The New Zealand Census:

Some Technical and Historical Aspects

A thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Statistics at Massey University

Shirley Ann Dixon (nee Coombes)

March 1989

Abstract

This thesis provides an overview of the New Zealand Census of Population and Dwellings. Certain critical aspects are examined in detail, including the collection phase involving questionnaire content and the enumeration process, the testing before and after, the preparation of the data for entry into a computer and the subsequent dissemination of the information. The information for this research was obtained from published material from overseas, from published and unpublished material from the New Zealand Department of Statistics and from interviews with some officers of the Department.

In each aspect, New Zealand is compared and contrasted with other major countries; specifically America, Australia and India. Because of its geographical proximity, any developments in Australia have an immediate impact on New Zealand. The US Bureau of the Census is often a forerunner in the development of census procedures and techniques. The procedures developed in India to cope with their own specific and peculiar problems in census-taking provide an interesting comparison with those of New Zealand. Where pertinent, aspects of censuses in other countries are also compared with those of New Zealand censuses.

New Zealand has adopted many of the procedures used in other countries, but limited resources have hindered or prevented census staff from developing and maintaining some of the procedures used in American and Canadian censuses. In particular, pilot testing of questionnaires has only recently been incorporated into the census procedures, and major post-censal evaluations are not conducted. On the other hand, the small size of the New Zealand population has facilitated innovations in such areas as data entry, editing and imputation.

The history of census-taking is covered to gain a perspective on the place of the census in modern society. Alternatives to

censuses were examined; specifically, regular major surveys, administrative records and data banks. It is found that surveys suffer a lower response rate than censuses and that the problems of differential undercoverage of various population groups experienced in censuses are exacerbated in surveys. Administrative records frequently do not contain sufficient detail, varying definitions are employed to categorise the data and the quality of the data cannot always be assured. Data banks provide a rapidly growing source of information, but currently also suffer from a lack of universal definitions, and many data banks do not incorporate strict quality control procedures as a matter of course. Moreover, strict confidentiality laws currently prevent access by census staff to administrative files and data banks.

It could be argued that censuses should continue to be taken because of the need to obtain current, detailed information on all members of any population for planning for present and future needs of that society. A census is the only vehicle for collecting information supplied by all members of the population at a single point in time.

If censuses are to remain credible and acceptable to the individual members of a population, challenges must continue to be addressed such as: the accuracy of estimates must be protected by obtained the highest possible response rate from all sections of the population; confidentiality of data must be guaranteed; the costs of the census operation must be kept within budget, while still maintaining high data quality and publication of data in a time frame that is acceptable to users of census data; universal definitions must be employed to minimise the redundancy between censuses, surveys and administrative lists; results of the census must be attractively presented to the public using a variety of media and accompanying analysis reports must be aimed at increasing the public awareness and of the importance and need for regular, successful censuses.

Preface

As nations grow, their composition and needs change. Young nations are typically small, with most of the population concentrated in rural areas and employed in agricultural activities. Usually, as the size of the total population increases, more and more people are attracted towards the towns and cities. This is known as "urban drift", and a trend towards non-agricultural employment marks significant social and economic changes in the structure of the population. Some towns and cities will grow at a faster rate than others, and will place varying demands on the social and financial resources of the country.

In addition to the geographic distribution of the population, the racial and the age distribution will also determine current and future needs. Different races have different social structures and these will affect their requirements for housing and education. For example, some races have "extended family" social structures, and will not place as severe a demand on housing as races which favour the "nuclear family" structure. In recent times, the trend in increasing family dissolution has resulted in increased pressure on existing housing resources.

Races within a nation frequently experience differing infant mortality rates, differing life expectancies, and varying susceptibility of races to infections and diseases. Knowledge of the racial components of the population will assist in planning for the health needs, on both a national and regional basis. Changes in fertility patterns will gradually alter the age structure of the population. A general decrease in the average number of children in each family will result in a gradual reduction of the number of persons entering the workforce, whereas a "baby boom" will place an increased demand on the educational resources in the near future. If the baby boom is followed by a gradual trend towards smaller

families, a "bulge" in the age structure of the population will be experienced, resulting in changing demands on the society as children of the baby boom era enter and leave school, then enter and leave the workforce. The age structure of the workforce is of particular importance, as this sector of the population must contribute to the national financial resources, to cater for the needs of the young and the elderly.

From the point of view of those who must plan and administer, a reliable source of information on all features of the population is essential in order to provide for the present and future needs of a nation. Such information needs to be as up-to-date as possible. A census is currently the only vehicle for the gathering of a detailed "snapshot image" of the total population. Surveys do not attempt to obtain complete coverage of the population, while administrative lists are seldom sufficiently detailed and often do not contain current information. Data bases may contain up-to-date information, but data on migrations within a country can only be obtained through keeping records such as all the financial transactions of each individual. Because of the problems of confidentiality, there is still some resistance to the prospect of data bases being utilised in such a manner.

The Encyclopaedia Britannica Micropaedia defines a census to be the enumeration of people, and also of houses, businesses, or other important items in a country or region at a particular time. Used alone, the term usually refers to a population census. The 11th Edition of the Encyclopaedia Britannica defines it as a term used to denote a periodical enumeration restricted, in modern times, to population, and occasionally to industries and industrial resources, but formerly extending to property of all kinds, for the purpose of assessment. The word is taken from Latin *censere*, to estimate or assess; connected by some with *centum*, *i.e.* a count by hundreds.

Early censuses were chiefly taken to assess the amount of tax to be paid by citizens, or to count the number of men eligible

for military service. In addition to enumerating the population, modern censuses are used to obtain information necessary to analyse social conditions and to assess the effectiveness of government policies.

A census provides a picture of the population as at a given moment. Although some of the questions relate to the individual's past - such as birthplace, education, and number of children born - each census is better as a current record than as a historical record. Historical trends are properly revealed only by a succession of censuses. Censuses are generally taken on a regular basis every 5 or 10 years. If censuses are conducted on a 10-yearly cycle, information from these censuses is usually supplemented by intercensal surveys.

For individuals in the population, the censuses may seem an unwelcome intrusion into their privacy. However, as well as research, census data is used for such important decisions as political representation, allocation of government funds, planning for educational facilities, housing needs, health facilities, and the siting of industrial plants. An accurate description of the population would require the enumeration of every member of the population, with all the questionnaires being completed without omissions or inaccuracies or deliberately erroneous information. In reality, such a happy state is unobtainable. Even if total coverage were obtained (that is, every member of the population is counted), the information supplied would not always be accurate or complete.

Decision makers and planners must rely on publicity and educational campaigns to persuade the public of the necessity of regular, successful censuses. Quality control procedures must be carefully monitored to ensure that the published census data is as free from errors and omissions as possible. If a census does not achieve a good response rate, then analysis of the data will produce estimates which are not accurate, and decisions made on the basis of these estimates may result in undesirable consequences.

Because of the high mobility of persons in the 15-30 year age group, problems are frequently experienced in contacting them for enumeration. Minority population groups are also prone to higher underenumeration rates than the rest of the population, possibly because they do not understand the purpose of the census, or they have fears about the usage made of the census data. Following any census, an evaluation should be made of the coverage achieved, both for the total population and subgroups of the population, and also the quality of census data.

Pretests and pilot tests should be performed prior to a census to evaluate the questions contained in the census questionnaire and the design of the questionnaire. Field testing and dress rehearsals should be used to evaluate the enumeration procedures and the final version of the census questionnaire.

Analyses, tables and diagrams should be made available using several types of media, and their availability widely publicised. This would help to to impress on the public the value of census data. The success of any census operation depends on the cooperation of the public, and no efforts must be spared in educating everyone about the value and necessity of regular, successful censuses.

Acknowledgements:

My sincere thanks to Dr Richard Brook, my Supervisor, who patiently assisted and guided me through the morass of literature, and to the following staff of the Department of Statistics, who generously gave of their time to furnish much of the information used in this thesis:

Terry O.'H. Papps	Executive Officer, Population Division
Jenny Ling	Executive Officer, Development
Frank Nolan	Manager, Mathematical Statistics
Len Cook	Assistant Government Statistician

Without their help and expertise this research would not have been possible, but the author accepts the responsibility for any errors or omissions in this review.

My thanks also to my husband, John, and my daughters Kirsten and Rachael for their patience, tolerance and support during the time this thesis was researched and prepared.

Table of Contents

List of Appendices

List of Figures

Chapter 1

HISTORICAL OVERVIEW OF CENSUS TAKING

Introduction

A census is a collection of information from every element of a population. The history of the population census dates back some three thousand years or so, but it was only in the last few hundred years that censuses were carried out at regular intervals and that extensive information was collected on each individual.

The term census stirs up strong feelings. This may seem strange, as it would appear to many of us to be a dry counting of the population, essential to the successful management of a nation.

Reasons for the reluctance of people to be counted are not difficult to imagine, for rulers usually instigated the census as a basis of taxation or conscription into the army. In more recent times, antagonism may arise as the census is seen as an unwelcome governmental infringement in their lives, particularly if they happen to be illegal immigrants or if they have overstayed their entry permits or have other reasons to fear the government of the day.

Early Censuses

Censuses have been conducted to obtain varying information from very ancient times and, looking back, we become aware of the fear and suspicion with which people viewed them. The Old Testament makes mention of the enumeration of the fighting men of Israel, males who were 20 years of age or older, and of the separate enumeration of the non-military Levites, who were

counted if they were at least 30 years old. Solomon had a similar census conducted in order to distribute the functions; this was reluctantly done by Joab at David's command (see 11 Samuel, Chapter 24). The antagonism of a section, at least, of the population against this enumeration is suggested by the comment that "God was angry with David" for this act. Apparently, a register of the population of each clan was kept during the Babylonian captivity, and the totals were published on the return to Jerusalem.

In the Persian Empire, some method was used to determine the resources of each province in order to fix the tribute, or taxation. In China, enumeration was an ancient institution in connection with provincial revenues and military liabilities. In Egypt, Amassis had each individual's occupation registered annually in order to aid official supervision of morals by discouraging disreputable occupations. In Greece, according to Herodotus, Solon introduced the annual registration of occupations into the Athenian administration scheme, and this later developed into an electoral record.

Roman Censuses

Rome first established the regular system from which the name of the enquiry, census, is derived. The original Roman census is attributed to Servius Tullius, who is said to have decreed that the population be enumerated every fifth year, along with the property of each family including land, livestock, slaves and freedmen. In those days, Roman society was rigidly structured into 6 main classes, ranging from patricians (noblemen) to plebians (commoners). Slaves occupied a place below all of these classes. The main object of the census was to ensure the accurate division of people into these classes and their respective centuries, based on considerations of combined numbers and wealth.

The word census comes from Latin, from the same stem as censor, which has strong moral implications. This seems very strange at first sight and, indeed, the link between morality and data collection is interesting: Enumeration of the people was only one of a group of many functions which were performed by the two magistrates of the highest importance in the Roman republic, called Censors. The Censors' duty was to take census of the citizens of Rome, to estimate their property, to accordingly impose taxes, and to punish offences. The Censor's responsibility included maintaining the standards of morality and the conventional requirements of Roman custom; bad cultivation of land, disreputable occupations, luxuriousness and celibacy were punishable offences. Any citizen who neglected his registration for the census risked high penal consequences.

The functions of the Censors were especially directed to the objects of public revenue, and apparently the enumeration of the people was not deemed of value as a source of statistical knowledge which might influence morals or legislation. This may be why so little is known today about statistical considerations, or questions, about the population and extent of the city of Rome itself in those times.

The census, being so important, was conducted every 5 years, and was followed by a religious sacrifice of purification or lustration, offered on behalf of the people by the censors or functionaries in charge of the classification. The term lustrum is now taken to mean a period of 5 years. The word census came to mean the property qualification of class as well as the process of registering the individual. It was later used in the sense of taxation, in which it has survived in the contracted form of cess, which is defined in the Oxford Dictionary as tax, levy or rate.

In the time of Augustus, 5 BC, the census was extended to the whole empire. According to the Gospel of St Luke, Augustus "ordered the whole world to be taxed" or, according to revised versions, "to be enrolled". Unfortunately, the compilation of

this, the most comprehensive enumeration yet attempted, was never completed due, apparently, to the emperor's death.

After the collapse of Rome, the practice of census-taking was discontinued until the modern period, with few exceptions. The Domesday inquest of England in 1086 was made to acquaint William the Conqueror of the land holders and holdings of his new domain. Less comprehensive censuses were the Breviary of Charlemagne and the almost complete count of the German city of Nürnberg (Nuremburg), made in 1449, under the threat of siege. Like the Roman census, the former two inquiries took little or no account of the population at large.

The Domesday Book

The aims of Domesday were to settle disputes over tenures after the wholesale transfer of estates from English to Norman hands, to establish exactly what services the aristocracy owed the crown, and to provide an up-to-date account of potential tax revenues. The survey provoked dissent and in 1086 Robert Losinga, Bishop of Hereford, wrote that *"the land was vexed with much violence arising from the collection of the royal taxes"*.

Relatively little written documentation from the 11th century has survived, and the current knowledge of the compilation of the Domesday Book has been deduced from a mere handful of documents. The Domesday Survey was a survey of the manors, not of individuals, and it is believed that the survey was conducted by dividing the counties of England into either seven or nine circuits, which were visited by teams of commissioners. The counties and possible circuits are displayed in Figure 1.1 on page 5.

Ten questions were put to the king's tenants-in-chief or their agents, and to be answered for three time periods: in the time of King Edward (that is, before the Norman Conquest); at the

Figure 1.1 Domesday Counties and Possible Circuits

Source: *Domesday 1086-1986: an exhibition to celebrate the 900th anniversary of Domesday Book, 1986* (Public Record Office)

time when King William gave the manors (after the Conquest in 1066); and at the time of the survey(1086).

The Domesday questions were as follows:

1. *What is the name of the manor?*
2. *Who held it in the time of King Edward?*
3. *Who holds it now?*
4. *How many hides are there? (i.e. what is its assessment for the geld tax? According to the Oxford Illustrated Dictionary (1962), a hide is a medieval measure of land, of varying extent, and originally meant the amount required by one free family and its dependents, or as much as could be tilled with one plough in a year.)*
5. *How many ploughteams, in demesne (i.e. on the lord's land) and among the men (i.e. the rest of the village)?*
6. *How many villeins(feudal tenants entirely subject to lord or attached to manor)?*
   How many cottars (peasants occupying cottages and labouring as required)?
   How many slaves?
   How many freemen (labourers who were neither slaves nor serfs [labourers not allowed to leave the land on which they worked])?
   How many sokemen (persons with right of local jurisdiction)?
7. *How much wood? How much meadow? How much pasture? How many mills? How many fisheries?*
8. *How much has been added or taken away?*
9. *How much was the whole worth? How much is it worth now?*
10. *How much had or has each freeman or each sokeman?*

The final question asked, relating to the time of the survey, was

11. *And whether more can be had than is had (i.e. can the manor raise more tax revenue?)*

The answers to these questions, either written or verbal, were subjected to the scrutiny of the hundred juries in the county courts. Following the proof of the information in court, it was copied and sent to Winchester, where the information for the whole country as far north as Yorkshire, but excluding Essex,

Suffolk and Norfolk, was edited to form Great Domesday. The entries for the eastern counties were not edited but simply bound together to form Little Domesday. Unfortunately, Great Domesday is not dated, and varying opinions exist as to when it was completed. However, the information for both volumes was probably assembled by the Summer of 1096 and possibly by 1 August, when a great assembly was held at Salisbury. If this is so, then production of the book could then have started in the previous winter. The Domesday Book had probably reached its final stages by 9 September, 1087, when William died in Rouen.

It is interesting to note that it is still valid for the Domesday Book to be cited in English courts of law when questions are raised of ownership of land or of some commercial practices.

Development of the Modern Census

The modern idea of a population census as being a complete enumeration of all people and their important characteristics, such as housing, marital status and employment, slowly arose in the 17th and 18th Centuries, due to the desire to understand the basic structure and trends of society. Contrast this with the aim of the earlier censuses: identification and control of particular individuals. There is no such thing as "the first modern census", as none of the early censuses embodied all the modern features.

Census taking as we know it today has evolved from three parallel developments:

(i) the gradual change to national enumerations for general scientific and governmental purposes;
(ii) the improvement of administrative machinery, techniques and accuracy of enumeration, including legal safeguards to assure confidentiality; and
(iii) the deepening and systemization of the types of information obtained.

