation of households and persons to non-responding housing units, a further imputation was carried out.

Examination of growth rates of population and housing units between the 2002 census, the 2011 pre-census, and the 2012 census suggested that some housing units were not captured during the enumeration. A decision was made to impute these unobserved housing units, and then to impute households and persons to the imputed housing units following a procedure similar to that used to impute households and persons to non-responding housing unit units.

The imputation of housing units that were not observed in census field operations, and of then imputing households and persons to these imputed housing units, is not a generally recognized international practice. We find no mention of it in the United Nations *Handbook on Population and Housing Census Editing*.

We recommend that this imputation not be part of processing of the 2012 census results.

5.11. Calculation of missing value rates

Missing value rates for the various kinds of information collected in a census or survey are an important generic indicator of data quality. Missing value rates for most 2012 population and housing census variables are presented in subsection 5.12 as evidence of the quality of field operations. This section describes the calculation of missing value rates.

For variables that apply to all persons, like age, sex, and relation to head of household, the calculation of missing value rates generally poses no difficulty. The rate for any variable is simply (a) the number of records in the data set that contain no value for the variable divided by (b) the total number of records in the data set.

Some variables, however, are defined only for subgroups of persons, these subgroups being defined by the values of other variables. Given the 2012 main census questionnaire design, which is not unusual in this respect, “Country of birth”, for example, item 23 in section E of the main census questionnaire, applies only to persons born outside Chile.

The correct denominator for the missing value rate for this variable is the number of persons born outside Chile. Figure 4.1 shows the place of birth item. If none of the three choices corresponding to codes 1-3 is marked, we do not know whether or not the person in question was born in Chile. We therefore do not know whether or not this record should be counted in the denominator of the missing value rate for country of birth. The missing value rate is indeterminate.

There are however two extreme assumptions. One is that *no* persons for whom no mark indicating one of the three codes was born in Chile. In this case the denominator of the missing value rate is the number of persons for whom code 3, “In another country”, is checked.

Figure 4.1. 2012 Census Place of Birth Question

The other is that *all* persons for whom no mark is shown were born outside of Chile. In this case the denominator of the missing value rate is the number of persons for whom code 3 is marked *plus* the number of persons for whom no code is marked.

We may therefore calculate two missing value rates and now that the true but unknown rate must lie between the two extreme values. This is how the missing value rate intervals in subsection 5.12 are calculated.

In this example there is a single variable, defined by the codes 1, 2, and 3, that determines whether or not there should be a value for the variable, country of birth, for which the missing value rate is calculated. In other cases, however, two or more variables may be used to “filter” the persons for whom a value should be available.

Consider for example item 41, number of surviving children, in section E of the main census questionnaire. This information is obtained only for (a) women who are (b) age 15 years old or older and who have (c) one or more children ever born. To know whether a value for number of surviving children should be available it is necessary to look at the values of three other variables—sex, age, and number of children born.

It is useful to have a general method that applies whenever knowing whether there should be a value for a given “filtered” variable depends on one or more “filter” variables. A general method is described below.

For any person, there are three possibilities.

Case 1. There should be a value for the given variable. This is the case only if the values of *all* the filter variables have values that indicate that there should be a value for the given variable.

Case 2. There should *not* be a value for the given variable. This is the case if *any* of the filter variables has a value that indicates that there should *not* be value for the given variable.

Case 3. It is unknown whether or not there should be a value for the given variable. This is the case if neither of the two preceding possibilities applies.

To calculate missing value rates in the general case the procedure is as follows.

Step 1. Create a “derived” variable with values 1, 2, or 3 for the above three cases.

Step 2. Create a derived variable for the given variable with value 1 if a value is present and 2 if no value is present.

Step 3. Cross-classify all persons by the derived variables defined in Steps 1 and 2.

Step 3 yields the table shown in Figure 4.2.

Filter Variables	Filtered Variable		
	Value	Missing	Total
Case 1: In domain	N[1,1]	N[1,2]	N[1,•]
Case 2: Unknown	N[2,1]	N[2,2]	N[2,•]
Case 3: Not in domain	N[3,1]	N[3,2]	N[•,•]

Figure 4.2. Calculation of missing value rates

Given the counts in this table, the bounds of the missing value rate interval are calculated as (1) $N[1,2]$ divided by $N[1, \bullet]$ and (2) $N[1,2] + N[2,2]$ divided by $N[1, \bullet] + N[2, \bullet]$.

5.12 Missing value rates for person variables (unedited data)

Table 5.1 shows missing value rates calculated as described in the preceding section for person data prior to editing for inconsistencies and missing values. These values provide a useful general indicator of data quality resulting from field work and data capture.

The United Nations *Handbook on Population and Housing Census Editing* states that

When the percentage of missing or inconsistent responses is low (less than one or 2 percent), any reasonable editing rules are not likely to affect the use of the data. When the percentage is high (5 to 10 per cent, or more, depending on the situation), simple, or even complex, imputation may distort the census results. (paragraph 65)

Table 5.1
Missing Value Rates (MVRs) for Person Variables,
Unedited 2012 Census Data

No.	Variable Name [Questionnaire ID]	MVR
1	Relation to head [D]	1.9
2	Sex [E19]	0.5
3	Age [E20]	0.4
4	Nationality Filter [E21a]	0.4
5	Foreign/Dual Nat. [E21bc]	(3,18)
6	April 2007 Resid. Filter [E22a]	0.0
7	Comuna April 2007 Res. [E22b]	(1,5)
8	Country April 2007 Res [E22c]	(1,35)
9	Place of Birth Filter [E23a]	0.0
10	Comuna of Birth [E23b]	(2,4)
11	Country of Birth [E23c]	(1,35)
12	Year of Arrival [E23d]	(41,61)
13	Indigenous Group Filter [E24]	0.0
14	Indigenous Group [E25b]	(3,11)
15	Language [E26]	-
16	Literacy [E27]	1.3
17	Highest Educ. Level Att. [E28]	0.6
18	Level Completed or Not [E29]	0.7
19	Highest Level Passed [E30]	1.1
20	Now Attending [E31]	0.6
21	Internet [E32a]	0.6
22	Email [E32b]	0.6
23	Disability-Motor [E33a]	0.6
24	Disability-Speech [E33b]	0.7
25	Disability-Mental [E33c]	0.7
26	Disability-Hearing [E33d]	0.7
27	Disability-Sight [E33e]	0.7
28	Marital Status [E34]	0.6
29	Relationship Status [E35]	1.3
30	Employment Situation [E36]	0.9
31	Employment Status [E37]	(0.3-1.8)
32	Comuna of Work/Study [E38]	(1,9)
33	Religion [E39]	0.9
34	Children Ever Born [E40]	18
35	Children Surviving [E41]	-
36	Month of Last Birth [E42a]	-
37	Year of Last Birth [E42b]	-

Note MVR = Missing Value Rate. Intervals show possible values depending on true but unknown values of filter variables. “-“ signifies values that could not be calculated for lack of time.

This indicates that missing value rates are important not only as indicators of data quality, but as guides to the development of editing specifications.

The missing value rate intervals in Table 5.1 provide information on the level of uncertainty regarding the missing value rate as well as information on the frequency of missing values.

Employment status, No. 31, has a missing value rate interval of 0.3% to 1.8%. The interval is wide in relative terms, but the upper limit is below 2%. The frequency of missing values is low, whence the uncertainty is relatively unimportant.

Children ever born, number 34 on the list illustrates a high missing value rate with low uncertainty. The question is asked only of women age 15 years old or older, but missing values for age and sex are very low, 0.4% and 0.5%, respectively. These low values account for the low uncertainty. The frequency of missing values, however, is high—18%.