The United Nations (1958, p.4) gave a definition of a modern population census. Six essential features of a census were listed:

1. A census must have *national sponsorship.*
2. A census must cover a *precisely defined territory*; boundary changes that affect comparisons between successive censuses should be clearly and explicitly stated.
3. *All persons* in the scope of the census must be included without duplication or omission. (If sampling is used, it must give every member of a stratum, or subgroup of the population, equal likelihood of selection.)
4. The people must be counted as of a *fixed time.* Persons born after the census date are to be excluded, and persons who die after the census date are to be included.
5. Census data must be obtained separately for *each individual.*
6. The data from a census must be *published.*

The Establishment of Regular Enumerations

The first modern effort to count everyone at successive intervals was made in La Nouvelle France, now known as Quebec, and Acadie, which now bears the name Nova Scotia, where 6 enumerations were made between 1665-1754. In Paris, domestic occurrences such as births, deaths and marriages in the locality were registered and periodically published, and in 1670 Colbert ordered the extension of the system to the rural communes. Colbert, a statesman under Louis XIV, was a remarkable man who reformed French financial administration, developed industry tariffs, virtually founded the French Navy, and founded the French Academies of Literature, Science, and the Fine Arts.

In 1749, the Swedish clergy were required to render returns of their parishioners, from which the total population of Sweden,

including Finland, was obtained. 1787 marked a similar development for Denmark.

Several Italian states conducted approximate enumerations, including Sardinia in 1773 and 1795, Parma in 1770 and Tuscany in 1766. From 1742 onwards, several German states conducted enumerations.

The British Board of Trade ordered 27 censuses in the North America colonies between 1635 and 1776. After independence in 1776, these former colonies continued to take censuses until the establishment of the United States of America. The first Federal census of the United States of America was taken in 1790. It is noteworthy both for the size of the area enumerated, for the effort made to obtain data on characteristics of the population, and for the political purpose for which it is undertaken, namely representation on the basis of the population.

The United States of America has continued to take censuses every 10 years since 1790. The first American census neglected to obtain information on occupation, birthplace, marital status and exact age. The 1800 Census included a 5-year classification of whites, and from 1850 onwards, the individual was used as a unit, rather than the family as had previously been the case.

In England in 1753, a private member of the House of Commons introduced a bill to provide for the annual enumeration of the people and of persons in receipt of parochial relief. Despite official support, the bill was violently opposed and it was denounced as *"sacrilegious"* , *"likely to result in some public misfortune or epidemical distemper"* and a *"project ..... totally subversive of the last remains of English liberty (and) calculated to reveal our weakness to our enemies"* . The bill was passed in the House of Commons, but was thrown out of the House of Lords.

However, by the end of the century, general opinion had changed to the extent that it was thought desirable to know the relations between an increasing population and the means of subsistence. A census bill was again introduced by a private member, and this bill was passed without opposition at the end of 1800.

The United Kingdom of Great Britain and Ireland was created by Act of Parliament in 1801 and the first census of Great Britain was conducted in 1801, but it did not include Ireland. France took censuses in 1800 and 1806, but the administrative machinery was poor, and it was another 30 years before the enumerations could be accepted as valid.

Censuses of England, Scotland and Wales were conducted every 10 years, but the first attempt at a general census of Ireland was not made until 1811. However, because of the antagonism of the Irish Catholics towards the British, it was a disaster, and the successive census in 1821 only achieved a bare enumeration, which is not believed to be accurate. At the time of the creation of the United Kingdom, Catholic bishops and dignitaries were banished from Ireland, and catholics were forced to contribute towards the support of the Anglican Church. Despite being the majority of the Irish population, Irish catholics had no vote, could not bear arms, could not seek higher education, practice law, buy land, or hold office. Rent was paid to absentee landlords in England, whose agents had absolute power over their tenants. Voting rights were not given to Irish catholics until 1829.

The 1831 Irish Census results were believed to be subject to overenumeration, as enumerators were paid according to the numbers they returned. The results were corrected in 1934, and made the basis of the new system of national education. The 1841 and 1851 Censuses were more successful, probably because police constables were employed as enumerators. The latter two censuses were notable for the valuable statistics collected on the rural economy of Ireland. The schedules for

the Irish censuses were completed by enumerators, whereas those for the English, Welsh and Scottish censuses were completed by the householders.

The English and Scottish censuses did not secure age data until 1841, and no information on marital status was obtained until 1851. A uniform system of registration of births, deaths and marriages came into operation in England and Wales in 1837, and this enabled the accuracy of returns to be checked; it also provided the appropriate machinery for census taking. Parish schoolmasters were employed as enumerators in the country districts of Scotland.

NEW ZEALAND HISTORY OF CENSUS TAKING FROM 1840 ONWARDS

Early Population Estimates

New Zealand was proclaimed a crown colony of Britain in 1840. In the few years prior to this, some statistics had been collected and evaluated. Reports from Select Committees on New Zealand, in the British Parliamentary Papers, 1837-40, contain several estimates of the European and Maori populations made by missionaries, traders and officers of passing ships. At that time, there were many Maoris in the north of the North Island and hence the bulk of the non-Maori settlement, which chiefly consisted of traders and missionaries, was localised in that area. These early estimates only related to the populations in that area. There were also pockets of non-Maori settlements in the south of the South Island, where industries such as sealing were practised, but there are no records of any estimates of the Maori population in that area. It is probably true to say that the sealers were not concerned about the welfare of the local Maori population, and saw no point in spending time on estimating the size of the native population.

Appendices 1.1 and 1.2 contain extracts taken from the Select Committee reports. Appendix 1.1 illustrates how crude these early estimates of the size of the Maori and non-Maori populations, and the bareness of other information gleaned about the Maori population, were. This information, though sparse, does yield a picture of the geographical distributions of the two populations, although it must be kept in mind that several of the testimonies in the reports were based on hearsay alone. Appendix 1.2 focuses on the character of the non-Maori persons in New Zealand, as attested to by the witnesses, and these clearly give the impression that the non-Maoris living in New Zealand at that time fell into two categories; those who were in New Zealand to attend to the spiritual well-being of the Maori population, and those who were there for financial or other form of gain, which included evading or escaping from authority.

First Population Censuses (Non-Maori)

Along with the status of the colony went the usual policy of the Colonial Office of carrying out a population census. Schedules were sent to Governor Hobson in 1840, who forwarded them to the resident police magistrate of each settlement or town. The Colonial Secretary was in charge of organising the collection of the information, and forms were sent to the police magistrates of each town or settlement for completion. The system was put into operation in 1842. In 1840 and 1841, population numbers were based on official estimates.

Between 1841 and 1851, resident magistrates regularly took official population counts in the settlements of the new colony. As for all British colonies, the data obtained from the census was carefully written up in the Blue Books, so called because of the colour of their covers. The information collected was supplied to the Secretary of State for War and Colonies, and included the following particulars:

Description of the county, district or parish
Area in square miles
Number of "Whites" and "Coloured" (excluding Maoris) by sex
Number of "Aliens and Resident Strangers"
Population to the square mile
Occupation status of the population
Number of births, deaths and marriages in that year.

The original Colonial Office scheme which commenced in New Zealand in 1842 included the collection of statistics on religious professions. In the main settlements, particulars of religious affiliations were noted when the census was taken by the police magistrates. Clergymen were also required to make annual returns, but these returns were incomplete. In general, the geographic boundaries were as for the population census, and religious affiliations of the military and their families were not included. The religions of minors were assumed to be the same as those of their parents or guardians.

However, these enumerations included only the non-Maori population, and even then only the population living within the settlements. Even for such a simple and straightforward census, the time, effort and money it consumed was quite considerable. As the early censuses provided no financial gains in the form of taxes for either the local Government or the English Government, they must have been taken in an effort to obtain knowledge of the colony for administrative purposes.

In comparison with present day methods, this first census was very simple, cheap, and somewhat arbitrary. Who carried out the actual enumeration? We must assume that this was one additional task added to the busy life of the local policemen. Information would have been tallied by first hand observation, by hearsay, or by any other method at his disposal. Generally, the information on each family would have been supplied by the head of the household, and particulars such as the residential address of each member of the family would have been supplied,

rather than the actual location of each person at the time of the enumeration. The result was a **census de jure**. It was somewhat rough and ready when compared with modern methods, particularly as the task of enumeration would have had to be fitted in as other duties permitted, and hence the data for each town or district was undoubtedly collected over a long period of time, and probably with varying degrees of enthusiasm. The current practice is to employ staff for the specific purpose of enumeration, and to conduct a simultaneous census throughout the country.

The households were enumerated street by street, as were the occupants of the gaol, the military and those residing in surrounding settlements. Occasionally, a return of residents engaged in shipping was furnished by the Collector of Customs. It was general practice not to include the military and their families in official Blue Book returns.

According to Simkin (1954), checks of the official records indicated a high level of accuracy for the settlements covered. This is quite remarkable, since the accuracy of the returns was entirely dependent on the attitude and ability of the police magistrates. Conditions in the Bay of Islands and Hokianga were unsettled in those early years following the transfer of the capital to Auckland; the outbreak of Heke's War forced the abandonment of the census in the north for 1845 and 1846.

The Blue Books were not designed for general circulation. From 1841-1847, the New Zealand Government Gazette provided a source of official statistics, but these were limited to particular parts of the Colony, or to particular periods or occasions. At that stage, New Zealand was separated into two provinces: New Ulster, which was that part of the North Island lying north of a line running east from the mouth of the Patea River; the remainder of New Zealand constituted New Munster (New Zealand Census of Population and Dwellings 1971 - The New Zealand People, 1973). In 1848, the New Ulster Government Gazette and the New Munster Government Gazette

replaced the single official publication. After 1847, the Blue Books were also produced separately for the two provinces of New Ulster and New Munster. Statistics of New Munster were printed in 1849 by order of the Legislative Council.

The Imperial Act of 1853 granted representative government to the expanding colony, and replaced the existing two provinces with six new provinces: Auckland, New Plymouth, Wellington, Nelson, Canterbury and Otago. Statistics of Nelson from 1843 to 1854, inclusive, were published by the Provincial Government in 1855, followed by a similar compilation for 1855, which was published in 1856. Statistics of New Plymouth, from 1853 to 1856, were published by the Provincial Government, and the census returns and various occasional publications in the Gazettes provided differing amounts of information on the provinces.

One of the actual census books used in the preparation of the returns has been traced. The Auckland Police Census Book is now lodged in the Auckland Public Library, and contains complete manuscripts for the Auckland province from 1842 to 1846. The information collected was more detailed than that required by the census form, and included the age distribution, religion, and number and type of houses occupied. This additional information was also collected in other districts, so was probably required by the Colonial Office, but was not included in the Blue Book returns, nor, apparently, was it regularly included in the Governor's despatches.

Other means of checking population movements were the regular returns and special reports often furnished by members of the three missionary societies. Also, when government officials made occasional journeys "to the interior" or along the coast, they were instructed to report on Pakeha, that is non-Maori, and Maori populations. A number of reports from the New Zealand Company's agents and correspondence to the New Zealand Journal were also used to check official population statistics. In most cases, the statistics collected by the several sources

agree, and where there are discrepancies, these are mainly a result of using different boundaries.

The most unsatisfactory official returns are those for "aliens and resident strangers". A lack of definition of these terms caused them on different occasions to be included with the British population or listed separately. Sometimes the military were included, and sometimes the figure quoted was merely an estimate. However, these criticisms apply more particularly to the earlier annual returns, as naturalisation later led to absorption of this category into the British population.

Possible sources for determining the populations of the more remote coastal and inland settlements which were not included in the official returns are old land claim files and other official papers, missionary records and private documents.

In the settlements of Auckland, New Plymouth, Wellington and Nelson, the usual practice was to enumerate and classify buildings when taking the annual census for the period 1840-1852. Although there was no Blue Book return for buildings, the Colonial Office probably required the collection of the information.

In some business areas, it was common practice to reside in rooms which formed part of shop premises. Hence some buildings were designated as houses, when they should really have been designated as commercial buildings.

These official building statistics do not include flour and flax mills, small factory buildings, public buildings such as churches and halls, military barracks, hospitals, courthouses and gaols. Statistics relating to the New Zealand Company settlements give a more accurate picture of the total number of buildings, and sometimes include specification of the material of the houses and buildings.

Quality of Early Census Data

How were the enumerations received by the settlers, and other Europeans? They were conceived and carried out by those with authority whose duties included maintaining law and order. Consequently, those persons fearing or distrusting the local authority would undoubtedly have avoided enumeration whenever possible, and as the enumeration was taken by the police, it is highly unlikely that a good coverage was obtained. Furthermore, the data collected would have been susceptible to a number of additional errors and biases. The census dates were not always uniform for the whole of New Zealand. Until 1851, the usual practice was to take the census in the last week of December or the first week in January (at the discretion of the local magistrate), but the period was occasionally extended from August to February. Various methods of enumeration were employed, and there was also variation in the scope of questions asked. Delay in forwarding returns occasionally caused their loss. The 1842 return from Nelson was never received, as the local police magistrate failed to forward the return at the proper time and was killed in the Wairau Massacre of June 1843.

Because the census boundaries were only defined for the main settlements, and non-standardised schedules were used, the early censuses did not provide a satisfactory basis for computing statistics for the whole country. Both details and whole categories of information were sometimes missing. However, although these early enumerations were incomplete as far as the total population of New Zealand was concerned, there is sufficient evidence to prove that early enumerations were entitled to rank as censuses. The isolation of towns and settlements and the difficulties of transport made collection of data a lengthy process, but the relative smallness of the population reduced the margin of error.

Development of Simultaneous Standard Census

By 1851, immigration had caused rapid growth of the colony, and it was clear that a simultaneous, uniform census throughout the country was needed. The Legislative Assembly passed a census ordinance in 1851, providing for a general census to be taken that year and "in every first, fourth and seventh year in every decade of years". The first national census was taken in 1851, and comprehensive enumeration of the non-Maori population was attempted, although the census was not taken simultaneously throughout the country. The actual enumeration was conducted by the provincial governments of New Ulster and New Munster. However, no second enumeration was ever taken under the ordinance. This was because the number of provinces, at different periods prior to 1853, varied from 6 to 10. In 1853, 6 new provincial governments were formed, each with its own Legislative Council, to be individually responsible for census enumeration. As mentioned above, these provinces were Auckland, New Plymouth, Wellington, Nelson, Canterbury and Otago. Each province (with the possible exception of Otago) passed an individual census ordinance in 1854 or 1855 and conducted its own census. A standard, based on the English model, was used to obtain information about new settlers, locations, names, ages, sex, infirmities, industries, fenced land and livestock. However, the obligation was not fulfilled, and the censuses of 1854 and 1857 were incomplete.

"Statistics of New Zealand, 1853, 1854, 1855 and 1856" was the first attempt to present one comprehensive and authorised compilation of the general statistics of the entire colony. However, in the introductory memorandum , the Registrar-General reported that the tables for the whole colony could not be even approximately completed, due to tardy, omitted or non-uniform returns. Certain returns were lost in the fire which destroyed the Wellington public offices, and of those returns received, the censuses were taken in varying months of the year for different provinces, and even in different months on successive occasions in the same province. *"A want of*

uniformity in the Schedules, not merely as to details, but as to important branches of information" was also reported. The Registrar-General also reported that information was *"taken at different times within the last two years in the several Provinces"* for the 1857 Census.

The Census Act of 1858 repealed the ordinance and instituted a national, integrated census to be taken at regular 3-yearly intervals, beginning in 1858. The Act also provided for the employment of census enumerators as official census collectors. Hence, by this stage, the enumerations were beginning to resemble the modern censuses, as they were now being taken on a national basis, at regular time intervals, and official enumerators were employed solely for the purposes of the census, rather than imposing additional duties on to the resident magistrates. The Registrar-General was able to report that the numbers given by the enumerators were *"generally correct in the Totals, and, indeed, in all that could be regarded as practically important"*, although the numbers *"were found on more minute analysis to contain minor discrepancies"* .

The Act of 1858 was amended in 1860, 1867, 1870, 1873 and 1876, and finally replaced by the Census Act of 1877. Prior to 1877, each province compiled its own census statistics and then forwarded them on to the Registrar-General. The 1877 Act repealed earlier legislation and provided for a census in every fifth year following the 1881 Census. The Registrar-General became responsible for the entire census operation and the census processing was centralised in Wellington, providing an opportunity for uniform interpretation and presentation of data. The quinquennial census cycle has been maintained since 1881 with only two interruptions; namely the abandonment of the 1931 Census, forced by the economic recession, and the postponement of the 1941 Census until 1945 and the subsequent cancellation of the 1946 Census, because of the outbreak of World War Two.

Unfortunately, the decision to abandon the 1931 Census has prevented a record of the social and economic status of the population in the depths of the worst economic depression up until that date.

Further amendments to the 1877 Act followed in 1880 and 1890, and an act of 1910 created a separate Census and Statistics Office, under the responsibility of a Government statistician. During the 1911 Census, a sub-enumerator was drowned while crossing a swollen river on horseback. In 1916, the Post Office agreed to cooperate with the enumeration of the census. The enumeration districts were redefined to facilitate the new enumeration procedure, and the Postmaster of a town centrally located in each district was appointed enumerator. Each enumerator was responsible for the appointment of sub-enumerators and the collection routine for their own district, allowing local knowledge to be used to full advantage.

Subsequent Statistics Acts were passed in 1926 and 1955, and in 1955 the Census and Statistics Office was extended to a separate department of state: the Department of Statistics. The current legislative authority for the census is the Statistics Act of 1975, in conjunction with the Amendment Acts of 1978, 1982, 1985 and 1986.

The Post Office continued to provide the staff and facilities for organising and controlling the census field work until prior to the 1986 Census, when a decision by Post Office Headquarters that the census duties would be conducted out of office hours, and at overtime rates persuaded the Department of Statistics to conduct all facets of the census operation itself.

Listed in Appendix 1.4 are the dates of New Zealand censuses from 1851 to 1986.

Current Legal Requirements of Census

The 1975 Statistics Act requires that questions on 9 subjects be included in all national censuses. The Act also mentions 25 topics that are considered to be of national value. Because of the importance of these topics, they are normally included in all censuses as a matter of course, and are referred to as standard questions. These two groups of subjects, namely compulsory and standard, form the basis of the New Zealand Census. Extracts from the 1975 Statistics Act are given in Appendix 1.3, and include a complete listing of compulsory and standard groups of subjects. The Act also provides for additional, unspecified questions to be asked where these would meet a specific need for information.

MAORI CENSUSES

Early Estimates

From 1769 to about 1830 very few estimates of the Maori population were made, and the correct sources of those estimates which survived is unclear. 'Counts' were made of some small areas, but these often were really estimates, ranging from those based on detailed personal knowledge of a district to those based on hearsay. In a few areas, some officials and missionaries took actual headcounts and attempted to distinguish between adults and children, and to tabulate tribes. Estimates are attributed to Cook (although some argue that the estimate was Forster's) and Nicholas.

Cook's estimate was based on settlements visited or coastal villages observed from shipboard. According to Lewthwaite (1950), Cook believed that the interior and the west coast of the North Island from Cape Maria van Diemen in the far north to Taranaki were uninhabited. Lewthwaite produced archaeological and other evidence which indicates that there

were in fact relatively dense settlements in parts of the interior. Nicholas's estimate was based on a brief visit, mainly to the Bay of Islands, in late 1814 to early 1815.

From approximately 1830-1840, non-Maori settlement was concentrated in the north of the North Island around the Bay of Islands, and in the south of the South Island. Missionaries, traders and visiting ships' officers estimated the size of the Maori population in the north. However, in the south, non-Maoris were engaged chiefly in industries such as sealing, and were not interested in enumerating the few Maoris in the area. Estimates for this period are attributed to Yate, Hinds, Bannister, Baring, Coates, Williams, Terry, Fox, New Zealand Company Report, Polack, Crawford, and Wilkes. Appendix 1.1 contains some of estimates of the European and Maori populations as supplied by Nicholas, Flatt, Watkins, Montefiore, Wilkinson, Baring, Coates and Beecham to the Select Committees on New Zealand (British Parliamentary Papers, 1837-40).

Both Yate and Williams were members of the Church Missionary Society. Williams had travelled widely and had made regional estimates of the populations around mission stations. Coates was Secretary to the Church Missionary Society, and his evidence consisted of a letter quoted from the Reverend Williams. Similarly, Hinds quoted missionary records and Bannister quoted a Reverend W. Gate (presumably Reverend Yate). Terry visited New Zealand for approximately a year, whereas Nicholas had spent 10 weeks in the country. Both men had published an account of the country. Fox merely quoted missionaries. Polack was a trader who had spent six years in New Zealand, and his estimate would seem to be the most reliable, but it must be remembered that it was still merely an estimate of the population. Crawford had been in New Zealand for a year or more and Hamlin's estimate was apparently from a combination of missionary records and personal experience. Wilks had only visited the country briefly, and it is likely that his estimate was based on hearsay. Flatt was a missionary

who spent two and a half years in New Zealand and Watkins, a private surgeon, had spent three months in the country botanising. Montefiore had chartered a ship for a pleasure tour, and had only occasionally gone ashore. Most of his evidence consisted of hearsay. The Reverend Wilkinson was an Anglican clergyman who had spent 4 months in New Zealand. Baring, an M.P., quoted correspondence as did Beecham, who was secretary to the Wesleyan Missionary Society.

From 1840 to 1857-8, further small-scale detailed estimates of small areas were made, as well as a number of estimates of the total Maori population. In contrast to earlier estimates, these were often based on widespread travel, and the more detailed estimates gave an idea of the geographical distribution of the Maori population. Estimates by Dieffenbach, Swainson, Shortland, Clarke, and Halswell indicate that Northland, Waikato-King Country, the Bay of Plenty and the East Coast had large populations, while the Firth of Thames and the southern regions had smaller numbers. Later estimates by Taylor, McLean and Fenton suggest a decline in the population. Estimates of the Maori population are also attributed to Grey, Fox and Thomson, while Fox's estimate was based on his analyses of various estimates of the period, and Thomson claimed access to official, but unspecified, sources.

The First "Census"

The first census-type enumeration of Maoris, attributed to Fenton, covered the whole of New Zealand, excluding Nelson. The figures for Nelson were derived from a provincial census of 1855. Fenton's Census took place over twelve months in 1857-8 and as well as enumerating the Maoris, it also tabulated 'adults' and 'minors' for some regions. However, suspicion and hostility on the part of the Maoris caused unsystematic enumeration or mere estimates in many areas. Fenton personally conducted the enumeration of the Waikato, a region on the verge of war, as he was well known to the Maoris there. In other

areas, under-enumeration occurred even when the census takers were systematic and thorough.

From 1860 onwards, warfare between Maori and Pakeha prevented any census being taken of the Maori population for at least 10 years, and an estimate was made for the North Island in 1867. In 1868, a census was made of the South Island Maoris, consisting of enumerations made in some South Island districts.

Establishment of Regular Enumerations

After commencement of regular census-taking in 1874, enumerations were taken in the same years as non-Maori censuses on every occasion, although the information collected was less comprehensive than that for the rest of the population. Considerable difficulties were experienced, due to the nomadic habits of some tribal groups and the constraints of language and literacy. Antagonism towards the Pakeha must have been another dominant factor. Comparisons of the 1874, 1878, 1881 Maori censuses with data from other sources indicates that underenumeration occurred, and enumerators involved in the 1891 Census reported fairly severe underenumeration in certain regions.

After the 1874 Maori Census, special books were supplied to the Native Departments in each district, and data was collected on numbers, sex, ages (in very broad categories) and tribes. The Maori censuses were conducted over varying periods of time, as it was considered impractical to attempt the enumerations on one particular night. From the 1886 Maori Census onwards, the classifications were by narrower age-groups and by sex for the total population, for regions and, until 1901, for tribes. The schedules were completed by officials who were usually district officers of the Native Department. For these censuses, Maoris who were not of full Maori blood were arbitrarily allocated to the categories 'Half-castes living as Maoris'

(included in the Maori census) and 'Half-castes living as Europeans' (excluded from the census), depending on the whim of each enumerator. We thus have an incomplete picture of the Maori population at that time, making it difficult to compare these earlier censuses with more recent ones.

The 1901 census enumerators reported that the enumerations were the most thorough up until that time, but the census still consisted of a mere headcount. The 1906 Census, conducted under the supervision of the Justice Department, used a rough classification of the Maori population in terms of age and mode of living. The two age groups were 'Under 15 years of age', and '15 years of age and over'. Maoris still living as members of tribes were enumerated separately from those living in "European" communities.

Gradual Integration of Maori and Non-Maori Censuses

In 1916, the first attempt was made to integrate the Maori and non-Maori enumerations, but the experiment was confined to the South Island. The same schedules and subenumerators as for the non-Maori census were used, and for the first time, household heads were responsible for the enumeration. For the 1916 Census, many enumerators reported experiencing difficulty in obtaining information because the Maori feared that the statistics would be used for military recruitment purposes. The dual system of integrated enumeration for the South Island, and the usage of special books, which were completed by enumerators, for the North Island was repeated for the 1921 Census.

The 1926 Maori Census marks the first enumeration of the Maori population on a specific night, and the first count of the Maori populations of towns. A special Maori schedule was introduced for the North Island and the Chatham Islands, and all questions and guide notes were given in both English and Maori. The schedule was more detailed than previous schedules, although it

was still not as comprehensive as the standard European schedule, which was supplied to the South Island Maoris, and, when requested, to North Island Maoris. The standard European schedules were supplied to Maoris living in the South Island and in Stewart Island because they were so few in number. In contrast to earlier Maori censuses, throughout the whole country household heads were now responsible for completing the schedules.

One side of the Maori questionnaire collected the following personal data: name, sex, race, marriage, trade or occupation, religion, and usual residence, The standard European questionnaire contained the same questions, as well as questions on the relationship to the head of the household, orphanhood, industry, grade of occupation (employment status), length of residence, dependents and income. The differentiation between the two types of 'half-caste' was abandoned; that is, 'living as Europeans' versus 'living as Maoris'.

The 1926 Census marked the first time that detailed information was collected on the dwelling characteristics and conditions of Maoris. The dwelling schedule was printed on the reverse side of the Maori questionnaire. Questions common to both dwelling schedules were the nature of the dwelling, the number of occupants, the number of rooms, and the tenure. The standard European dwelling schedule also contained questions relating to the materials of outer walls, rent, whether the dwelling was permanent or temporary, flats, habitual residents, location (which was measured by the distance to the nearest Post Office), and the stock of poultry.

The 1926 Census marks the beginning of modern, comprehensive and completely integrated census-taking in New Zealand, and the procedure was repeated for the 1936 and 1945 censuses. From 1951 onwards, the enumeration procedures were identical for the Maori and non-Maori sectors of the population, and published census tabulations of the New Zealand population

included data on the Maoris. In 1951, a small number of schedules were printed in Maori, and were to be issued on request to Maoris in the North Island. There is no record of any of these schedules having been used.

Minor illogicalities regarding persons defined as 'part Maori' and 'part other non-European blood' were gradually but not entirely eliminated. The change in definition between the 1921 and 1926 censuses created the difficulty of comparing earlier censuses with recent censuses, as did further changes in racial classification after the 1951, 1976 and 1981 censuses.

Chapter 2

THE NEED FOR CENSUSES

Introduction

In modern times, the term **census** is used to denote an enumeration restricted to population, houses, businesses, or other important items in a country or region at a particular time. Used alone, the term usually refers to a population census. The word is taken from Latin *censere*, to estimate or assess, and it formerly extended to property of all kinds for the purpose of assessment. Over time, the concept of a census has slowly evolved from being a mere headcount of people and possessions for taxation purposes to its current position of providing an invaluable source of detailed varied and time-related information on the popuiation and its characteristics. Such information is necessary to assess the various needs of the population, both currently and in the future. Planning for new transport facilities, the siting of new schools, community centres, hospitals, shopping centres, extensions to public services such as electricity, gas, water, sanitation and postal services are undertaken using current census data. Data from the latest census can be compared with that from previous censuses to determine trends in migration, fertility, family patterns and employment. These can then be used to predict population figures for the future.

The Birth of Statistics as a Tool to Measure Social Development

Prior to the nineteenth century, the need for data collection in the social area was not recognised. During the nineteenth century, the concept of statistics as the collection of data and very basic data display and manipulation was gradually seen to

be a tool to analyse social conditions and the effectiveness of public policy. In 1814 Quételet, a Belgian astronomer-statistician regarded by many as the founder of modern social statistics, organised Belgium's central statistical bureau, which was adopted by other countries as a model for similar agencies. The emergence of statistics as a tool for social analysis was facilitated by the development of the mathematical theory of probability, the beginning of the modern state with its agencies for collecting information on its citizens and their activities, and the theoretical interest of political economists in finding causes for human social behaviour.

Although Florence Nightingale is best known for her work in relieving the misery of wounded British soldiers in the Crimean War by drastically improving their healthcare, she was also one of the earliest statisticians and a follower of Quételet. Florence systemised the chaotic record-keeping practices, used medical statistics to calculate mortality rates and to highlight the impact of disease, and pioneered the graphical representation of statistics by inventing polar-area charts. She also used graphs to compare the relative mortality of English soldiers in Crimea and of English male civilians of corresponding ages living at home.

Under Florence Nightingale's instructions, data was collected on the number of patients in a hospital at the beginning and end of each year, the number admitted during the year, the number of patients who had recovered or who had been discharged as incurable or dismissed at their request, the number of patients who had died and the average duration of their time in hospital. However, even the most basic personal information was not kept on the patients, such as their names, the illnesses or wounds suffered, and whether the soldiers recovered or died.

An article on Florence Nightingale by Cohen (1984) included illustrations of her diagrams and these have been reproduced in Figures 2.1 and 2.2 on pages 30 and 31. Figure 2.1 displays an

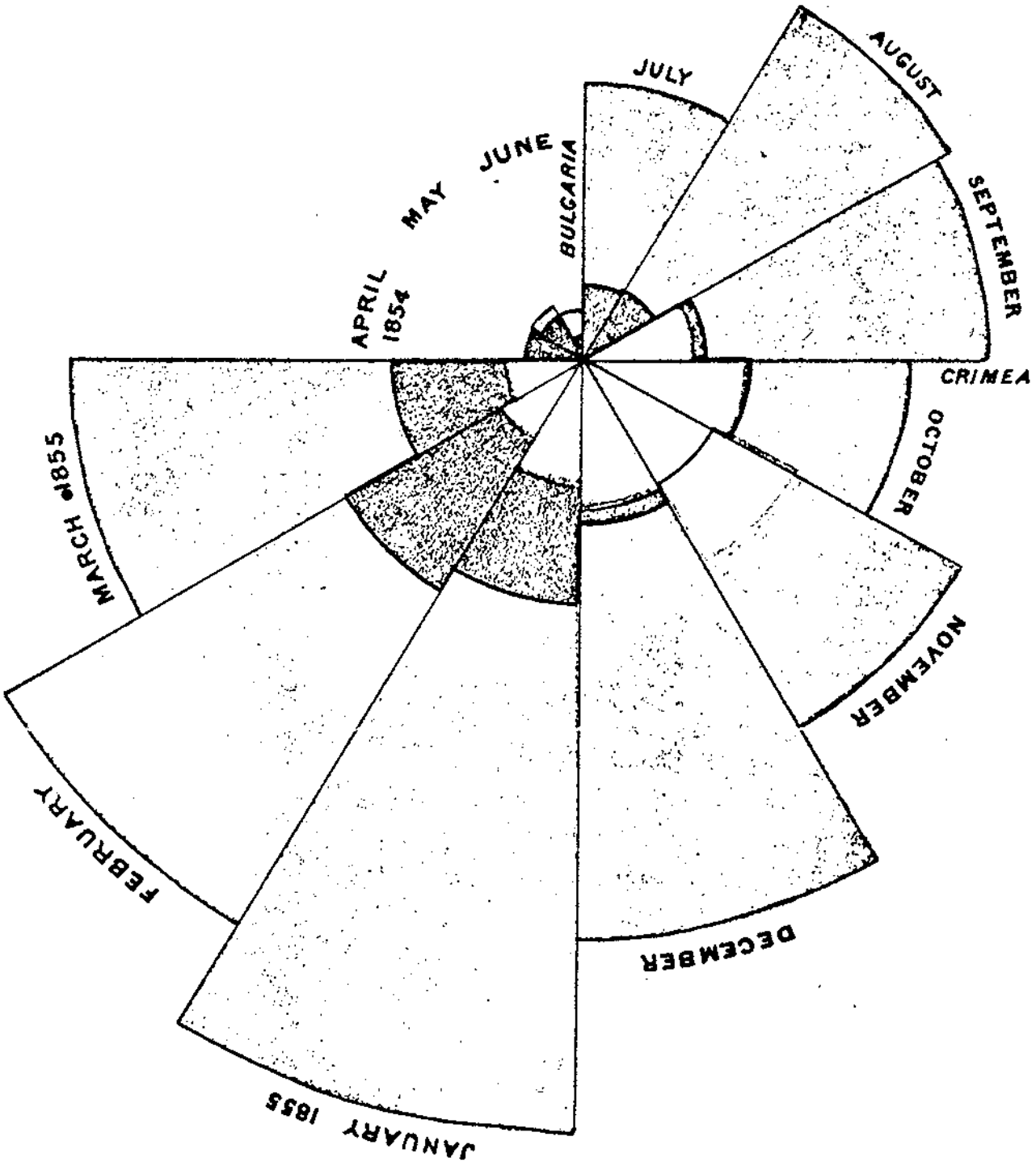


Figure 2.1 Polar Area Diagram invented by Florence Nightingale

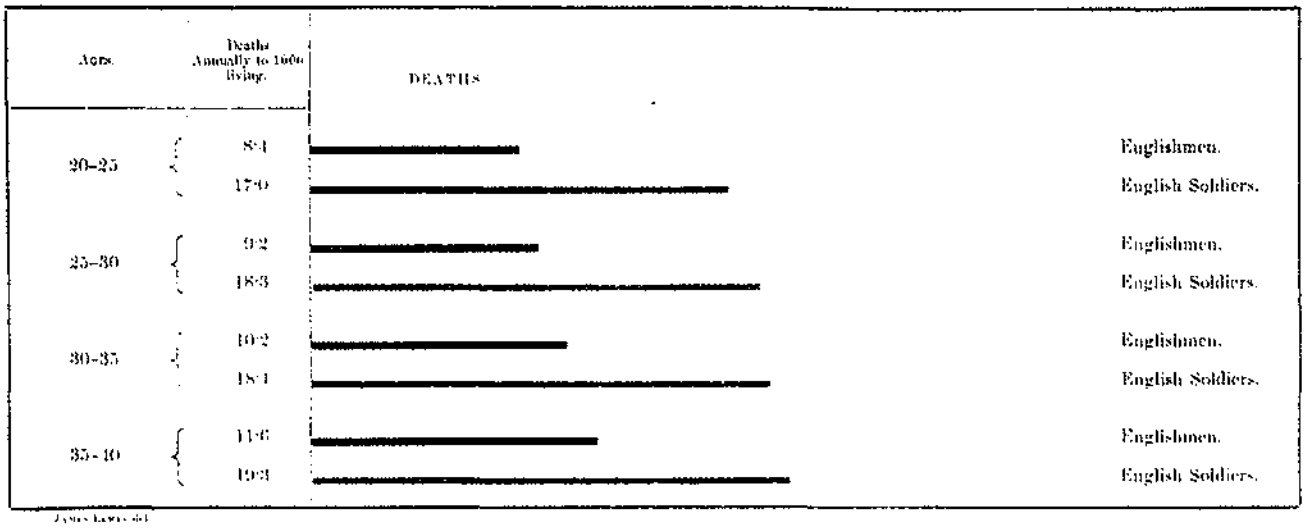
Source: Cohen, I.B. "Florence Nightingale" *Scientific American* March 1984, Vol 250 No 3

The area of each coloured wedge, measured from the centre, is proportional to the statistic beir measured. Mortality in British Hospitals peaked in January 1855, when 2,761 soldiers died contagious diseases, 83 of wounds and 324 of other causes.