Country of April 2007 residence, No. 8, and country of birth, No. 11, both have missing value rate intervals of 1-35%. The level of uncertainty is so high that it is impossible to say whether the frequency of missing values is high or not.

Year of arrival in Chile, No. 12, illustrates a high frequency of missing values with substantial but far less extreme uncertainty—an interval of 41-61%. Perhaps surprisingly, indigenous group, No. 14, with an interval of 3-11%, shows a high level of uncertainty combined with moderate to high frequency of missing values.

It is notable that the worst cases all involve variables describing persons who were born in another country or who were resident in another country as of April 2007. The contrast with the variables comuna of April 2007 residence (No. 7) and comuna of birth (No. 10) is striking. These variables show less uncertainty and a far lower frequency of missing values, 1-5% and 2-4%, respectively. It may be hypothesized that the high level of uncertainty for the country of April 2007 residence and country of birth questions reflects a reluctance of foreigners resident in Chile whose resident status may be doubtful or problematic to answer these questions.

The overall picture provided by the missing value rate information in Table 5.1 is reasonably good. Missing value rates for 23 of the 37 variables listed are less than 2%. Some of these variables are defined for all persons. Others are defined for subgroups, but have missing value rate intervals so small as to disappear on rounding to tenths of a percent.

The rate for children ever born is so high as to call the usability of the data into question, and similarly for the country of April 2007 residence, country of birth and, to a lesser extent, foreign or dual nationality. These results are regrettable, to be sure, and suggest that focused methodological research would be useful to determine whether results can be improved in the next census, but they represent only five of the 33 person variables for which missing value rate information is provided.

5.13. Analysis of the main stages of data processing for selected areas

Following our field visits we met with the INE data processing staff to request special tabulations of data for the areas we observed. For each enumeration area in the areas we visited we received counts of

- Total housing units
- Occupied housing units, response
- Occupied housing units, no response
- Unoccupied housing units
- Persons

for

RO Data received from the regional office

and for the following data received from the INE data processing staff in Santiago.

D1 Data produced by the process described in sub-section 5.6. At this stage the census data has been processed sufficiently to allow construction of records for persons, households and housing units, but no editing for inconsistencies between variables or missing values for variables has been done.

D2 Data after editing for inconsistencies and missing values. This process will not usually change the number of housing units or the number of persons.

D3 Data after imputation of households and persons to non-responding households. This process converts non-responding housing units to responding housing units, so the numbers of responding units, and the number of persons, will increase if there are any non-responding housing units in the enumeration area.

D4 Data after imputation on unenumerated housing units. This will change the number of housing units and the number of persons whenever a housing unit not observed during field work is imputed to the enumeration area.

As an example, for the enumeration area with portafolio code 2956958, located in Tomé comuna of Concepción, the five counts for the five data sets are as follows.

	RO	D1	D2	D3	D4
1. Total housing units	15	15	15	15	15
2. Occupied housing units, response . .	5	5	5	15	15
3. Occupied housing units, not response	10	10	10	10	-
4. Unoccupied housing units	0	0	0	0	-
5. Persons	-	19	19	63	63

There are 15 total housing units from the regional office, **RO**, and at each stage of central office data processing, **D1** through **D4**. All 15 units are occupied, but 10 of the 15 are non-responding (viviendas particulares personas ausentas). The **RO**, **D1** and **D2** numbers are equal, as expected, because the processing that produces **D1** and **D2** will not usually change the number of housing units.

Between **D2** and **D3**, however, the number of occupied housing units with response jumps from 5 to 10 and the number of occupied units with no response falls correspondingly. Note that total housing units in row 1 of the table should equal the sum of the numbers of units in rows 2-4, providing a check on the counts.

This change is the result of the imputation of households and persons to non-responding housing units. All 10 of the non-responding units have been converted to responding units. The number of imputed units in this case is twice the number of units for which data has been obtained from respondents.

That there is no change in numbers of housing units between **D3** and **D5** shows that no housing units have been imputed in this enumeration area.

The numbers of persons in the last row of the table jump from 19 to 63, this the result of the imputation of persons to non-responding housing units.

Annex 6 contains these counts for 114 comunas in San Pedro de la Paz and Tome comunas in Concepcion.

This data may be used in this way to see how each stage of data processing affects the counts of housing units of each type (occupied responding, occupied non-responding, unoccupied). If anomalies are encountered, the process may be reviewed to see how they may be explained. We observed a very minor number of anomalies.

6. Census products and services

6.1. Introduction

The United Nations *Principles and Recommendation for Population and Housing Censuses* (Revision 2, 2008) states the following.

- (1) A census should cost-effectively meet the demonstrated needs of users.
- (2) User needs are established through consultations with existing and potential users.
- (3) Census products and services should meet legal obligations and users' needs with stated quality standards.
- (4) Census outputs should be produced with a minimum of error suitable for the purposes for which the data are to be used.
- (5) Standard outputs should be provided for main results.
- (6) Services should provide for outputs customized to user needs.

(7) A census is complete when census information is available in a form suited to users' needs.

Points 1-2 are made in paragraph 1.15, points 3-6 in paragraph 1.17, and point 7 in paragraph 1.206.

6.2. Methodological and administrative report

Principle 3 of the *Fundamental Principles of Official Statistics* states that

“To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.”

Chapter III, “Planning, organization and administration of population and housing censuses”, Section J, “Systematic recording and dissemination of census experience”, of the *Principles and Recommendations* states that

It is recommended that every country should prepare and, if possible, publish a *methodological and administrative report* providing specimens of the census questionnaires and forms, instructions for the enumeration, and detailed information on the cost of the census and on the implementation of the census budget, as well as information on the manner in which the census was planned, organized and conducted, the important methodological and other problems encountered at the various stages of the programme, and points to be considered in future censuses. It is important that the report be as comprehensive as possible, covering all stages and aspects of census planning and operations, including fieldwork, processing, analysis, dissemination, evaluation, and so forth. This report would both assist the users of the census results in appraising and interpreting the data and facilitate the proper planning of future data-collection programmes, including population and housing censuses. (paragraph 1.406; emphasis added)

We recommend that INE prepare a methodological and administrative report on the 2012 census, that this report be published concurrently with any future released of 2012 population and housing census results, and that no census results be released in advance of this report.

Information in the methodological and administrative report may be classified under three broad headings.

1. A narrative report of the planning for and implementing of the census, from the very beginning through the first release of census results.

2. Matters of record, such as the census budget, records of user consultations, numbers of enumerators and supervisors recruited, trained and deployed, missing value rates for each variable, imputation rates for each variable, planned products and services with delivery dates for each, and a census dictionary.
3. References to more detailed documentation, such as census questionnaires, enumerator and supervisor training manuals, coding manuals, editing specifications, tabulations plans, reports of field testing, quality assurance reports, and contracts for contracted out work.

The methodological and administrative report may be published in two stages if this facilitates early release of the census results. The first part of the methodological and administrative report, will contain information to enable users to “appraise and interpret” the data. This will include details of field operations and data processing similar to those provided in section 4 and section 5 of this report, and it will also include imputation rates for all census variables. This report must be published concurrently with the first release of (non-preliminary) census results.

The second part of the methodological and administrative report will contain information to “facilitate the proper planning of future data-collection programmes, including population and housing censuses” (quoted phrases in this and the preceding paragraph from paragraph 1.406 of the *Principles and Recommendations*). This report should be published before preparations for the next census begin.

6.3. Standard tables

Tables have traditionally been the primary products of a population census. The United Nations *Principles and Recommendations* devotes over 100 of its 442 pages to recommended and additional tabulations for population censuses and for housing censuses.