Legend:

- blue wedges (the outer wedges) deaths from "preventable or mitigable zymotic" diseases (contagious diseases such as cholera and typhus),
- pink wedges (the inner wedges) deaths from wounds
- grey wedges (the middle wedges) deaths from all other causes

Representing the Relative Mortality of the **Army at Home** *and of the* **English Male Population** *at corresponding Ages.*



Representing the Relative Mortality, from different Causes, of the **Army in the East in Hospital** *and of the* **English Mal Population** *aged* 15—45.

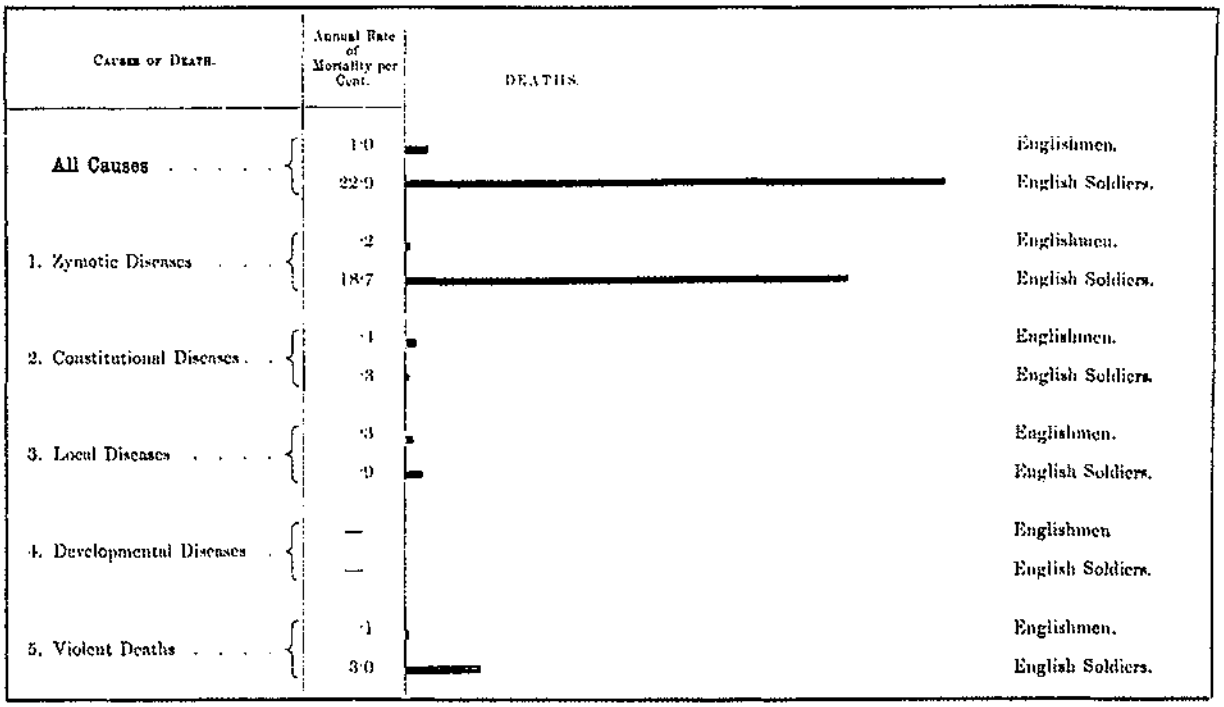


Figure 2.2 Line Diagram produced by Florence Nightingale

Source: Cohen, I.B. "Florence Nightingale" *Scientific American* March 1984, Vol 250, No 3

example of polar-area diagrams, which Florence called "coxcombs" and Figure 2.2 displays line diagrams taken from the Royal Commission Report, comparing conditions in the army to those in civilian life. As shown in the top diagram, mortality on the peacetime army in Britain was nearly twice as high as it was among civilians. The bottom diagram in Figure 2.2 compares the prevalence of the "zymotic" diseases as the main cause of death in the Crimean War with that in England. Figures in the top diagram are percentages; those in the bottom diagram are per 1,000.

Florence's logical arguments and usage of these graphics and statistics were so effective that, despite determined resistance on the part of the British Army, she managed to influence the allocation of resources for improved sanitary conditions and a series of physical improvements in military buildings. It is worth noting that half a year after she arrived at the main British hospital in Crimea, the mortality rate in the hospital had dropped from 42.7% to 2.2%.

The use made of statistics has burgeoned to the situation today, where statistics gathered on the population size and the number of houses are used to obtain information on housing needs, the population composition in terms of age groups indicates the current and future needs such as education and medical requirements, and the study of successive enumerations will yield trends in population growth and migration, both internal, such as from rural to urban and external, such as the nett increase or decrease of the population of the country as a whole.

One important feature which is emphasised when advertising the census to the public is the impact and importance of census data in the fields of business, administration and research, and for determining electoral boundaries. The location of a factory may well be determined by the distribution of the population, as well as the proximity of related industries, the ease of distribution of the products using present or planned transport

facilities. Planning for the present and future needs of the population is greatly assisted when trends in the shifting of populations from one area to another can be identified, and research into the level of prosperity of the population, including the oft-quoted 'standard of living' can be conducted using socio-economic indicators, such as the ownership of certain amenities. New transport facilities, the siting of new schools, community centres, hospitals, shopping centres, extensions to public services such as electricity, gas, water, sanitation and postal services.

General Justification for Census Questions

When planning for the needs of the population in the near future, such as education, housing and medical assistance, a mere headcount of the population is insufficient. It also necessary to know the distribution of the population, its composition in terms of sex, age and race, and changes in the population distribution and composition. Censuses generally have the following separate enumeration units: individuals, families, households, and dwellings. As a census collects data such as each person's age, sex, ethnic origin, location and income, any census item can be analysed with any other census items; for example, it is possible to obtain information on income classified by age, sex, occupation and geographic location. Thus a census provides a very detailed picture of the population at a single point in time. Area classifications are vital to census data, both for cross section and time series analyses of census trends.

A census collects two main sets of data. The first set is **demographic data.** This is the information on the population and its characteristics such as geographical distribution, age distribution, ethnic distribution and level of education. This data is comparatively constant over time, and relies heavily on consistent questions having been asked in previous censuses and in censuses carried out elsewhere in order to detect any trends.

The second set of data is **socio-economic data** which can be used to determine the level of prosperity of the population, such as the median annual income, the average number of cars per family, the ownership of boats, caravans or holiday homes. Because socio-economic data tends to change with needs of the time, it is independent of historical data. Any question which shows that the level of ownership of that particular item has reached near-saturation level should be discarded, as the data will no longer be useful as an indicator of prosperity.

When the most recent census for a country is compared with previous censuses, it is often possible to detect trends of fertility, mortality, migration from rural areas to urban areas, and migration from specific areas of a country to other areas or overseas. In New Zealand, the regional distribution of the population has become increasingly unbalanced, due to some regions consistently experiencing higher growth rates than others. The proportion of the Total Population living in the North Island has steadily increased in the North Island's favour from 52.9% in 1901 to 73.1% in 1981. The Rural-Urban drift is clearly displayed by the increase from 67.93% of the Total Population living in Urban areas in 1926 to 83.59% in 1981.

In general, between 1971 and 1981, the northern-most regions of both islands experienced the highest population growth rates. The only exceptions to this pattern were Hawkes Bay and Horowhenua in central and southern North Island, and the Clutha-Central Otago region in southern South Island. The latter regional growth can be explained by the construction of a major hydro-electricity dam, which brought a large influx of construction workers and administrators into the area. New Zealand experienced a net migration gain from other countries between 1961 and 1966, and between 1971 and 1976. Between and following these periods, net emigration has been experienced.

Present day need for census in NZ

In New Zealand, the information yielded by the censuses is used by government departments, quasi-government organisations, local authorities, banks, financial institutions, private enterprise companies, research organisations, universities, libraries and individual members of the public. The data is used to ascertain the current and future population estimates on a national and sub-national basis and, following each New Zealand census, boundaries of the general electorate are revised. In addition, the census provides important demographic and socio-economic characteristics of the population, such as the age structure, the number of households, the structure of the labour force, median incomes, characteristics of the unemployed, racial distribution, population growth, migration trends and family formation trends. Again, because detailed information is collected from all respondents, the data is available for various cross-classifications such as geographical area, age, sex and race. Appendix 2.1 lists just some of the uses made of the information gained from the New Zealand 1981 Census of Population and Dwellings.

The Census Questionnaires contain several questions which seek the same information which has been, or is being, collected by other government departments. However, these government departments are legally bound to keep their own information about individuals confidential to themselves. Moreover, information held by particular government departments is often based on differing classifications, and seldom relates to a common time period, and hence cannot be analysed in detail. For instance, in New Zealand, records of social security and war pension payments to individuals are kept by the Department of Social Welfare, and details of gross (taxable) income are collected from individuals by the Inland Revenue Department. However, as the Department of Inland Revenue does not request the date of birth of tax payers, it is not possible to use their

data to generate statistics on income distributions of those persons aged 15-19 years, 20-24 years, 25-29 years, and so on.

Prior to each New Zealand census, users of census statistics are invited to make representations to the Department of Statistics on what statistics they require from the census. The Department recognises that the questions asked in the census must be constantly reviewed and additions, alterations or deletions made to ensure that the census satisfies the needs of the users. Government departments, local authorities and universities, and members of the public are encouraged to make submissions. Chapter 4 details the criterion which any potential census question must satisfy before being considered for possible inclusion in the census, but we can note here that the census questions must be applicable to New Zealand as a whole, be of importance to the community and not be controversial or engender public resistance to the census.

A computerised information service known as INFOS (Information Network for Official Statistics) was made available to the New Zealand public in 1982. The INFOS data base provides data on many aspects of economic and social activity in New Zealand, currently ranging alphabetically from Age Estimates to Wool Price Index. INFOS allows users to analyse and tabulate data through interactive access to a dedicated mainframe computer in the Department of Statistics. Some of the facilities available to INFOS users include display and printing of data selected by users and simple graphics. More complex statistical data manipulations and high quality colour graphics are also available through linkage to the SAS System. Users can also produce their own tables through linkage to TGS (Table Generating System developed by the Department of Statistics) or TPL (Table Producing Language developed by the US Bureau of Labor Statistics), or write reports using the SCRIPT package. For those users of census data who do not have access to INFOS, ad hoc services of user requirements are available on request.

In addition to the traditional publications on New Zealand census data in standard tabular format, 1986 Census Summary Files, either on tape or through INFOS, may be purchased by members of the public. These are computer-readable files of aggregated record data, and users of these files can specify the amount of detail required. Because these summary files contain aggregated data, they preserve confidentiality of information supplied by individual respondents, and hence do not need to be subjected to the procedure of random rounding.

Until recently, users of New Zealand census data enjoyed the privilege of being supplied with the data or services such as table production or statistical analyses for free, or at worst, at a minimal charge which did not cover the costs. However, users must now pay commercial rates for the supply of data or services. While such a practice will ensure that data or statistical analyses will not be requested lightly, the cost of such services means that census data is no longer so readily available to individual members of the public, and will undoubtedly have a flow-on effect to the attitude of the public in future censuses. While individual members of the population are legally required to complete census questionnaires, they are given no remuneration for this service. Consequently, users of census data may well resent having to pay commercial rates for data which they themselves contributed towards.

Chapter 3

ENUMERATION

Introduction

The quality of the census data will only be of a high standard if the fullest possible coverage of the population is achieved and the data collected is accurate. While censuses are necessary to provide a source of detailed information necessary for administering a country, they do mean an invasion of privacy for each individual respondent. What is more, individuals are not financially compensated for their time and effort in supplying the required information.

Members of the public are more likely to be receptive to the idea of a census and thus more willing to cooperate if they realise the purpose and importance of census data, and if a guarantee is given that personal privacy will be protected. Common methods of preventing published information from being traced back to the individual respondents are to withhold small-area data which does not have a sufficiently high number of respondents, and to randomly round or mask all census data as a matter of course prior to publication.

Needless to say, the success of a census depends on the dedication of all the persons involved. The questions asked must be well thought out and thoroughly tested before being presented to the public to ensure that the wording used does elicit the desired responses. For example, the questions used should be neither ambiguous nor likely to provoke hostility. To ensure that the data collected is what was required, the questionnaires can be pretested, using a small group (sample) of respondents. Personnel involved in investigating the information given by the respondents must be carefully selected and thoroughly trained to ensure their questioning is

efficient. If it is found that one or more of the questions needs to be modified, this can be done before the actual census takes place. Ideally, the modified questionnaire should also be pretested prior to the census.

It is also important that data collected in a census are published as soon as possible, while still ensuring that the published data is accurate. Official publications which come several months or even years after a census was conducted will not receive the same attention, or be as useful, as more timely publications.

Publicity Campaigns

To achieve cooperation from the public, an extensive publicity campaign is necessary prior to the census. The purpose of such a campaign is to make everyone aware of the impending census and to explain what type of questions will be asked, why each question is being asked, and what the information will be used for. If the public relations aspect of the census is successfully carried out, then this will assist in the widest possible coverage being achieved. Experience has shown that census response rates are always higher than those achieved for surveys, and this must in part be due to the extensive publicity generated prior to a census. Undoubtedly, the knowledge that individuals are legally required to provide the required information for censuses, will also affect the response rate. Persons conducting surveys have no recourse to legal action, should individuals refuse to participate.

To be successful, the publicity campaign must reach all members of the population, which is not an easy task. In addition to advertisements and features on television, radio, newspapers, popular magazines and periodicals, special efforts should be made to reach members of ethnic minority groups by liaising with their spokespersons. Such spokespersons should

be able to identify any special assistance or specific problems which may occur.

Assistance should be made available to those who experience difficulty completing questionnaires, including the provision of interpreters when necessary. Questionnaires should be made freely available to members of the public who for some reason have not been supplied with a questionnaire.

Because most societies today are multicultural, leaflets printed in the appropriate languages need to be made freely available, to promote as full a response as possible. These leaflets can serve the dual purpose of publicising the imminent census, and explaining how the questionnaires should be completed.

Traditionally, minority groups have had a high rate of underenumeration; possibly because of language barriers or fear of officialdom. However, government funding is often allocated on the basis of the population distribution, and in such cases, any section of the population which has been undercounted will be disadvantaged.

Confidentiality of Data and Coverage

Fear of what the data will be used for is another factor which influences the attitudes of respondents. Persons living in a building where children are not permitted will naturally be loathe to indicate that they have children. Illegal immigrants will attempt to avoid being counted because they fear detection and deportation by the Government. Shortly before the 1976 New Zealand Census, the Government announced that it intended to deport illegal overstayers. The figures obtained from the 1976 Census were well short of what was expected. For instance, for the Total Population, the enumerated figure was 3,129,383, whereas population projections using previous census data and migration data had indicated that the figure

should have been in the vicinity of 3,148,774; a difference (undercount) of nearly 20,000 persons.

Persons living in condemned houses may fear eviction and possible prosecution for trespassing. Others may fear curtailment of welfare or social security payments, should the true nature of their living arrangements be revealed. All such unreported data will contribute towards an undercount of the population, and in the latter example, also misclassification of data.

All respondents must be aware of their legal obligations to the census, and they must also be assured of the confidentiality of the information gathered in the census. The New Zealand 1975 Statistics Act provides that census information cannot be disclosed in any way which could, directly or indirectly, link the figures to a specific household or person. For the 1981 and 1986 New Zealand publications, all cell values, including totals, were randomly rounded to base three using a simple random rounding procedure. This precautionary measure ensured that information could not be gleaned from tables which involved small categories, thus preserving the anonymity of the respondents, while permitting the breakdown of census data into smaller geographical areas. Summary files are also available to the public, but these are not subjected to the rounding procedure since they only contain aggregated data.

Legal Obligations

In order to make it clear precisely what information is required from every member of the public, the New Zealand 1975 Statistics Act includes a listing of the mandatory questions which must be asked at each New Zealand census and authorises the Government Statistician to obtain information relating to any topic which is considered to be in the public interest. The Act provides for a penalty of up to $250 for neglect or refusal to complete a Population and Dwellings Census questionnaire or

to answer questions or inquiries lawfully addressed by an authorised employee of the Department of Statistics. There is also a penalty, not exceeding $10 per day, if such default continues after conviction.

While the threat of legal action for nonresponse is not conducive to the supply of high quality data, it is nonetheless necessary to provide some sort of legal stricture for nonresponse, and to give enumerators the authority to collect the required data.

The problem of coverage not only includes deliberate non-response by those who are antagonistic towards the census, but also people living in areas which are isolated or difficult to access. It is not difficult to imagine the problems involved in enumerating people who live in mountainous regions such as Tibet, or even the problems encountered by enumerators in early New Zealand history, when much of the inland areas had only been partially explored, and considerable distances had to be travelled on foot.

The Statistical Bureau of China's Communique on Major Figures in the 1982 Population Census reported that the total population was 1,031,882,511. This figure included the island populations of Fuji and Taiwan provinces, and compatriots in Hong Kong and Macao. Of this total population, some 800 million live in the countryside, and development is uneven from one region to another. The mammoth task of enumerating a population of this size is awesome, and it is not surprising to learn that censuses in China are not conducted at regular intervals. In fact, the two previous national population censuses were taken in 1953 and 1964. However, mobility of the Chinese population is reported to be relatively low, and both a household registration system and a regular system for collecting population statistics have been in existence for some time.

Following the national Chinese census in 1953, a new household registration system was established, and a government decree requires every citizen to register at his of her regular place of abode and give information on sex, age, nationality, education and occupation. All births, deaths, immigration and emigration movements must also be reported to the registration office and this data has been used to produce the population figures released by the State Statistical Communiques on National Economic and Social Development.

Census Maps

For a census to be successfully conducted, it is essential to have correct geographic information for two reasons: firstly, complete coverage is facilitated by the enumerators having correct and legible maps which list every housing units in each census district; secondly, the correct geographic information permits the assignation of each housing unit and its occupants to the appropriate land area.

The three main geographic tools used in census taking are **maps**, **address reference files** and **geographic reference files**. Address reference files are used to associate addresses with their geographic locations, whereas geographic reference files catalogue the various geographic areas and define their relationships, facilitating an ordered presentation of census data by area.

There are two general categories of maps: statistical or legal boundary **"outline" maps** and **thematic maps**. Outline maps are produced to assist those who work with census data in locating the legal and statistical jurisdictions to which the data refer, and include **field maps**, which are used to guide data collection. Thematic maps are used to present the spatial distribution and relative magnitude of a given set of data; in other words, they show statistical data in pictorial format. Field maps are used to ensure that no area of land is either

omitted or duplicated during enumeration. **User maps** include maps defining meshblock/area unit boundaries and those presenting statistical data. Examples of thematic maps are given in Chapter 9.

In preparation for the mapping operation, the entire country is divided into districts which will be covered by individual enumerators. The Encyclopaedia of Statistical Sciences (1982) lists the following criteria for the delineation of these Enumeration Districts:

1. *In the interests of speed of making contact and of the simultaneity of the whole census operation, no district should be larger than can, reasonably, be covered on foot by an enumerator in 1 or 2 days.*

2. *The boundary of each district must be clearly recognisable on the ground. Each enumerator will be given a map of his or her district, and there must be no ambiguity as to where this district ends and another begins.*

3. *Each district must be completely contiguous with the other. There must be no gaps between defined districts leading to underenumeration, and no overlapping which would heighten the risk of double enumeration.*

4. *Because much of the census information is required for local as well as central government purposes, the districts must be so designed that they can be aggregated to exactly complete local administrative areas.*

The latter criteria is followed by the note that, for planning purposes, statistics are often required for small areas that do not appear to conform to local administrative area boundaries, but which are capable of being fitted into a grid coordinate system. In such cases, the data processing is facilitated if the maps make it possible to apply a grid reference to each dwelling. Coordinates which are recorded on the maps can then

transferred to the census schedule for each dwelling. The occupants of each dwelling can then be allocated, as part of the computer processing, to any required combination of grid squares.

De Facto Enumeration versus De Jure

If a resident of a particular area is in another part of the country at the time the census is taken, should she be counted as part of the local population, or should her count contribute towards the population of the area where she lives? Both approaches have their advantages and disadvantages. The **de jure** population is defined to be the normal population of a locality and the de jure census population of a particular area includes all residents of that area, wherever they may happen to be at the time of the census. Overseas residents who were temporarily in the country at the time of the census are excluded from the de jure population counts.

The de jure approach is most easily applied where the population being enumerated is a structured or close-knit society, and the head of the community or unit can provide the details of all the members of his community or unit, such as the relationship of each member, where they belong in the society and where they are located at the time of the census. Examples of such a society which spring to mind are the Maori tribes in earlier New Zealand history, and the society of rural China, where peasants live within their local village, and each village forms its own tight community.

When the society being enumerated is a highly mobile society, the population is better suited to a **de facto** approach, in which the members of the population are enumerated at the places they were on census night. However, while the de facto method is undoubtedly easier to apply to a society such as ours, care must be taken to ensure, as much as is practicable, that the snapshot image obtained is an accurate picture of the

population. For instance, in an area where the resident population is small, an overnight school trip to some other area at the time of the census may seriously affect the population count, particularly in the school children age-grouping, which, in turn, may affect the Government allocation of education funds. Conversely, the population of a rural area may be grossly inflated by the presence of military personnel on a national exercise, or people attending a "rural retreat" or some other form of camp or activity which artificially concentrates them in a certain area at census time.

The New Zealand census is a census de facto, and the questions on residence in each Personal Questionnaire are as follows:

Full address on Census Night *(Do not give PO Box or Rural Delivery Numbers)*

....

Number in street, and name of street, road, etc. *Name of suburb or rural locality (if any)* *Name of city, town or county*

Usual residential address: *(Tick box which applies)*

Same as address given above ☐ *NZ resident with no fixed residential address in NZ* ☐ *Usually resident overseas* ☐

Other fixed residential address in NZ ☐ *Specify:* *(i)* *Number in street, and name of street, road, etc* .....

(ii) *Number in street, and name of street, road, etc* .....

(iii) *Number in street, and name of street, road, etc* .....

The Australian Bureau of Statistics employs a Householder's Schedule, on the front of which the following is requested:

Signature of Person (Householder) .....

Address:

No. and street .....

Suburb, town or locality ..... *Postcode* .....

Date .....

On the inside of the Australian schedule, information is requested about all occupants of the household on Census Night, and includes the following:

Where does each person usually live?

At the address shown on the front of this form (tick box)

If elsewhere,

No. and street .....

Suburb, town and locality .....

Name of local council .....

State ..... *Postcode* .....

Country (if usual residence is overseas) .....

The Australian census is a census de facto, but the above question allows the production of population estimates on a usual-residence basis if required.

The American census is conducted on a de jure basis, and the questions on residence listed on the householder's schedule are as follows:

What is the name of each person who was living here on Tuesday, April 1, 1980, or who was staying or visiting here and had no other home?

List

- *Family members living here, including babies still in the hospital*
- *Relatives living here*
- *Lodgers or boarders living here*
- *Other persons living here*
- *College students who stay here while attending college, even of their parents live elsewhere*
- *Persons who usually live here but are temporarily away (including children in boarding school below the college level)*
- *Persons with a home elsewhere but who stay here most of the week while working*

Do not list

- *Any person away from here in the Armed Forces*
- *Any college student who stays somewhere else while attending college*
- *Any person who usually stays somewhere else most of the week while working here*
- *Any person away from here in an institution such as a home for the aged or a mental hospital*
- *Any person staying or visiting here who has a usual home elsewhere*

Enumerator-completed Questionnaires versus Respondent-completed

Another vital aspect of enumeration which must be seriously considered is by whom are the census schedules or questionnaires to be completed? In New Zealand, sub-enumerators are employed to distribute and collect the census questionnaires, but the questionnaires are completed by the respondents. Leaflets describing the census in Maori and Polynesian languages are available on request from the sub-enumerators. In many other countries, illiteracy and resistance to the census have necessitated the questionnaires being completed by the enumerators. In a country such as India, which has high levels of illiteracy and several different languages, it is apparent that the questionnaires should be completed by an enumerator who is not only fluent in the local language, but also understands the questions being asked and is sufficiently motivated to endeavour to obtain the correct responses. It cannot be stressed too often that the success of any census is inherently dependent on the quality of data collected. The most sophisticated statistical analyses cannot compensate for false or incomplete data.

Other countries, although not as greatly affected by the problems of illiteracy, have been forced to adopt the method of enumerator-completed questionnaires because the response rates when voluntary surveys were undertaken had been found

to be markedly lower than that observed for enumerator-completed censuses.

One problem inherent with enumerator-completed questionnaires is the reaction of the respondent to the enumerator. If the respondent feels overwhelmed or intimidated by the enumerator, then the quality of data supplied will most probably be seriously affected. Equally, the enumerator may be intimidated by an aggressive respondent, or may simply lack the necessary skills to obtain the required information to a sufficient depth. Respondents who are not sure what a particular question means will usually ask the enumerator for guidance, and it is very easy for an enumerator to influence the responses given.

Enumerator bias can result from an enumerator influencing the supplied responses in a systematic manner. Enumerators can also obtain data which are unreliable because of inconsistent behaviour towards respondents. Unreliable data will inflate the standard deviation of the estimates because the overall accuracy of the data will be decreased.

In order to minimise the effects which enumerators can have on census data, the following procedures should be observed when recording responses:

1. Census questions should be read exactly as they are written.

2. If a response to a question is considered to be inadequate in the sense that it does not supply the required information, any ensuing questions asked by the enumerator should be standardised and nondirective so as not to influence the eventual response.

3. In cases where respondents are asked to answer a question in their own words (as opposed to selecting one of several

options), the enumerator must record all such responses verbatim.

4. The enumerator should conduct the interview in an unbiased manner; in particular, the enumerator should not present information about himself or herself, or comment on the respondent's answers in ways which could indicate a preference for some answers over others.

If the questionnaires are completed by enumerators, then the selected persons should be trained to ensure that they achieve the best possible responses from the public, and the quality of their work should be carefully monitored. Systematic assessment of enumerators' work would also ensure that enumerators do not fabricate responses when frustrated by failure to contact interviewees, or for any other reason, such as the time required to obtain responses. The term "curb stoning" was derived from the practice of enumerators completing the form for a particular address while sitting on the curb.

Personal Privacy

Another problem is the lack of privacy when a respondent is interviewed by an enumerator or a single questionnaire is used to collect information on all members of a household. In New Zealand, prior to 1921, a single large schedule was used for each dwelling. Personal schedules were only used for those persons in boarding houses or prisons. For the 1921 census, in an effort to ensure good quality census data, the Department of Statistics adopted the policy of using household schedules for all private households and personal schedules for every individual in communal living centres such as hospitals, camps, boarding houses and ships. The personal schedules contained the same questions as the household schedules, but offered greater confidentiality because they could be completed in privacy and enclosed in a separate envelope by each respondent. This innovation proved to be tremendously successful, and since

1945, personal schedules have been completed for every individual, and a dwelling questionnaire completed for each dwelling occupied on census night. This approach is in contrast with that used in India, where an enumerator will arrive at a village, and interview each person in the presence of the other villagers who will often contribute items of information or challenge responses given by the person currently being interviewed.

If the questionnaires are self-administered, as in New Zealand, staff involved with distributing and collecting the questionnaires must endeavour to ensure that everyone receives and completes a questionnaire. Special care must be taken to ensure that persons in transit at the time of the census, or immediately before or after, are enumerated.

We must not overlook the financial constraints of a census. While most people would agree to the necessity of a regular census, few of us are keen to pay more than is necessary for this. While some methods of operation may be ideal in theory, they may not be financially viable, or, conversely, they may be cheaper financially, but produce data of inferior quality. For the last two American censuses, the Bureau of the Census has sent out questionnaires to most areas on a mailout/mailback basis. However, the cost savings achieved by postal enumeration are to some extent offset by the cost of conducting follow-up operations in an effort to reduce the nonresponse rate. The initial attempts to contact nonrespondents are made by telephone, and enumerators will make field visits if telephone contact cannot be established. In New Zealand, respondents are aware that all delivered questionnaires will be collected and checked by the local sub-enumerator, and the Department of Statistics makes every effort to ensure that everyone in the country on census night completes a Personal Questionnaire, and that a Dwelling Questionnaire is completed for every separate dwelling which is used as living accommodation on census night. Sub-enumerators make up to 2 further visits if residents of a household are not home during

the initial collection phase. If both of these visits are unsuccessful, a stamped addressed envelope is left in the letterbox, with an accompanying request that the questionnaires be posted back to the Department.

As is discussed in more detail in Chapter 6, nonresponse can occur when one or more members of a household are not enumerated, and it is often difficult to detect. However, experience of follow-up operations has shown that the data obtained from persons who were initial nonrespondents can vary greatly from the data supplied by respondents. For this reason, it is important that every effort be made to obtain data from nonrespondents. Unfortunately, such follow-up operations are expensive, and it has often been common practice to assume that data from nonrespondents would have followed the same distribution as that of respondents. Conclusions drawn from analysis of data from respondents may not apply to nonrespondents, and it is important that any published data is accompanied by an assessment of the coverage obtained, and an explanation of how the nonresponse was treated.

Another potential problem relating to census coverage is that of multiple ownership of private dwellings. For cases where an individual owns more than one house (for instance, that person may also own a holiday home), the sub-enumerator must endeavour to guard against the possibility of counting that person twice, possibly by annotating the relevant questionnaire if it is thought possible that the person may have already been enumerated elsewhere.

Timeliness of the Published Data

One way of reducing the time lag between the enumeration and publication of the data is to release unofficial figures, these being data which have not been extensively checked and verified. However, because the census data is used so

extensively in planning by businesses, researchers and governmental agencies alike, any published data will be seized upon, despite any accompanying qualifications on the accuracy of the data, and there is no guarantee that later published amendments to the unofficial figures will be used to update the earlier calculations. Detection of processing and geographic coding errors will force the revision of figures for both the total population and the population of smaller geographical areas, and it is preferable to delay publication of census data until it has been thoroughly checked. Fortunately, the time required for coding and processing the data can be drastically reduced by using well-designed questionnaires, and automated coding and editing procedures.

Another method employed to speed up the publication of data (and also to reduce the data processing costs) is to analyse a sample of the returns. Provisional census data are often produced on this basis, and the New Zealand Department of Statistics used a 10% systematic random sample to produce preliminary statistics for the 1976 and 1981 Censuses. Sampling can also be employed to select households which will be supplied with different census questionnaires; these may be contain all the questions asked in the shorter questionnaire as well as additional questions or, if the questionnaires are all to be of the same length, they may consist of alternative questions. The U.S. Bureau of the Census employed two main questionnaires in the 1980 Census. The short form contained the basic population and housing questions which were asked of all persons and housing units, and the long form contained the basic items plus questions only asked of a sample of the population and housing units. One sixth of housing units were supplied with the long form, except in smaller areas (less than 2,500 persons), where one half of the housing units were sampled in order to ensure the accuracy of small-area data.