The tables that can be produced depend on the topics included on the census questionnaires, a subject to which the *Principles and Recommendations* devotes another more than 100 pages.

Two volumes of 2012 census tables were published, one containing tables for persons, the other containing tables for housing units. The publications are

1. *RESULTADOS XVIII CENSO DE POBLACIÓN 2012, Características demográficas, sociales, culturales y económicas de la población (Tomo I)*
2. *RESULTADOS VII CENSO DE VIVIENDA 2012, Características de las Viviendas y Hogares (Tomo II)*

Review of the tables included in the first of these volumes, in particular the tables on marital status (Table 1.3), children ever born (Table 3.1), and the scarcity of tables for small areas, suggests that user consultations on the development tabulation plan may not have been sufficiently thorough.

It may therefore be appropriate for INE to consult briefly with selected users to determine whether the existing tabulation plan should be revised and to publish standard tables based on a revised plan if the results of the consultations so indicate.

Whatever plan is finally decided upon, standard tables should be disseminated as print-published reports, as digital facsimiles of the print-published reports available on the INE/census website, and as table-specific files in a format easily readable by computer applications users use to work with the data.

At this writing, an intranet page accessible only within INE provides digital facsimiles of the two print-published reports noted above as well as Excel files of the tables contained in the reports, in conformity with the preceding paragraph. Tables 3.1, 3.2 and 3.2 in the first published report do not appear in the list of tables in computer spreadsheet format, however, and the hyperlinks to Tables 9.1 through 9.4 are non-functional.

The *Principles and Recommendations* states that

The basic feature of the census is to generate statistics on small areas and small population groups with no/minimum sampling errors.” (paragraph 1.2(c); emphasis added).

It may be considered cost-ineffective to print-publish detailed tables for comunas and smaller areas, but it is likely that they may be cost-effectively published as computer files on the INE/census website.

6.4. On-demand tabulation service

For all but the smallest countries, the number of tables that may be produced from a population and housing census is effectively infinite. This follows from the census being a complete enumeration of the population, the number of possible geographic subdivisions, the number of variables, the number of possible combinations of variables, the number of possible groupings of values for variables, and the possibility of combining variables for persons, households and housing unit units.

It is therefore unlikely that any collection of standard tables, no matter how extensive and carefully planned, will satisfy all user needs. It is therefore fortunate that the development of information and communication technology in recent decades makes “on-demand tabulation” possible and cost-effective.

On-demand tabulation refers to a product or service that allows users to specify and receive tables they would like to use in real time, subject to restrictions necessary to ensure confidentiality. Computer software applications to implement on-demand tabulation are available commercially and non-commercially.

Web-based on-demand tabulation is one option, but substantial numbers of users may not have Internet connections adequate for efficient use of a web-based service. For these users a CD or DVD one-demand tabulation product is preferable. It is appropriate to provide both a web-based service and a CD/DVD product.

The service or product should provide ready access to all pertinent documentation, including the report on census planning and implementation, questions, enumerator and supervisor training manuals, coding specifications, and codebooks.

On-demand tabulation service should not be confused with products or services that provide access to a large number of pre-made tables. These products and services are valuable for making it possible to cost-effectively provide users with large numbers (tens or hundreds of thousands) of standard tables, but they do not allow users to specify a particular table customized to the user's need.

6.5. Statistical briefs

The products and services discussed in preceding sections tend to satisfy the needs of relatively sophisticated users of census data. Even traditional print-published census tables, with their display of thousands upon thousands of numbers, are likely to overwhelm many potential users of census data who lack the expertise to extract useful information from them.

Additionally, both standard tables and on-demand tabulation are omnibus products, covering most or all topics included in the census. This can make it difficult for users with specific interests to find what they are interested in. This in turn inhibits full use of the information resource represented by the census results.

Statistical briefs are short reports presenting descriptive data on particular topics covered by the census of interest that are of interest to particular groups of persons.

Statistical briefs may be published in print and digitally on the INE website. When published digitally, it may be advantageous for them to be in a format accessible to Internet search engines so that people with interest in the topics can more easily find them.

Producing statistical briefs does not require extensive resources or high statistical sophistication, but it does require a sound understanding of the subject matter and of the larger social, economic and cultural context. To serve their function they must be produced to a high statistical and editorial standard.

It may be appropriate, and indeed advantageous, for a statistical office to produce statistical briefs in cooperation with statistically capable members of the user community interested in the topic of the brief.

Cooperating users benefit by gaining access, if only indirectly, to data and data processing resources not otherwise available to them. The statistical office benefits by strengthening subject matter expertise in the topic and by establishing a potentially valuable relationships with different segments of users. In this way they may provide a cost-effective approach to developing national statistical capacity of producers and users of census and related data.

6.6. Evaluation studies

National statistical offices are often identified primarily as producers of data, but data analysis, particularly analysis aimed at assessing data quality is an equally important role. The United Nations *Handbook of Statistical Organization* identifies analysis as one of the core functions of a national statistical office.

Analysis ... should be carried out by the national statistical office as well as by those outside the Government. It is important to the intellectual vitality of a statistical office that members of its staff critique the process and models used to produce the data as well as discuss the strengths and weaknesses of the data. This is important for the growth of individual staff members and for the office in its quest to improve the quality of its data. ... (paragraph 389)

Analyses of the quality of census data are essential, both for users to know the strengths and limitations of the data and for the statistical office to learn how to improve the quality of the products and services it produces.

Analyses leading to formal written reports, reviewed by competent authorities in or outside the statistics office, will ideally be part of the routine work of the office.

Analysis reports are most likely to benefit users if they are formally published. Publication tends to higher quality reports because knowing that the report will be made public can be a powerful incentive to high quality work. In this way publication benefits the statistics office as well as users.

6.7. Public use microdata sample

A public use microdata sample (PUMS) is a sample of census records for persons, households, and/or housing unit units. The sample records are “anonymized” to protect confidentiality. National statistical offices may produce PUMS for population and housing censuses.

Public use microdata samples enable users to carry out on census data the kinds of statistical analysis applied to sample surveys. A PUMS for the 2012 census may be indicated if user consultations indicate a demand.

7. Post-enumeration survey

7.1. Introduction

The United Nations *Principles and Recommendations for Population and Housing Censuses* states that

It is universally accepted that a population census is not perfect and that errors can and do occur at all stages of the census operation (paragraph 1.380)

and that evaluation of census results should accordingly be undertaken to serve some or all of the following objectives (paragraph 1.382).

1. to provide users with some measures of the quality of census data to help them interpret the results
2. to identify as far as is practicable the types and sources of error in order assist the planning of future censuses;
3. to serve as a basis for constructing a best estimate of census aggregates, such as the total population, or to provide census results adjusted to take into account identified errors

The post-enumeration survey, discussed in Chapter III, Section H, Sub-section 4 (paragraphs 1.393-1.398), of the *Principles and Recommendations*, is one of several methods for census evaluation.

7.2. Definitions

The *target population* of a population and housing census refers to the population of persons (households, housing units, collective quarters) the census is intended to enumerate. The target population is defined by the census reference time, the national territory to be covered (usually but not always the entire national territory), the type of enumeration, specifications concerning such special populations as foreign diplomats, international visitors, and foreign crews of ships in port, and possibly other factors.

Omission refers to persons (households, housing units, collective quarters) in the target population that were not enumerated.

Omission rate refers to omissions as a percent of target population

Incorrect enumerations refers to duplicate enumeration of persons in the target population, enumeration of non-existent persons, and enumeration of persons not in the target population. Similarly, as applicable, for households, housing units, and collective quarters.