Whenever sampling is used, the data is analysed under the assumption that the sample is representative of the population. In other words, the conclusions drawn from the sample data can

be applied to the whole population. Careful sample design can produce reliable estimates for the population, with the added bonus of reducing respondent burden, increasing the scope of questions asked, and reducing the time lag between enumeration and publication of the data. However, as for every part of the census operation, an intensive review should be conducted to evaluate the success of the procedures, to pin-point problem areas, and to examine possible methods of future improvement.

New Zealand Enumeration Procedures

(To facilitate continuity of the text of the following section, definitions of the technical terms have been restricted to the glossary at the back of the thesis.)

The New Zealand Census includes all persons who are in New Zealand at midnight on the Census night. Diplomatic personnel, persons on a vessel between New Zealand ports, tourists and other temporary visitors from overseas, fishermen and others normally resident in New Zealand, but who are temporarily working off shore, are included. A count only of New Zealand Armed Forces overseas at Census date is supplied to the Department of Statistics. However, members of the Armed Forces of foreign (non-commonwealth) nations on warships or in New Zealand military camps, and New Zealand residents who are overseas at Census date are excluded. Persons under the age of 15 years or any overseas visitors who will not work in New Zealand during their stay or who are not members of families of overseas residents who are working while visiting New Zealand need only complete Part A of the Personal Questionnaire .

The census also covers all dwellings in New Zealand at that time. Questionnaires are only completed for occupied dwellings (by the occupier or person in charge of the dwelling).

Unoccupied dwellings and dwellings under construction are enumerated, but questionnaires are not completed for them.

Appendices 3.1-3.3 give a brief background to the relative timing of various steps in the the 1986 New Zealand Census of Population and Dwellings, the organisation of the enumeration phase, the duties of the Census Supervisors and Enumerators, and the stratification of land into units for compilation of the census data. Appendix 3.4 shows the stratification of geographical areas of New Zealand into Statistical Areas for 1986 Census purposes, and Figures 3.1 and 3.2 on pages 56 and 57 display the Statistical Areas of the North and South Islands. For the purposes of the New Zealand Census, the country is divided into **Districts**, ranging from Kaitaia to Riverton. Each District is in turn divided into Sub-districts, which are further divided into **meshblocks**. For the 1986 Census, each District was controlled by a Census Supervisor and an Enumerator was appointed to each Sub-district to perform the house-to-house enumeration.

A seven digit number was used to identify each area in **New Zealand**, the first three digits identifying the District, the next two digits identifying the Sub-district and the last two digits identifying the meshblock (New Zealand Census of Population and Dwellings 1981 Enumerator's Handbook, 1980). For example, the number 006/02/09 refers to District 6, Sub-district 2, meshblock 9. To ensure that no area of land was either omitted or duplicated during the enumeration, each Enumerator was supplied with a map of his/her Census Sub-district. These field maps displayed the included meshblocks and other information, depending on the type and scale of map used. Generally, only one map was supplied for urban areas but, as necessary, Enumerators in rural areas were supplied with one map showing the whole Sub-district, with one or more larger-scale maps showing the densely populated areas.

In order to supply maps for the Supervisors and Enumerators, the Department of Statistics consulted with Chief Postmasters

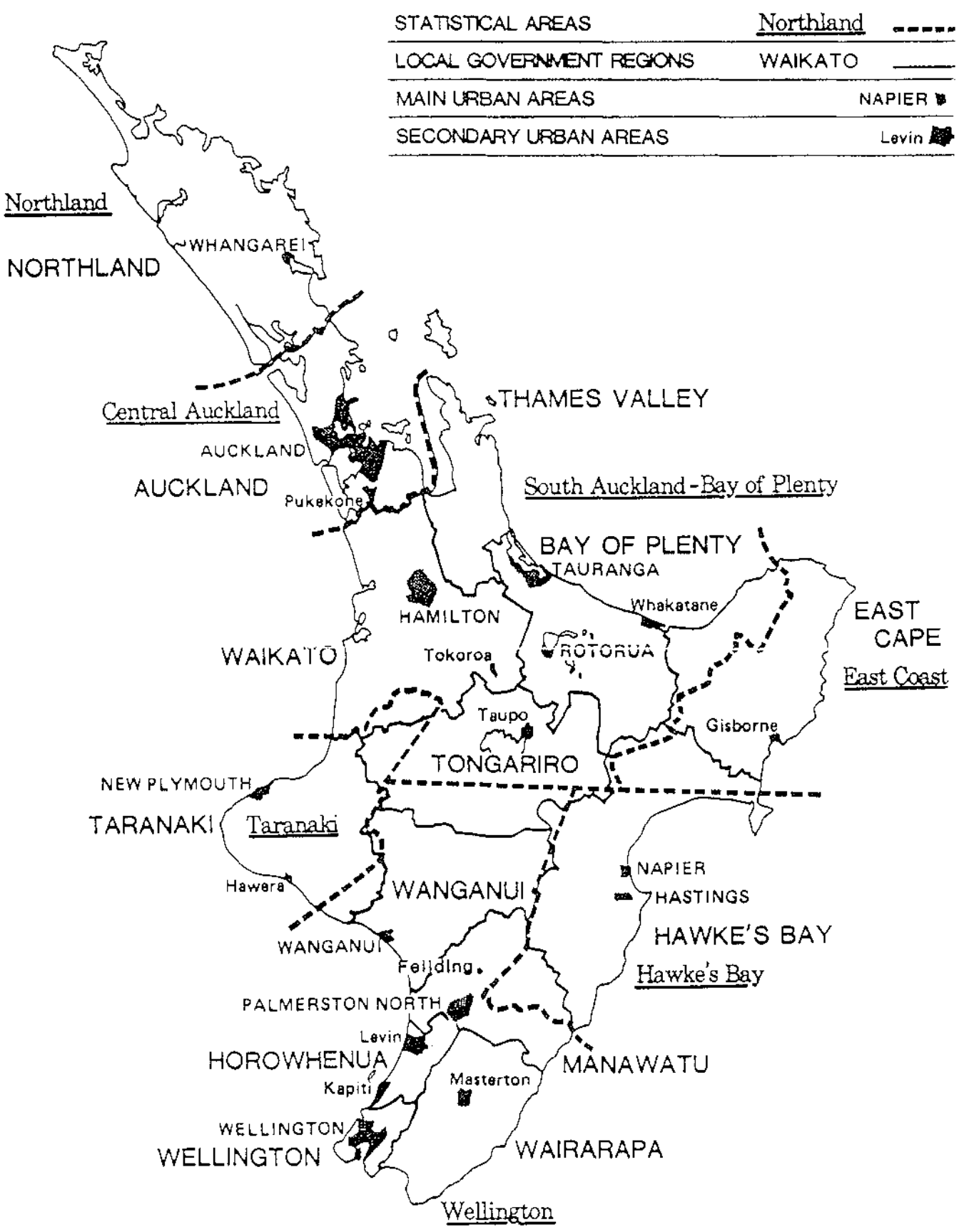


Figure 3.1 Statistical Areas of North Island, New Zealand for Census Purposes

Source: *New Zealand Census of Population and Dwellings 1986* Series B Report 24 (Department of Statistics, Wellington)

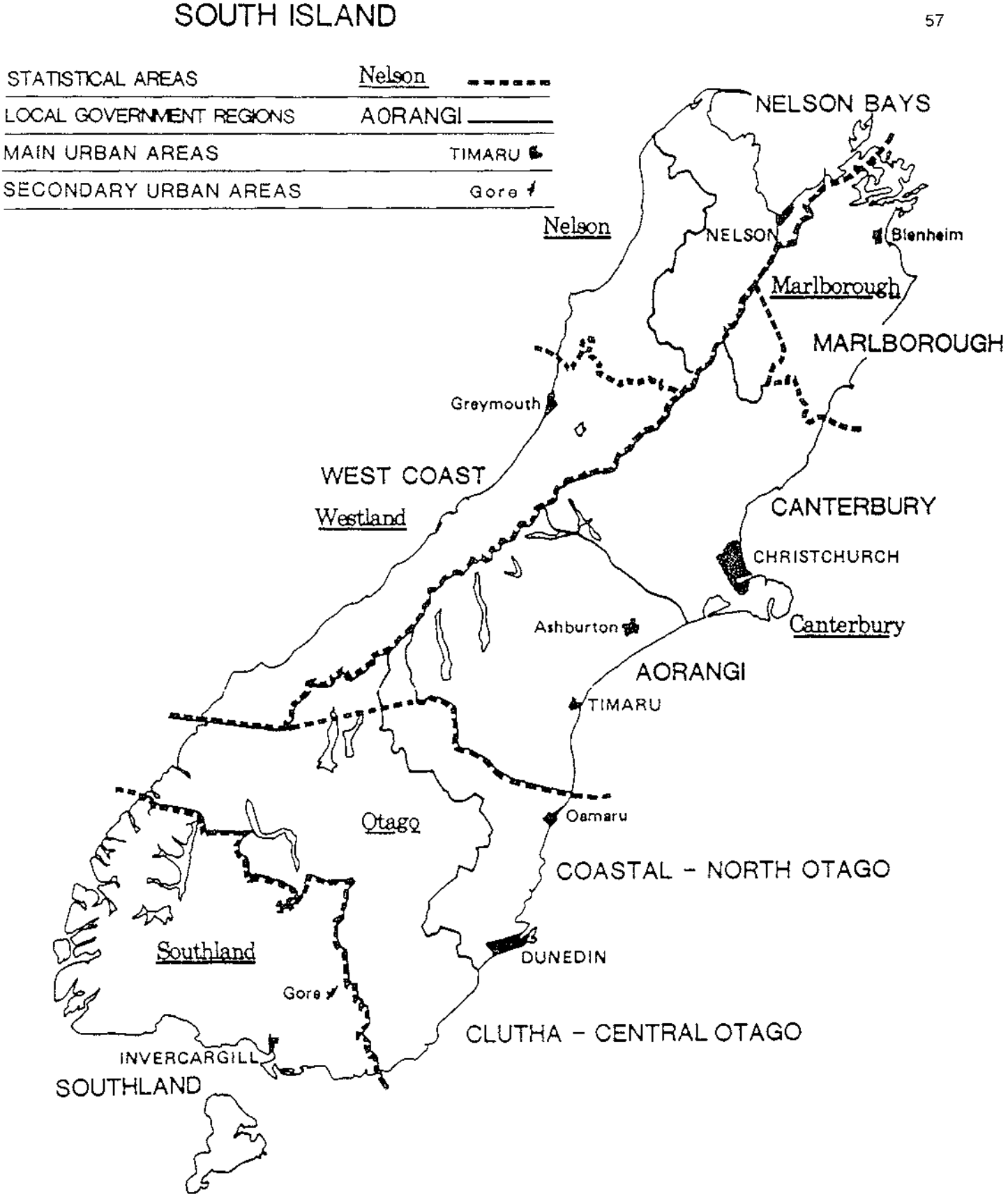


Figure 3.2 Statistical Areas of South Island, New Zealand for Census Purposes

Source: *New Zealand Census of Population and Dwellings 1986* Series B Report 24 (Department of Statistics, Wellington)

to determine the location of Census District boundaries. On advice of these boundaries, the Department of Lands and Survey prepared two sets of Enumerator maps, on receipt of which the Department of Statistics divided each District into a number of Sub-districts, taking into account the estimated population and dwelling counts for each Sub-district, reports from Supervisors and Enumerators involved in the previous census, and any other available relevant details, such as aerial photographs. Appendix 3.5 lists the guidelines used during the creation of the Sub-districts, and each Sub-district contained approximately 250-350 dwellings in urban areas, and 200-250 dwellings in rural areas. When necessary, special Sub-districts were created for large institutions such as hospitals, prisons and Army camps. The Department of Lands and Survey then used these finalised Sub-district boundaries to prepare a set of Sub-district maps for each District.

Prior to the 1986 Census, Districts were the responsibility of Enumerators and Sub-districts were administered by Sub-enumerators. Figures 3.3-3.5 on pages 59-61 display examples of a Sub-enumerator's map and aerial photographs for the 1981 Census. The aerial photographs in Figures 3.4 and 3.5 identify the meshblock boundaries depicted in Figure 3.3 and show a general view of the actual appearance of the meshblocks. Figure 3.4 shows a meshblock of urban nature, and Figure 3.5 shows a rural meshblock. The field organisation was altered for the 1986 Census, with Census Supervisors being responsible for Districts and Enumerators administering the Sub-districts. Figure 3.6 on page 62 shows a typical Urban Enumerator's map for the 1986 Census.

Prior to the actual enumeration, an extensive publicity campaign is carried out. Foremost in the public eye are the television commercials and films about the census, regional television programmes and radio programmes featuring the census. Obligatory listings of Census Supervisors' names and office addresses are published in the public notices columns of the major newspapers, and weekly magazines and trade journals

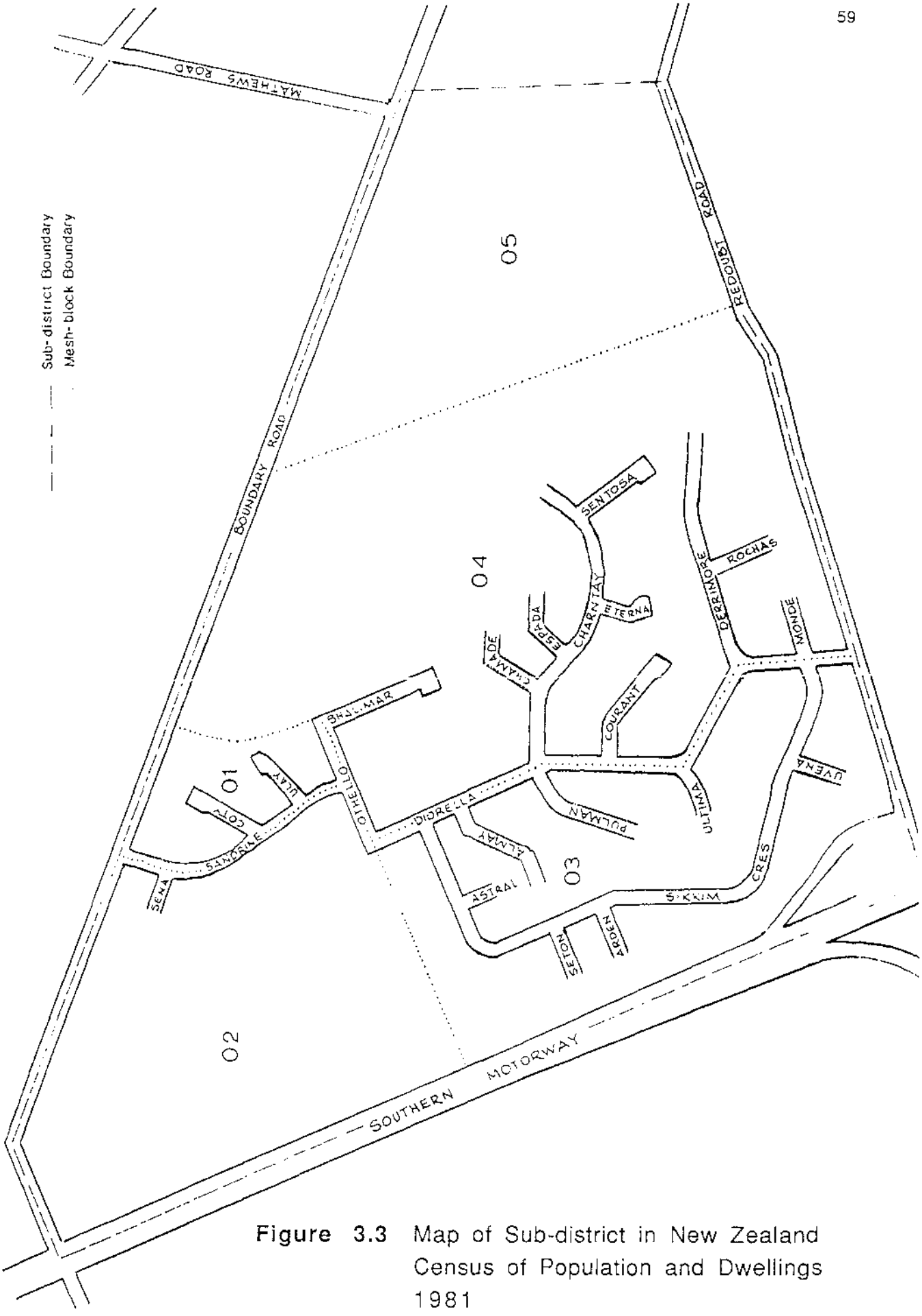


Figure 3.3 Map of Sub-district in New Zealand Census of Population and Dwellings 1981

Source: *New Zealand Census of Population and Dwellings 1981* Sub-Enumerator's Reference Manual (Department of Statistics, Wellington)

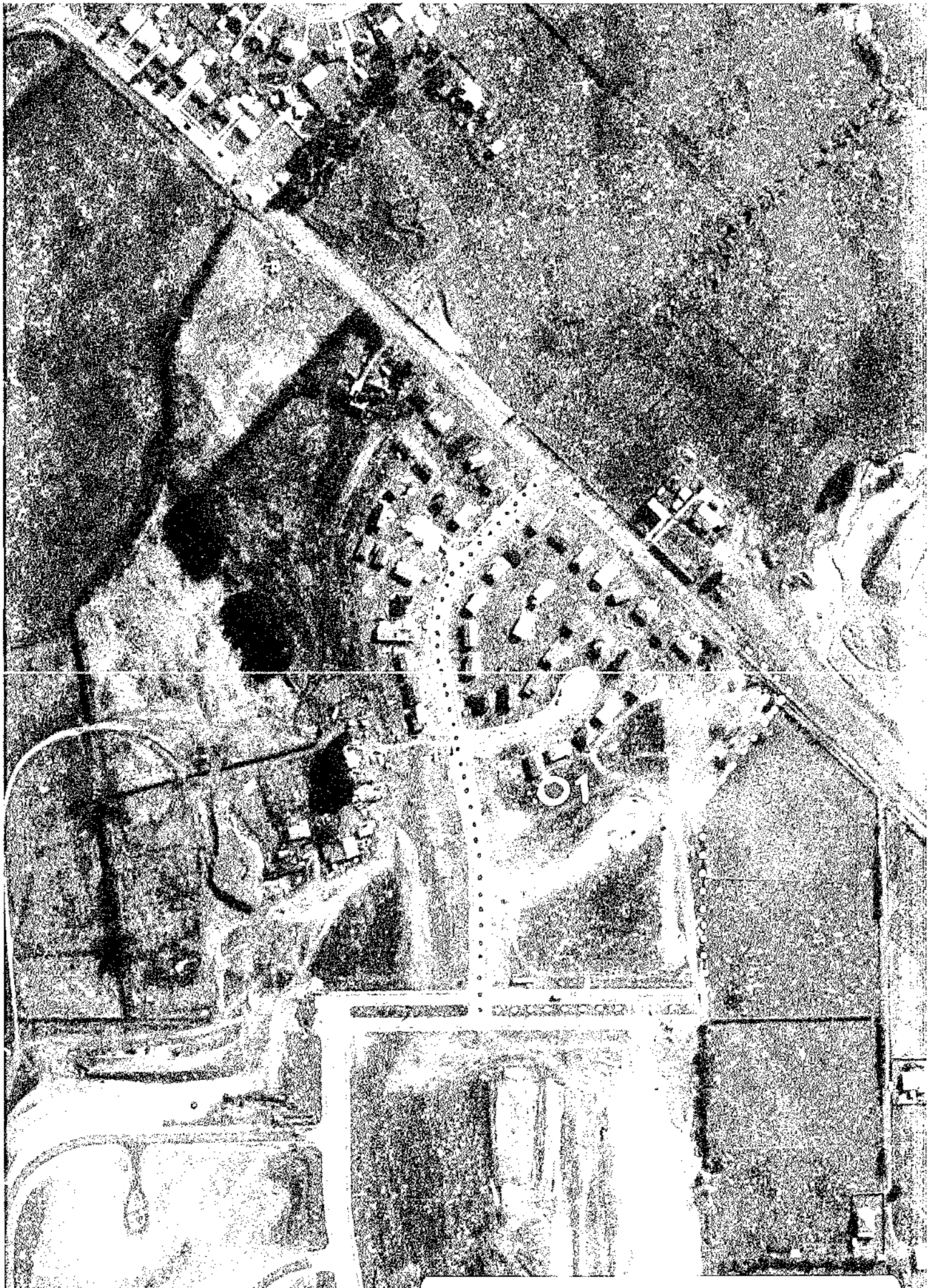


Figure 3.4 Aerial Photograph of Urban Meshblock of fig. 3.3

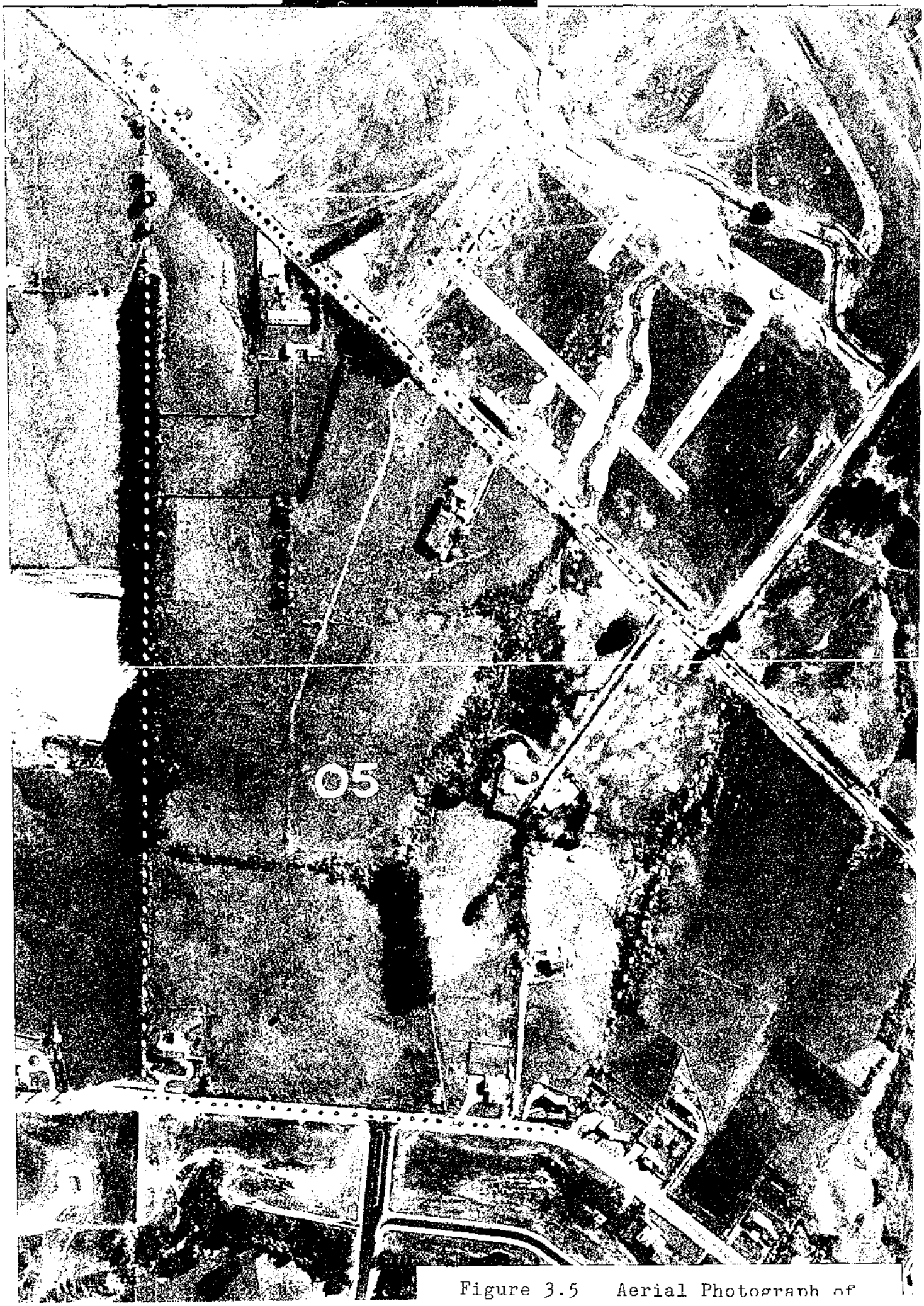


Figure 3.5 Aerial Photograph of

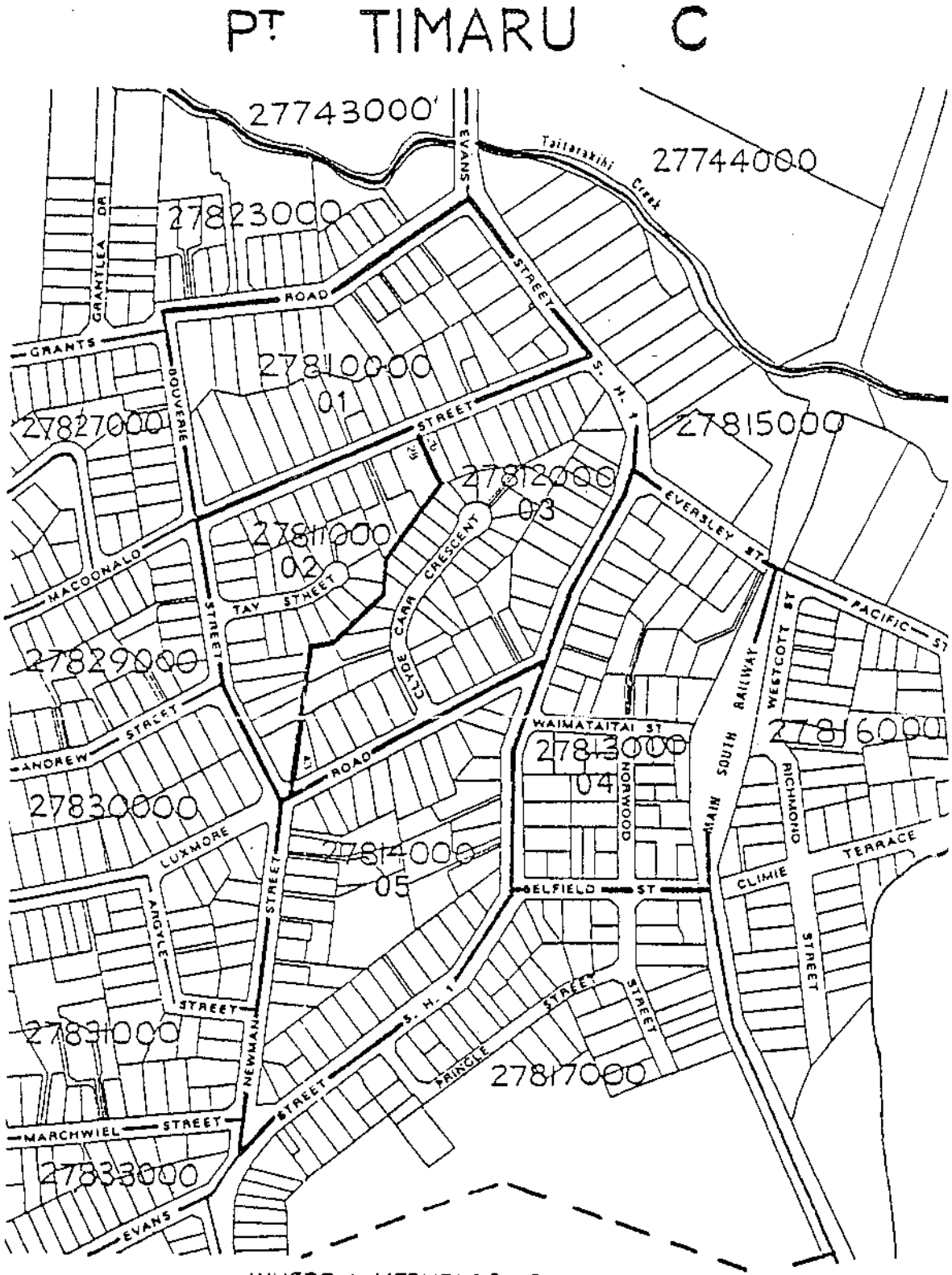


Figure 3.6 Typical Urban Enumerator's Map for New Zealand Census of Population and Dwellings 1986

166

Source: *New Zealand Census of Population and Dwellings Census '86 Enumerator's Handbook* (Population Division, Department of Statistics)

Map shows District 166, Sub-district 02. The two-digit number is the administrative meshblock number. The eight-digit number is the standard meshblock number. Hence meshblock 01 is 27810000, meshblock 02 is 27811000, meshblock 03 is 27812000.

feature articles about the census. Census Household Leaflets are distributed to all occupied homes and Language Leaflets, contained translations of the main contents of the household circular into New Zealand Maori and 5 Pacific Island languages are now also made freely available. A special Secondary Schools leaflet is distributed, which includes a competition for the best essay, poem and poster. Posters, notices, stickers and small labels are featured in prominent places in schools, universities, polytechnics, retail outlets, Post Offices and other Government departments. These leaflets contain telephone numbers of local area supervisors, and Enumerators are also made available to answer householders' queries. Telephone enquiry centres are established in Auckland, Wellington, Christchurch and Dunedin to provide an information service.

United States of America

In the United States of America, the 10-yearly censuses are conducted on a de jure basis, as each person is enumerated as an inhabitant of their usual place of residence. The US Bureau of the Census uses Householders Schedules, which request basic information on every person in the household on Census night. For the 1980 census, a mail-out, mail-back method was used for 95.5% of the population, and conventional door-to-door canvassing was used for the remainder. Response rates were then compared for the two different methods of enumeration in the post-census evaluation phase. Chapter 6 gives a more detailed discussion of the methods of evaluating census coverage used in the United States of America.

For the mail-census areas, a period of 2 weeks was allowed for the receipt, sorting and matching of questionnaires to the address register, which consisted of the mailing lists used in the distribution of the questionnaires. For those housing units from which no questionnaires were received, the enumerator attempted to contact the householders. If contact was

established, the questionnaire was collected if it had already been completed by the householder, or completed by the enumerator with the assistance of the householder. If no-one was contacted at home after 4 visits, the enumerator attempted to complete the questionnaire by talking to neighbours, landlords and building superintendents.

For the population questions, the information required was the name, sex, age, race and marital status of each person in the household, and the relationship to the household member in whose name the home was owned or rented. The housing questions were completed for both occupied and vacant housing units. If non-response housing units were vacant, the enumerators answered only the questions on housing.

For the conventional areas, postal carriers delivered unaddressed short-form questionnaires to each known housing unit, which contained instructions to the householder to complete and retain the questionnaire for collection by an enumerator. Enumerators then canvassed their assigned districts, listing the addresses of the housing units and collecting the questionnaires, completing them when necessary. Using a pre-printed sampling pattern in their address register, each enumerator also collected information from selected households on long-form questionnaires.

A supplementary American Indian Questionnaire was completed for every housing unit on Indian reservations that was designated as a short-form housing unit and had at least one American Indian, Eskimo or Aleut occupant. These supplementary questionnaires were also used in mail-census districts which contained Indian reservations and in parts of Oklahoma that were formerly Indian reservations, but excluded those in urbanised areas.

US Automated Geographic Support System

The U.S. Bureau of Census prepared the maps, address reference files and geographic reference files for the 1980 Census in independent manual operations. For the 1990 Census, the Bureau intends to have the geographic support system automated. This automated support system has been named **TIGER**; an acronym standing for **T**opologically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing system. It will have a large data base, built on a computer-readable or "digital" map, which will cover the entire area to be included in the census.

All the geographic information that was previously produced separately in the 1980 Census will be integrated into one file, which will relate all mappable features, address information and geographic codes. Thus any change to one item will be reflected in all contexts simultaneously, and the geographic products and services for the 1990 Census will be produced from one consistent data base. Computer-driven plotting devices will be used to generate high-quality maps that match the geographic areas in the Bureau's tabulations. The Bureau aims to have the TIGER system operational for use in the dress-rehearsal censuses in 1988.

To prevent duplication of cost and effort, the Census Bureau and the U.S. Geological Survey have entered into a cooperative agreement for the creation of the initial cartographic database of the contiguous 48 states and the District of Columbia. For Alaska, Hawaii, Puerto Rice and the outlying areas, the Census Bureau is "digitising" information from printed maps on a point-to-point basis.

Each record in the TIGER file will be for a geographic feature, such as a road, street, railroad, waterway or political boundary. The record will contain the name and type of the feature; the coordinate values for each intersection point along the feature, such as the latitude and longitude for the point where one street intersects another; the range of addresses located

between intersection points for streets and roads, in addition to the post office and zip code associated with each address range; and, for each feature segment, the codes for all geographic areas in which the segment is located.

India

In India, the Census comes under the jurisdiction of the Home Office, and is taken once every ten years. The Census Commissioner belongs to the Indian Administration Service. The Census is run by the States under the direction of the Union, with officers in charge being drawn from the Indian Administration Service.

The 1981 Census of India was conducted on an extended de facto basis. Persons who were away from their normal place of residence throughout the enumeration period, 9-28 February with a reference date of 1 March 1981, were enumerated at their actual location during this period.