Errors in census results are of two types. *Coverage error* refers to omissions and incorrect enumerations. *Content error* refers to “errors that arise in the incorrect reporting or recording of the characteristics of persons, households and housing units enumerated in the census” (*Principles and Recommendations* paragraph 1.380).

7.3. The 2012 census post-census survey (PCS)

Following the 2012 census INE conducted a post-enumeration survey, referred to as the post-census survey (PCS), in urban areas only. INE staff provided information in four different meetings on post-census survey goals, questionnaire, sample design, matching and estimation methodology. The PCS goals were defined as estimating omissions in census and evaluating quality of answers to selected questions asked in census.

Evaluating the quality of responses to questions is also defined as evaluation of content errors.

According to the post census survey (PCS) methodology, an enumeration sample called E-sample is selected from the census enumeration of the target population independent of E sample. A second sample called P-sample is selected from the target population. The PCS methodology requires two independent (P- and E-) samples. Also P-sample data collection should be as close as possible to the census date so that a good quality data is obtained for the target population and hence, reduce matching difficulties. However, the data collection operations for P-sample while closer to the census day should not adversely affect independence between P- and E-samples. Therefore, the data collection operations for the two samples should not overlap in any geographical area beside other factors to maintain independence. The PCS methodology then compares through matching operations, the P-sample records with E-sample records to estimate the true target population, and coverage errors in census.

The PCS methodology estimates the target population. The Dual System Estimation model postulates that each person has a probability of being either included in the census or not included in the census as well as either included or not included in PES. This can be described as in the following table.

Table 7.1. Dual System Estimation Model

	Persons in Census	Persons not in Census	Total persons
Persons in PES	N_{11}	N_{12}	N_{1+}
Persons not in PES	N_{21}	N_{22}	N_{2+}
Total persons	N_{+1}	N_{+2}	N_{++}

In the above table, all the cells are observable except N_{22} and all the marginals that included N_{22} . The model assumes independence between the census and the PCS. Hence the probability of being in ij^{th} cell P_{ij} is the product of two marginal probabilities that contains ij^{th} cell. Thus, under the independence assumption, the estimate of total population is

$$DSE = N_{++} = \frac{(N_{+1})(N_{1+})}{N_{11}}$$

Thus, the total population can be written as function of number included in census, in PCS and those included in both. This model is applied within each estimation domain usually called post-stratum.

In practice, the components of DSE are estimated from samples. The observable entries in the above table are the verified correct enumerations based on matching operations. The matching operation is very complex and high quality matching results are extremely critical for the success of PCS. The matching operations should identify correct enumerations for E-sample persons and match rate for P-sample persons who are correctly enumerated in the PCS as a census day resident. Correct enumeration has four dimensions -- appropriateness, uniqueness, completeness, and geographic correctness. Appropriateness means that the person should be included in the census. Uniqueness means that the person is only once in the census. Completeness means that the person has sufficient information recorded to be identified correctly. Geographic correctness is that the person is enumerated at his/her right location.

7.4. PCS sample design

The INE planned and implemented post census survey (PCS) to measure coverage errors in 2012 census. INE designed the PCS to provide omission rates only for persons living in private housing units in urban areas. Accordingly, a sample of 22,500 housing units was selected to represent the urban area private housing unit population. PCS sample was a two stage stratified systematic sample. INE determined the sample size at regional level. The first stage selected blocks proportional to their number of housing units in 2011 pre-census listings. The second stage selected about 25% of the housing unit in a block. An enumerator received instructions to select a subsample within a block with a given random start.

PCS collected information on housing unit identification, address of the housing unit and the housing unit type. For persons, it collected the first name of a person, relationship to head of the household, sex, age, date of birth, residence location at comuna or higher geographic level, highest level of educational attainment and other related questions, disability status, ethnicity, marital status, usual residence when census was taken, and information on residents of the housing unit. No census day reference was used to collect information. Instead residents were asked their usual residence when the census was taken which was conducted during April 9, 2012 through June 2013. This data was not collected to corresponding to a specific date of census.

7.5. PCS data collection

The enumerators were hired based on their experience with census or survey data collection. They received three- day training along with their supervisors and national coordinators. In addition, the national coordinators and the supervisors received three day training prior to the enumerators' training. The enumerators and other field staff who

also worked on census were not assigned to the same block where they worked during census.

The enumerator materials included PCS forms, and a map of the sampled block. The map of the block did not include any listing of the housing units. Enumerator first canvassed private housing units in the block starting from the northwest corner of the block. Then starting with the pre-specified random start, selected the sample housing units for interview.

The supervisor selected a 10 percent sample of the work completed previous day to check the quality of enumerators' work. The supervisor also checked the entire work of an enumerator if he/she suspected poor quality of his/her enumerator.

Based on the above discussions, we concluded that the sample selection, and field operations including enumerator assignment and their materials did not violate required independence between the P- and E- samples.

Census data collection occurred during April through July 2012. The PCS data collection occurred during November through December 2012. This delay in PCS data collection according to the PES team members was due to late decision to conduct PCS.

7.6. PCS matching methodology

The matching methodology has not been fully developed and implemented yet. However, PCS team is currently working on it. The Team has formed three geographical groups for matching and estimation. The first and second groups includes six regions each and the third group includes the remaining three regions. The first group includes smaller region and the third group includes the larger regions. The matching and estimation for the second group will start after completing the first group, and the third group will be proceeded after the second group. According to the schedule, the first group will be done by December 2013, second group by March 14, and the third group by June 2014. Thus, matching would be about 18 to 24 months after 2012 census was conducted.

Matching methodology would first match based on the responses to questions to geography, first name, sex, and age of a person. If needed, secondary auxiliary items consisting of marital status, ethnicity, and educational level would be used to resolve match status of a person. Both census and the PCS collected only the first names of persons residing at a sampled housing unit. Without the full name, it would be difficult, if not impossible, to determine correct enumeration status of significant number of sample persons. The matching operations do not include field follow-up for those persons in P-sample that did not match with E- sample persons and vice versa.

Of the occupied housing units, 9.2% were absentee housing units in PCS. The census had about 3.03% absentee housing units. This means that based on absentee housing units alone, at least 9.2% and as large as 12.2% of persons in two samples would not match, hence, may not be identified if they were correctly enumerated. In addition, many persons who were interviewed would not be identified as correct enumeration due to lack of their

full names for matching. Lack of full names of sample persons would make almost impossible to identify duplicates within block.

Furthermore, a significant number of persons would not match even if they were enumerated only in one sample while their housing unit was enumerated in both samples. This would happen when a person living in the housing unit was not recorded in one of the two samples while his/her housing unit was enumerated in both samples or the occupants who were living in a sample housing unit at the time of census have moved out and were replaced by in-movers. The mover rates in urban areas was not available. Hence, it is not possible to quantify percent of movers in the sample who would have moved between the census and PCS interview period.

7.7. PCS field follow-up operation not undertaken

The PCS did not include a field follow-up operation. This has implications for estimating correct enumeration and, hence, omissions in census. According to the current plan, if person is enumerated only in one of the two samples, the person is considered correct enumeration in that sample. Thus, if the person was enumerated in P-sample only, then it is considered omission in census and correct enumeration in P-sample even if it was not confirmed as correct enumeration in P -sample. This person could have been included erroneously in P-sample. Considering the above information, it would not be possible to determine correct enumeration for a large proportion of persons in one or both samples. This proportion just based on absentee housing unit could be as large as 12.2. This percent would be higher due to movers in the sample between census and PCS interviews.

Furthermore, modifying the matching operations to include field follow up of those persons who did not match during the current planned matching phase will not have significant pay off. This is due to the fact that the data collection for census and PCS had been over a year and many persons have moved away from the housing unit they were occupying during the census or PCS enumeration. Hence, no additional data would be obtained for these movers during the field follow-up. Second due to the long time lag, a good quality data that would help matching most likely would not be available due to recall bias. Third, since only the first name is recorded in both samples, the field follow-up would not be as successful if the full names were recorded. Therefore, efforts to get additional information during the field follow-up operations that would help matching and identify correct enumeration would not be very productive.

7.8. PCS estimation methodology

INE developed the PCS estimation methodology which is not statistically defensible. INE defines percent omission rate as

$$\text{Percent omission rate} = (100 - \text{coverage rate}),$$

Where percent coverage rate is defined as

$$\frac{\text{No. of persons in both samples} + \text{No. of persons only in PCS}}{\text{No. of persons in both samples} + \text{No. of persons in only in E - sample}} * 100$$

This formula includes erroneous enumeration in both samples, hence, it does not provide a statistically defensible estimate of omissions.

More appropriate rate of omission is the percent difference between the estimates of target population and correct enumeration in the target population. That is ***Census Omissions = True Population – Correct Enumerations***

This computation of omission rate is described in United Nation *Post Enumeration Surveys – Operational Guidelines*, page 62.

Omission rate obtained this way is statistically defensible and many countries have computed omission rates by applying this methodology.

7.9. Conclusion and recommendation

The preceding discussion shows that the PCS cannot be used to estimate omission in the 2012 census. We recommend accordingly that any ongoing work aimed at this objective cease.

If resources permit, however, INE might use the PCS data for methodological research to enhance future censuses and PCS. This research might focus on, for example, the following.

1. Content errors for matched persons in P- and E-samples
2. Missing value rates for one, two or three match variables
3. Match rates based on the current methodology
4. Sources for causing not match
5. What worked/did not work in matching methodology

Knowledge gained in this way could improve future censuses, post-enumeration surveys, and other data collection operations.

8. Demographic equation estimates

8.1. Introduction

The United Nations *Principles and Recommendations for Population and Housing Censuses* states that

“Demographic analysis offers a powerful methodology for evaluating the quality of a census and countries are encouraged to use demographic analysis as part of their overall census evaluation methodology.” (paragraph 1.389)

This section discusses the use of demographic equation estimates for evaluating completeness of enumeration in the 2012 census.

8.2. Demographic equation estimation

The demographic equation may be expressed most simply as

$$(1) \quad \textit{Ending Population} = \\ \textit{Initial population} + \textit{Entries to the population} - \textit{Exits from the population}$$

where *Initial Population* and *Ending Population* refer to the numbers of entities in the population at the beginning and end of some time period and *Entries ...* and *Exits ...* refer, respectively, to the numbers of entities entering the population during this time period and the number of entities exiting the population during this time period.

When the entities comprising the population are persons—they may also be, for example, households or housing units—, *Entries* and *Exits* may be expressed as

$$(2) \quad \textit{Entries to the population} = \textit{Births} + \textit{In-migrants} \textit{ and} \\ (3) \quad \textit{Exits from the population} = \textit{Deaths} + \textit{Out-migrants}.$$

Formulas (1-3) are tautologies, meaning that their truth follows from the definitions of their terms.

Estimates of population numbers based on these and similar equations are referred to as *demographic equation estimates*. Formulas (1-3) are patently elementary, but they are valid only if their terms refer to the same clearly defined population. This requirement, simple in principle, frequently complicates calculation of demographic equation estimates.

If *Initial Population* is the *de facto* population of some geographic territory, for example, *Ending Population* must be the *de facto* population of the same territory, *Births* and *Deaths* must refer to events occurring within this territory, and *In-migrations* and *Out-migrations* must refer to physical movements across the boundary of this territory.

If *Initial Population* and *Ending Population* are based on usual residence, the definition of usual residence must be the same. It will also be necessary for vital statistics numbers of births and deaths to be available by usual residence, of the mother in case of births and of the deceased person in case of deaths, and the definition of usual residence used by the vital statistics system must be the same as the definition used by the population censuses or other source of the initial and ending population numbers.

Demographic equation estimates for national populations typically involve four main ingredients.

1. An *Initial Population* from the most recent population census, perhaps adjusted for net omission, referring to time *t1*.
2. *Births* between time *t1* and time *t2* estimated from vital statistics, perhaps adjusted for incomplete birth registration or other errors, where *t2* is the time for which the estimated of population is wanted.

3. *Deaths* between time $t1$ and time $t2$ estimated from vital statistics, perhaps adjusted for incomplete death registration or other errors.

4. *Net in-migrants*, defined as *In-migrants* – *Out-migrants*, between time $t1$ and time $t2$, estimated from whatever data is available on international migration.

The demographic estimate is

$$(4) \quad \text{Estimated population at time } t1 + \\ \text{Estimated Births} - \text{Estimated Deaths} + \text{Estimated net In-migrants}$$

where the last three terms refer to events between time $t1$ and time $t2$.

If the population is closed to migration, the last term of (4) drops out. For many countries, net in-migration, though not zero, is small relative to the difference between births and deaths. For these countries, demographic equation estimates of the national population may be robust against errors in the estimated number of net in-migrants.

Formula (4) applies to subnational as well as national populations. Application requires estimates of net in-migrants, however, and for subnational populations net in-migrants may be large relative to the difference between births and deaths. Accurate demographic equation estimates therefore require accurate estimates of net in-migrants. Data on internal migration (migration within the national territory) is for many countries unavailable or problematic.

8.3. Demographic estimate of the 2012 census population

A demographic equation estimate of the 2012 census population may be calculated from the 2002 population count, estimates of inter-censal births and deaths calculated from registered births and deaths, and an estimate of net inter-censal net migration. Various complications must be dealt with, but these are familiar to demographers and do not require further mention here.

The Chilean demographers we have met believe that the vital statistics system in Chile registers nearly all births and deaths. This may be true, but the United Nations *Principles and Recommendations for Vital Statistics Systems* states the following.

As far as practicable, qualitative or quantitative indicators of the degree of completeness and timeliness of registration should be given for each geographic reporting area (see Chap. IV) and also, where pertinent, for various significant segments of the population, (e.g., various ethnic groups). (paragraph 471)

Evaluation or performance monitoring should be part of the operation of civil registration and vital statistics systems. If this function is not already a component of civil registration and vital statistics systems,

there should be a designation of an evaluation unit within each system, as a component part of an improvement strategy. This unit should be responsible for organizing evaluation studies, using the appropriate methodologies described in chapter V below, as well as external evaluation, internal evaluation, pilot studies and demonstration area projects. (paragraph 526)

The evaluation of quality of data should address, *as a minimum, the level of completeness of civil registration ...* (paragraph 548; emphasis added)

Because of the importance of civil registration information and register-based vital statistics, both on an individual record basis as well as in aggregated form, the maintenance of high standards of quality should be a major and continuing concern to those responsible for the administration of the systems. Therefore, adequately funded evaluation activities of civil registration and vital statistics systems must be considered as essential components of the management, operation and maintenance of such systems. (paragraph 549)

The responsibility for the establishment and execution of methods of critical evaluation should be vested in an independent agency capable of undertaking evaluation. ... (paragraph 550)

So far as we have been able to determine, no evaluation study providing an estimate of the level of completeness of birth and death data between 2002 and 2012 is available.

The statements we have heard attesting to the completeness of birth and death registration in Chile are in fact hypotheses about the behavior of millions of Chilean men and women. To be in accord with international recommendations, the vital statistics system must provide statistical validation of these hypotheses.

Demographic equation estimates of the total population of Chile as of the 2012 census may be made using vital registration numbers of births and deaths (it appears that international migration numbers may be too small to have an appreciable effect). Lacking the requisite evaluation of data quality, however, these estimates do not, based on international recommendations, provide a statistically sound basis for estimating 2012 census omission.

8.4. Demographic estimates of the 2012 subnational population

Demographic equation estimates may be made for subnational populations if numbers of births, deaths and net migrants for *de facto* (for consistency with the 2002 census *de facto* enumeration) subnational populations are available.

The resulting estimates will of course be of the *de facto* population, and so not strictly comparable with the 2012 census counts of usual residents. Analyses to establish the relationship between *de facto* and usual resident population in subnational areas would be required to know how large the differences are. Differences might be statistically

negligible for some large comunas in the Santiago region, for example, but large for relatively small comuna in relatively remote areas of the country.

All of the issues raised in the preceding subsection for national population estimates apply as well to subnational population estimates. The critical issue for subnational populations is data on internal migration. To estimate the population of a subnational area it is necessary to have accurate estimates of the net number movements between subnational populations during the intercensal period.

So far as we have been able to determine, there is no source of migration data adequate for this purpose. It is possible that 2012 census data on internal migration might be used, but so far as we have been able to determine, this has not been done. A potential problem is that missing values on the relevant migration variables, though perhaps few in number relative to the total population, may be large in relation to numbers of migrants into and out of subnational areas, particularly smaller areas.

Given this lack of internal migration data, it is not possible to produce demographic equation estimates of subnational population as of the 2012 census of sufficient accuracy estimate 2012 census omission at the subnational level.

8.5 Demographic equation estimates and projection estimates

Population projection estimates and demographic equation estimates are related, but different. There is a potential for confusion because the same demographic equations may be used for both.

Population projection estimates are typically produced by the national statistical office or another responsible entity following the release of population census results. Projection estimates for every country in the world, as well as population estimates back to 1950, are produced every two years by the United Nations Population Division. Results of the most recent projections are available at esa.un.org/unpd/wpp/Documentation/publications.htm (visited 18-Nov-2013), the Population Division website.

The inputs to population projections are typically population numbers from the most recent census, perhaps adjusted, and anticipated future numbers of births, deaths, and migrations, anticipated directly as numbers of events, or indirectly as rates to be applied to population numbers, or some combination of the two.

Population projection estimates are not demographic equation estimates because the future numbers of births, deaths and migrations are mostly or entirely unknown at the time the projection is made. Projection estimates are empirically grounded, and great efforts may be made to make them the best possible estimates of future population, but they make no pretense of being more than plausible guesses about future numbers of births, deaths and migrations.

When a future population census produces numbers of persons different from those in projection estimates based on the previous census, the presumption is that the census numbers are superior to the projection estimates, because the projection estimates are

anticipations of an imperfectly known future, while the census results are the result of a recent data collection operation.

9. Sampling of non-responding housing units

9.1 Sub-population sampling of non-responding housing units

During the 2012 census 173,787 occupied housing units were not interviewed due to either unavailability of the residents or they refused to be enumerated. For these residents, INE proceeded with imputation procedures to account for the persons living in these housing units. Since the imputation procedures assumes that the non-interviewed residents are similar to those who were interviewed which is not necessarily true. Therefore INE considered to a sampling based approach to possibly enhance their approach to account for these non-respondents. The samples based approach is briefly discussed below.

9.2 Sample Design

173,787 housing units were classified as “non-responding housing units” in 2012 census in 346 comunas covering the entire country. 16 comunas did not have any non-responding housing units. 162 comunas had 45 or less housing units. 168 the remaining comunas had more than 45 housing units. The sample was selected from these 168 housing units using region, urban/rural, comuna and the district as stratifiers. The sample to a comuna was allocated proportional to number of housing units in a comuna. According to INE, 75,701 absentee housing units were selected from 168 comunas to estimate a proportion of 0.5 with margin of error 5.0% with 95% confidence level.

INE expects response rate of 80%. Also, INE estimated implementation time of 4 months and 15 days which includes administrative arrangements, training of staff, field work including regional and national level supervision, and data preparation for release. The estimated cost for this approach is about USD 12 million.

The commission reviewed the information provided on methodology. The methodology in theory is acceptable as long as census legislation allows the use of sampling to produce the total population of the nation. Our concern with the plan is succeeding in achieving the margin of error for most of the groups and operational aspects of the plan and are briefly discussed below.

9.3 Margin of Error Aspect

The proportion used here is to estimate the proportions of persons belonging to a group. For example, proportion of persons 60 years of age or older. Another example is proportion of married persons over 15 years of age. The 0.5 proportion is used to compute sample size since it gives the largest sample size. In practice, most of the characteristics of interest would have proportion less than 0.5. Hence the sample size would provide reliable estimates for all proportions.

Alternatively, one could design a sample to get an average household size of persons living in absentee housing units to get a better estimate of total population.

9.4 Operational Aspects

The percent of refusals is unknown among the non-responding housing units. Those who refused to be interviewed during the census would most likely refuse again.

INE could not interview people in non-responding housing units during the census over about three months in spite of attempts made by supervisors and/or special teams. It is not clear what would encourage respondents to cooperate now. The INE assumed 80% response rate of the sample. However, INE herself raised concerns if they would succeed in getting such high response rate.

Over more than 18 months after census, the economic conditions of residents may have changed. Hence, the sample would estimate characteristics of different time frame and not of the census time.

Annual mover rate is not available at this time. Thus, it is not clear how many of the current residents did not live in the non-responding housing units at the time of census. One can assume that the persons who moved in have similar characteristics as the residents of census day but this assumption is questionable. This is because it is possible that due to changes in various local conditions (due to economic, earthquake and tsunami effects), the non-respondents may have different characteristics than at the time of census.

The sample estimates are subject to variances. Thus, the released data should have ranges on estimates. This would complicate release and understanding of the census data for data users.

INE selected sample in only 168 comunas. INE should explain satisfactorily to the remaining comunas why they were excluded from their plan.

The commission does not feel that INE would succeed using this approach based on the above factors. Therefore, INE should not implement this approach. However, if INE decide to implement sampling approach to account for non-respondents, they should develop and test their approach to improve significantly the response rate to ensure it is worth pursuing. The imputation approach that INE has already developed and implemented would be less risky. Also, imputation is accepted approach and had been implemented by many countries.

Alternatively, INE could collect the intended information using a national demographic survey that is already in place. For example, collect the information on residents of housing units who were interviewed during the first attempt, second attempt, third attempt, fourth and later attempts by the number of attempts. Interviews completed during the fourth or later attempts would be a proxy for those who did not respond after three attempts in census. Therefore, obtain the information from those who were interviewed during the fourth or later attempts to get the household size, sex and other information to improve census. This approach would require modification to the current interviewing procedures such as making four or more attempts to interview a housing

unit, recording and capturing additional information but would not be a major change. Also, the existing survey questionnaire does not have to be changed. Instead create a form to record limited information needed to improve the census.

Even if the alternative approach is not used for 2012 census, it would be important information for designing future INE programs such as censuses and surveys.

10. Chile's next population and housing census

10.1. Introduction

This section presents recommendations and suggestions for the next population and housing census in Chile. Many of the issues discussed have been touched in preceding sections.

10.2. Planning

The United Nations *Principles and Observations* states that

A population and housing census (or a population census by itself) is perhaps the single most extensive, complicated and expensive statistical operation, consisting of a complex series of interrelated steps that a country undertakes. (paragraph 1.77)

and further that

To ensure that the diverse operations occur in their proper sequence and in a timely manner, the entire census and its various component steps must be planned for carefully in advance. An apparently minor oversight in planning may lead to serious defects in the census results and to costly inefficiencies in the census operations. Careful planning is therefore critically important to a successful census, not only in countries with comparatively little statistical experience but also in those with a well-developed system of statistics. (paragraph 1.78)

The demands of census planning have been compounded in recent years by development of new approaches to census taking involving the use of administrative data and advances in information and communication technology. New methodologies are discussed in Section E, "Methodological approaches" of the *Principles and Recommendations*. Given the speed of technological change, however, other information sources, such as papers presented at international and regional meetings, may be an equally important source of information.

New approaches may appear to provide advantages for improving census coverage, reducing respondent burden, early delivery of census products and services, and lower costs. It is however essential that the decision to move to a new methodology be carefully researched, analysed and tested. If the necessary conditions—such as for example a

housing unit register that is well maintained and regularly updated, and whose quality is assured by thorough statistical evaluation—are not in place, the advantages will not be achieved.

The magnitude of the task of planning a population and housing census may be suggested by listing briefly some of the areas that must be covered. This is done in the following paragraphs.

10.2.1. Project planning A population and housing census is an extremely complex endeavor involving the coordinated work of thousands of people over many years. Meticulous project planning is therefore critical to success. Major milestones and their completion times must be identified. It is important to involving the staff that will be doing the work in each area.

10.2.2. Legal Basis Consider the legal mandate and any legal requirements specified in the legislation to conduct the census. This may also identify the type of census and the year of census to be taken. Also, consider and propose changes if needed for the future census.

10.2.3. Administrative organization Forming an administrative organization with clearly described responsibilities is essential for successful census. The organizations must be adequately staffed.

10.2.4. Mapping and demarcation of enumeration areas Census maps covering the entire national territory must be produced, even if there is no population in certain part of the country. The national territory must be divided into enumeration areas and a map produced for each area.

10.2.5. Quality assurance and improvement programme Quality assurance measures must be developed for all processes, including data collection, data capture, data processing, and design and production of census products and services. The United Nations *Handbook on Census Management for Population and Housing Censuses* states that

Quality is therefore not just the outcome of mechanistic applications of predetermined measures but relies on a combination of:

- (a) Established, documented processes;
- (b) Systems to monitor the outcomes of these processes;
- (c) Active encouragement by management to involve staff undertaking the processes in identifying and resolving deficiencies with quality. (Introduction to Section C, “Quality Assurance”)

10.2.6. Census budget Good budget estimates are very important for success of a census. Good estimates reduce major changes to census plans while the census is underway. Preparing the budget involves listing all work that must be done and determining starting dates, milestones and completion dates, and identifying staffing and other resource needs.

10.2.7. Methodological research Methodological research is required to respond to changing conditions and to improve future censuses. Research areas must be identified and a research program, possible involving data collection, developed. A reinterview program, for example, may be inexpensive way of evaluating content error. The results may be used to improve questionnaire design.

It may be useful to explore the use of analysts and researchers in other government agencies, universities and other research organizations, both to supplement necessary resources and to provide different perspectives. Interagency agreements could be made to ensure confidentiality of the data. A further benefit may be the development of working relationships with user.

10.2.8. User consultations Consultation with other government ministries and key stakeholders helps to ensure the collection of appropriate and useful data. Start this process from ground zero. Each organization requesting data should clearly identify the reason for collecting it, how it was used in the past and how it would be used in the future, any policy usage of the data including the specific policy should also be clearly identified. Also, the consequences should be explained if the requested data are not collected.

It may be difficult to resolve disagreement in a specific situation. Therefore, also identify the sole authority who resolves disagreements and makes the final decision on data to be collected.

10.2.9. Designing census products and services “The census should produce statistics that are relevant to data users. This is the core statistical objective of undertaking a census. Every action in a census must be directed towards producing relevant output that meets the needs of users.” (opening paragraph of United Nations Census Management Handbook, page 1).

Develop data publication plans in consultation with other government ministries and key stakeholders. This would include the data release schedule, published tables, electronic files, and accompanying documentation.

10.2.10. Questionnaire development Once the need data is identified and agreed upon, the census questionnaire must be developed. Well-designed and tested questionnaires are critical to the success of the census, as is determining the most appropriate data collection method. Designing questionnaire should include testing of a questionnaire at a small scale (also called hot house testing) before conducting a large scale experiment. If more than one census questionnaire is designed and tested, develop an evaluation and decision plan. If a revision to a questionnaire is needed after the test, revise and test again to see if the changes accomplished what was intended.

Other forms to monitor census process are crucial for the success and understanding of the census operations. Therefore, develop and test other forms needed to monitor selected census operations. The purpose of each form should be clearly stated.

10.2.11. Enumerator and supervisor training materials Develop manuals for field operations, including manuals for hiring, training, supervision, office operations, quality assurance checks, and so on.

10.2.12. Recruiting, selecting, training, supervising and compensating field staff Develop recruiting and administrative plans for field operations. Develop training plans and train the field staff. Include practice exercise as a part of the training. The training should include all level of staff starting from enumerator to the director of the regional office.

10.2.13. Data collection Develop data collection plans including schedules, contingencies for unexpected situations, monitoring plan for daily work progress, and quality check plans. Data collection period should be planned only for a week or so. However, in certain situations such as difficult to enumerate areas, it could be extended for extra days.

10.2.14. Data capture and processing Prepare a detailed plan describing the data capture and processing including various steps of the process.

10.2.15. Comprehensive test of all census procedures Test the entire census operations one year before the census. Every census process should be tested, beginning with field operations, through data capture and processing, production of census products and services, and evaluation and analysis. The test is carried out in selected areas. It is like a mini census—it is sometimes called “dress rehearsal”—conducted in only a few localities. Localities with specific characteristics, such as urban, difficult to enumerate areas, to assess whether census procedures give acceptable levels of non-response.

10.2.16. Census evaluation Census evaluation plans should include both internal and external evaluation. The post-enumeration survey is one recommended method. Demographic estimates are another method, but for these it is essential to have estimates of completeness of registration of births and deaths, as detailed in the United Nations *Principles and Recommendations for a Vital Statistics System*. INE may consider initiating studies to estimate completeness of births and death registrations data in cooperation with the other organizations involved. The census budget should include a census evaluation component.

10.2.17. Continuity with previous censuses Many uses of population and housing census data involve comparisons with previous censuses. Such comparisons are complicated by changes in the topics included, the classifications used, and the type of enumeration.

10.2.18. Comparability with censuses of other countries Population and housing censuses are international as well as national resources and may be used for comparing conditions in different countries. International comparisons are facilitated by adhering to international standards and recommendations. National needs take priority, but where there is no conflict following international recommendations is appropriate. Comparability with other countries in the same region of the world is important, but comparability considerations should not be limited to one region.

10.2.19. Conformity to international recommendations The United Nations *Principles and Recommendations for Population and Housing Censuses* is a valuable resource for international

comparability. The recommendations adopted by a country will depend on many circumstances, and some recommendations may not be followed, but they should be considered.

10.3. Type of enumeration

Both *de facto* and *de jure* censuses have been conducted successfully around the world. Either census can be successfully conducted if it is well and efficiently designed, its implementation is of high quality, and quality process is implemented in almost all steps of census process. The following factors should be considered in deciding the type of enumeration.

10.3.1. Comparability with previous census The census results are often compared with the previous census for various reasons. For example, policy makers are interested in knowing the socio-economic changes that have taken place at the national and sub national levels. Two different type of censuses cause complications in analyzing two difference census results while in some cases it does not permit a valid analysis. For example, subnational level analysis will not be possible since they represent different populations for a sub-national location.

10.3.2. Comparability with Other Countries Will there be any possibility in future to compare census results with other countries? The comparison with other countries would shed light on how the countries are progressing in comparison to each other. Comparison could be of migration from rural to urban, poverty level, economic progress, growth of population, etc. Comparison of census results based on similar type of censuses would be simpler and require fewer assumptions. Therefore, it is worth conducting the same type of census as the other countries. Annex 7 shows the type of enumeration in the most recent and previous census for 20 Latin American and Caribbean countries.

We recommend that INE decide the type of enumeration, *de facto* or *de jure*, for the next census based on an analysis of what will best serve user's needs as determined by consultations with users.

10.4. Reference time

Planning for the next population and housing census should identify a census reference time and plan census operations to obtain information on persons, households and housing units as of this reference time.

10.5. Length of enumeration period

We recommend that, if the next population census is not a one-day census, the length of the enumeration period should be about 2 weeks.

10.6. Census date

The decision on when the next population and housing census will be conducted involves a wide range of factors particular to the country in which it is conducted. It would not be

appropriate in our view for the International Commission to recommend that the next census be taken in a particular year.

We recommend rather that the decision be made by giving due consideration to the time required for planning and preparatory work indicated in sub-section 8.2 above and in more detail in Chapter III, “Planning, organization and administration of population and housing censuses”, of the United Nations *Principles and Recommendations for Population and Housing Censuses*.

Specific international recommendations on how much time is required to prepare for a population and housing census are not available, in part because of wide differences national circumstances. International practice is unreliable as a guide because censuses may be taken with inadequate lead time due to delays in allocation of resources.

Our expectation is that a minimum of five years will be required for preparations for the next population and housing census in Chile. Given the problems, actual and perceived, of the 2012 population and housing census, we recommend that a full census test be conducted one year prior to the census, as recommended in the United Nations *Principles and Recommendations for Population and Housing Censuses*.

11. Conclusions and recommendations

The International Commission scrutinized field operations and data processing for the 2012 population and housing census intensively and in detail. Our general approach, reference points and implementation details are described in preceding sections of this report.

We received information primarily from meetings at INE in Santiago and during field visits to regional offices and enumeration areas in the Santiago Metropolitana, Antofagasta and Bio Bio Regions. We investigated all important procedural aspects of the census enumeration and the data processing that followed, as well as selected preparatory work leading up to the enumeration.

What we learned about the process of conducting the census, our assessment of the work, and our reference point’s international principles, recommendations and guidelines is reported in detail in Section 4 on field operations and Section 5.

We observed that the rate of missing values for most of the questionnaire variables is low. The non-response rate of occupied housing units and the percentage of unoccupied housing units is not far from the same rate experienced by other countries. Countries with similar non-response rates have used their census data.

The report on the data capture prepared by INE indicates a data capture error rates of 0.14% for housing unit data and 0.29% for person data. These rates indicate good quality of data capture.

During our investigation we found several problems in the census process such as the imputation of housing units not observed during the enumeration, the length of the

enumeration period and no census reference time, and have commented on them in preceding sections of this report.

We are also aware that positive evidence of the quality of census results is important, but lacking. The design and implementation of the Post-Census Survey (PCS), as implemented, precludes using it to estimate omission of persons that should have been enumerated in the census.

Demographic analysis to establish omission at the national level is problematic because the evaluation of the completeness of birth and death registration stipulated by the United Nations *Principles and Recommendations for a Vital Statistics System* has not, so far as we have been able to determine, been carried out. There is no empirical basis for estimating census omission at the subnational level. Indeed, this would require inter-communal migration data for the inter-censal period, which is not available.

However, the outcomes of the investigation shows that the census was conducted within the level of standards and practices of other similar census operations conducted in other countries.

On the basis of these findings we conclude that the 2012 population census results will provide information for describing and assessing Chile's "economic, social and demographic circumstances and for developing sound policies and programs aimed at fostering the welfare of a country and its population" (United Nations *Principles and Recommendations*, paragraph 1.23). They will be useful as well for producing population projection estimates, for research organizations, commercial enterprises and civil society organizations of many kinds.

The 2012 housing census results provide information that may be used to assess the quality and characteristics of housing units throughout the country, to formulate housing policy and programs, and for disaster preparedness planning. They will be useful as well for research organizations, commercial enterprises and civil society organizations.

We accordingly recommend that the 2012 population and housing census data be disseminated to best serve users' needs such as social policies, housing programs, and commercial enterprises as soon as possible.

It is essential, however, that the census results not be disseminated without detailed documentation that will enable users to assess the fitness of the data for the uses they intend to make of it. The documentation should include field operations, data processing, tabulation and any other information regarding data quality.

Short-term recommendations:

- 1) The Commission recommends not to use census data derived by imputed unobserved housing units.
- 2) Imputation of households and persons in non-responding housing units is practiced in some countries and not in others. For instance, United States, Canada, and United Kingdom use imputation in such situations. However, there is no

international standard. We recommend that INE decide its approach based on what will best serve its users' needs. In case INE decides to use imputation it would not be out of line.

- 3) In any case, to account for population in non-responding housing units, we do not recommend taking a new sample of these units. This would be very expensive and most likely would be unproductive.
- 4) Dissemination of census results, with full documentation of census methodology and with appropriate cautions to users on data quality concerns, will promote analysis of the quality of data on specific topics. We recommend that the 2012 census data be disseminated and used, as soon as the necessary methodological documentation is ready for publication.

Medium and long term recommendations:

- 5) Data quality of the 2012 census may be investigated over a large period of time by analyzing the census results in relation to preceding censuses and other data sources.
- 6) While planning the next population and housing census, it is recommended to devote adequate time to gain knowledge of different type of censuses conducted in other countries, in relation to the use of administrative data, and in terms of type of data collection, to improve data quality, reduce costs and respondent's burden.
- 7) It is recommended to spend sufficient time for planning and developing a quality assurance program for the next population and housing census, to document and evaluate all the census phases and census results. This will improve data quality and will reinforce capacities and professionalism of INE.
- 8) The Commission recommends to plan and develop for the next census a digital mapping program in line with international recommendations and best practices experienced in other countries, with use of GIS tools and remote sensing data in order to support the improvement of the coverage and evaluation. This process will require several years.
- 9) Given the problems, real and perceived, of the 2012 census, we recommend that a full pilot census be conducted before the next population and housing census, as suggested in the United Nations *Principles and Recommendations for Population and Housing Censuses*. In view of this, and considering the preparatory work required for a successful census detailed in section 9 of this report, we recommend that a minimum of five years be allowed for planning and implementing the next census. Given that this is written in mid-November 2013, we recommend that the next census not be taken before 2019, and that it will be conducted the same year as per the majority of censuses carried out in the region for data comparability.
- 10) In the course of our work we found insufficient attention to user consultations to establish user needs to guide decisions on census methodology and content. The

United Nations *Handbook on Census Management for Population and Housing Censuses* makes the point succinctly in its opening paragraph: “The census should produce statistics that are relevant to data users. This is the core statistical objective of undertaking a census. Every action in a census must be directed towards producing relevant output that meets the needs of users.” We recommend that INE develop and implement a rigorous, systematic and continuing process of user consultations to learn in detail their needs and capabilities and to use this information to design census and other statistical products and services that will best serve all current and potential new users.

Further observations, suggestions and recommendations on other aspects of the 2012 population and housing census and the next population and housing census are presented in the preceding sections of this report.