India has no permanent listing of houses. Prior to the census, in an operation called house listing, an attempt was made to locate and identify all places that were occupied or were likely to be occupied by people. The houselists were also used to allocate census staff to different areas. During house listing, data was collected on the purpose for which each house was used, the number of persons normally residing in each household, and the numbers of totally blind, totally crippled or dumb persons in each household. Details of entrepreneurial activities being conducted in the houses and in the open were also recorded in a so-called enterprise list as part of an economic census.

Each supervisor estimated the number of persons in his or her area, and is allocated between 6 and 12 enumerators, who are usually primary school teachers. India has 17 major languages, and a high level of illiteracy. Because of widespread

illiteracy, the canvasser method of enumeration was used, where the enumerators interviewed the households and completed the household schedules and individual slips. The problems of so many different major languages in India are compounded by the multitude of local dialects, many of which are not static. Undoubtedly, misinterpretations of questions and responses will have occurred during the enumeration phase. On Census night, the enumerators continued their work until all areas have been covered. In certain inaccessible areas such as the snowbound regions of Jammu and Kashmir, enumeration was carried out nonsynchronously as weather conditions permitted.

The primary school teachers, predominantly female, are given holidays and a small remuneration for their work. Unfortunately, because the enumeration duties must be conducted in the evenings, many primary teachers earn extra money tutoring and incur a financial loss, since they cannot continue tutoring during the enumeration period. The enumeration system apparently works well in rural areas, as the teachers live in the community and know the other residents. In urban areas, municipal workers are co-opted as enumerators because many female teachers are reluctant to canvass at night. The Post-Census check estimated the omission rate as being around 1.8% and duplication 2.9%; the duplication estimates being 2.7% for rural areas and 3.5% for urban areas. Apparently, there is minimal antagonism towards the census, as the public are enthusiastic about being involved and like to feel that they have been entered on an official form. In country areas, neighbours gather around and help the person being interviewed. The fact that illegal immigrants openly admit admit to being from Bangladesh or Pakistan supports the assertion that people believe that confidentiality of the census data will be honoured.

The household schedule consisted of 2 parts, the first containing questions relating to the head of the household (such as religion, membership of scheduled caste or scheduled tribe), the language mainly spoken in the household, the ownership of

the house, the predominant construction of the house and the materials used, the facilities available, the number of living rooms occupied by the household, the number of married couples usually living there and ownership or tenure of the land. The second part of the household schedule required a listing of the members of the household, their sex, age, marital status and their relationship to the head of the household.

The individual slip which was to be completed for every member of the household also consisted of 2 parts, the first part consisting of cultural and economic questions. The second part was a sample slip which was used for a sample of 20% of the enumerator blocks, and contains questions on fertility and migration. A special form, called the degree holders' and technical personnel card, was also distributed to the 20% sample of enumerator blocks to obtain data on technical and professional personnel.

Australia

The Australian Census of 1981 was conducted on a de facto basis. Collectors delivered a householder's schedule to every household in their census districts prior to census night, making personal contact wherever possible, and later collected and checked the forms.

The householder's schedule contained comprehensive questions on all persons in the household on census night, including the age, sex, marital status, gross income, religious denomination, occupation and, for each woman in the household, the number of babies born to her. The schedule also contained questions on the dwelling, including the number and types of rooms, whether the dwelling was owned or rented, and the number of registered motor vehicles owned or used by members of the household.

The collectors also recorded the type of dwelling (whether a separate, semidetached or terrace house, one of a block of

flats, a caravan, houseboat, etc), the material of the external walls of the dwelling, and if the dwelling was unoccupied, the reason why.

Special envelopes were made available to those persons who did not wish to have their census forms examined by the collector. Special collectors (usually persons in authority at the non-private dwelling) were employed to distribute personal census forms to persons in non-private dwellings. Each person to whom a form was delivered was listed on a special form which was then used to obtain a summary of persons enumerated in non-private dwellings. Similar procedures were used for persons in transit on census night on ships, long-distance trains or buses. These persons were allocated to a special census district designated "migratory" within the respective state. After the enumeration, schedules from absent parents were not collated with the schedules received from the rest of the family. Aboriginal communities living in their traditional lifestyle were enumerated by aboriginal interviewers.

West Germany

In West Germany, population censuses were conducted in 1950, 1960 and 1970, and the most recent census was originally planned for 1980. However, conflict between the federal government, political groups and communes over the funding of the census forced its postponement, initially for three years. Much of the resistance was generated for political reasons. Pacifists and armament opponents advocated a *"population census boycott for ecology and peace"*, arguing that the Government intended *"to force the citizen to provide information, while itself wanted to withhold vital information from the citizen"* (Schroeren, 1982). These groups hoped to force the Federal Government to provide exact information on the planned locations for the stationing of intermediate-range missiles. Other groups joined the movement against the population census, some hoping to block the missile

deployment, others objecting to the invasion of privacy and fearing that the statistics supplied could not be safeguarded against abuse, particularly in light of the great advance in the technical development of automated data processing.

A broadly designed press campaign against the population census, including biased presentations containing allegations effected irreparable damage. Some presentations contained incorrect or unsubstantiated allegations, and advocates of the population census were not given sufficient opportunity to present their case. It was also unfortunate that the final publicity activities of the Government agencies responsible for the census operation coincided with the final phase of the election campaign for the tenth German Bundestag (Federal Parliament). Inevitably, discussions concerning the Population Census became an issue in the political campaign, and several political candidates publicly stated that they would withhold information being sought in the census.

Public opinion polls taken before the census showed that both opponents and advocates of the population census were concerned that confidentiality of the data may not always be maintained, and that the activities of persons boycotting the data and either refusing to supply information or giving false information would bring into question the usefulness of the census results, and hence the high financial expenditure must also be queried. The opinion polls showed marked resistance to the census, particularly among young people and those with higher levels of education. A great number of complaints were lodged to make void the 1983 Population Census Law, which was promulgated on 25 March 1982, as unconstitutional in that the rights guaranteed by the Constitution were being violated. On 13 April 1983, the Federal Constitutional Court finally granted a preliminary injunction suspending the operation of the census, pending judgement on the main issue.

On 15 December 1983, the decision of the Federal Constitutional Court on the complaints against the 1983 Census

Law confirmed the methods used to produce official statistics, stating that statistics were of essential importance for a modern industrialised country; that the obligation of the citizens to answer questions for statistical purposes was in conformity with the Basic Law (constitution). The court also stated that it was inadmissible to use for other purposes the personal data originally collected for statistics. Further opposition led to a ruling in December 1984 by the Federal Constitutional Court that parts of the legislation covering the census were unconstitutional. A modified law was approved in 1985 by all the parties in the *Bundestag* except the Green Party (Keesing's Record of World Events, 1987). The census was finally undertaken by the federal government on 25 May 1987, but a concerted campaign led by the Greens to boycott or otherwise sabotage the operation has cast doubt on the validity of the census results.

Opponents of the census were unconvinced that the information given would remain confidential, despite an official undertaking that names and addresses of respondents would be removed within a few weeks of the enumeration. A shortage of volunteers forced the conscription of Army personnel as enumerators, and those refusing to complete the census forms were liable for fines of up to DM 10,000. On May 16 approximately 20,000 people protested in various West German towns against the census, and on several occasions census takers were robbed of the questionnaires which they were delivering.

Although the legal decision was in favour of the population census, irreparable damage has undoubtedly been done to the public relations side of the census, and it will be well worth examining the follow-on effect to the response rate and to the quality of the census data over the next 10-20 years, both in Germany and in countries elsewhere, particularly those close to Germany.

Chapter 4

QUESTIONNAIRE DESIGN

Introduction

The census has the potential to collect a wide range of information, thus providing a vehicle for linkage of both diverse and related data, all of which has been collected on a singular time frame. Since the census is currently the only vehicle for collecting data which will be used to produce a snapshot picture of the population at the time of the enumeration, it is desirable to make the most efficient use possible of potential census questions.

To obtain the maximum amount of information, the questions asked should be as comprehensive as possible, but with the important and rather limiting restriction that the questionnaire can still be simply and quickly completed by the respondents. As discussed in Chapter 3, the attitude of the respondents to the questions asked may well affect their responses to the questions. If the responses are incomplete, inaccurate or omitted, then no matter how sophisticated the techniques used, the quality of the census data can never be improved.

Questionnaire Content

Keeping in mind the necessity of high quality data, the number of questions asked must be restricted, so that the questionnaires are not too long. It must always be remembered by those planning and conducting a census that the success of such an operation depends not only on the expertise and techniques employed, but, perhaps even more importantly, on the quality of the data which is collected. Every effort must be made to ensure cooperation from the public, and at times,

additional information may have to be sacrificed in order to achieve a high level of response and good quality data. In recent British censuses, questions seeking information on income and ethnicity have been excluded in an effort to avoid accusations of intrusion of privacy.

Each question should be easily answered, without the respondent frequently having to refer to personal documentation or invoke memory recall. If a question is not easily understood, or is ambiguous, then, with the best will in the world, the respondent can only supply the information he thinks is required. Frustration or confusion with a question may well result in no response being given at all, and every effort must be made to rephrase or eliminate those questions which may cause difficulty. An ideal method of identifying troublesome questions is to have a selected sample of persons complete the questionnaires at some stage before the actual Census. The subjects' responses to each question can be examined in detail, as well as their responses to the questionnaire as a whole. This procedure is often referred to as **pretesting** or **pilot testing**, and if it is believed that a particular sub-group of the population (such as the elderly, or a minority racial group) may have problems understanding the questionnaire, then the sample may be deliberately selected so that the particular sub-group in question is heavily represented. In other words, the sample will not be a typical cross-section of the population, but will have a higher than normal proportion of the sub-group in question. Chapter 5 deals with the topics of pilot testing, pretesting, field testing and dress rehearsals.

The wording and content of census questionnaires has altered over the years, to keep with the current terminology and trends. In earlier New Zealand censuses, questions were asked about matters such as the number of domestic servants employed, whether the household had a piped water supply, the total numbers, by sex, of fowls, ducks, geese and turkeys, the number of beehives and the total poundage of honey and beeswax

produced during the year. Classification of occupations included Gentlemen, Milliners and Straw Bonnet Makers, Wheelwrights, Millwrights, Tallow Chandlers, Tinmen, Gingerbeermakers, Coopers, Cutlers and Gunsmiths. Nowadays, it would be inappropriate to use such terms.

In a similar vein, questions asked about amenities must be constantly reviewed if such questions are to be used as indicators of prosperity. When it is apparent that a certain amenity is possessed by virtually all households (or virtually none), then the data supplied will no longer be useful as an indicator of prosperity, and any questions relating to that amenity should be discontinued, or replaced by questions about a different amenity. However, it is unfortunate that changes in the content and the wording of questions raise difficulties in making comparisons of responses over time. Changes in classification procedures mean that adjustments must be made to the data to obtain estimates of the numbers in "common" categories before comparisons are made. Two examples of changed classifications in New Zealand censuses are given in the following section on New Zealand History. Further examples of questions in New Zealand censuses which have changed over the years are listed in Appendix 4.1.

However, continuity of questions asked in successive censuses is also of importance, particularly to social scientists. Insight into future trends can often be gained by studying both current and historical data. Such linkage of data depends on the same, or very similar, questions being asked in a regular, periodic cycle, and any changes in the definitions of response categories need to be taken into account when comparing current and historical data. Chapters 6 and 7 discuss methods used to evaluate census coverage and to ensure, as much as possible, that the data obtained is accurate. One technique employed is to compare current census counts with estimated counts, which were obtained using prior census data which has been adjusted for migration, births and deaths. Any significant discrepancies between the current census counts and the

anticipated figures are investigated, in an effort to resolve the discrepancies. Comparison of current and historical data will hampered by changes in the definitions of response categories, such as age groups, or ethnic origin.

Sample Survey versus Census

As censuses are generally taken only every 5-10 years (intercensal periods vary from country to country), the questions asked must not elicit data which may vary from season to season, or become rapidly outdated. Data which is subject to seasonal fluctuations should be obtained using sample surveys taken during successive seasons, and over a period of several years, so that **time series analysis** can be used to determine the overall trend of the data, as well as estimating the average seasonal fluctuations.

Data which dates rapidly is not suitable for inclusion in censuses, due to the length of time required to process and analyse census data. Again, sample surveys will be more efficient, since the data can be collected, processed and analysed in a shorter time frame.

Since censuses target the entire population, the questions asked must be of sufficient importance to warrant inclusion in the census, they must be in the public interest, and they should apply to respondents throughout the country. If a question is only relevant to a certain region, or to a small proportion of the population or dwellings, then a separate survey covering the region or the population group in question should be conducted. Moreover, if a question could be asked in a separate sample survey, it is often more efficient to do so, rather than including it in a census questionnaire.

It is perhaps not generally understood that although a census entails complete coverage of the population in question, it does not necessarily provide more accurate statistics than a sample

survey. Because a sample survey does not cover the entire population, it is often possible to devote more resources to following-up nonresponses to the survey. For instance, it may be economically feasible to make several attempts to contact persons who were not at home, or declined to answer the questionnaire. Provided the sample selected for the survey is representative of the population in question (in other words, the **target population**), the data collected should present a true image of the responses which would have been obtained, had the entire population participated in the survey.

Once the target population has been identified, sampling methods such as **random sampling** or **stratified random sampling** are usually employed to obtain samples for surveys. The technique of randomly selecting the sample should ensure that every member of the population has an equal chance of being selected in the sample, and thus the sample should be representative of the population. **Cluster sampling** can also be used, but care must be taken to ensure that the savings achieved in terms of time and traveling expenses are not offset by a reduced number of independent observations.

However, while it is usually desirable to restrict the census questions to those for which there is no alternative method of obtaining the desired information, it is often difficult to collate responses from many smaller surveys. A cleverly designed census can contain sufficient questions to relate the desired information. In contrast to the New Zealand approach of including questions in a census only if the information cannot be obtained from surveys, the Australian Bureau of Statistics gives priority to information which is not readily available from other sources. For instance, if information being sought could be obtained from a large survey, but no such survey existed, then the Bureau would consider the Census to be an appropriate method of obtaining the information. It uses the census to provide opportunities for cross-classification with other dwelling and household information. The US Bureau

of the Census uses its censuses to provide a sampling frame for 'follow-on' surveys.

Reading Level

Attention must also be paid to the reading level of the questionnaire. For instance, should a 10 year old child be able to read and understand the questionnaire, or should the questionnaire be aimed at the 15 year age group and above? Educational researchers recommend that the questions should be designed so that they can be read and understood by children in the 12-14 year age group. Since the target of the census is the entire population, if a 12-14 year old child can comprehend and respond to the questions asked, then virtually all members of the population who are aged 15 years and above can cope with the questionnaires. Particular groups of the populations to keep in mind when designing the questionnaire are the elderly, who often feel intimidated by such forms, and minority racial groups, which may contain many members who have not achieved a good grasp of the English language. As mentioned in Chapter 3, for the more recent censuses in New Zealand, explanatory leaflets about the census are printed in Maori and Pacific Island languages, in addition to those printed in English.

The questions should be unambiguous, but where it is decided that guidance notes are necessary, then they should be sufficiently comprehensive to cover the respondents' needs. The guidance notes must not make the questionnaire too heavy or voluminous, nor should they be shortened to the extent that the instructions become impersonal or ambiguous. As an illustration, in the 1981 New Zealand Census, a question asking each woman for the number of children born had the accompanying instruction *'If a male tick box'*. The intention was that male respondents should place a tick in the box for coding purposes. However, because the instruction was so terse, it was misunderstood by many women, who ticked the box because one or more of their children was male.

Ideally, as far as the respondents are concerned, the guidance notes for each question should be included as part of the question, or at least be situated very close to the question. Of course, this may conflict with the desires of those involved with collecting, checking and coding the data, as these persons will want to be able to quickly and easily identify the responses. From their point of view, it would be preferable to have all the guidance notes contained in a separate section, which could later be detached and discarded. It is of interest to note that for the 1981 and 1986 New Zealand Censuses, the guidance notes for the Personal Questionnaire and the Dwelling Questionnaire were presented on separate, throw-away sheets. Only the completed questionnaires are collected, thus minimising the weight and volume of paper handled by the sub-enumerators, enumerators and coders.

Questionnaire Design

Another important factor influencing the response to the questions asked is the design of the questionnaire. The questionnaire must be produced in a format which is easy to read, understand and complete. Answer sections or answer boxes need to be clearly indicated, and the questionnaire should be include explanatory notes which clearly and concisely explain how the questionnaire should be completed. It is also helpful if the end of each question is easily identified (a common practice is to insert several blank lines between questions), and it is generally more convenient for the respondent if questions on a related theme are grouped together.

The technique of pretesting the questionnaires can be used to gauge the response rate to varying layouts of the questions, differing placements of answer boxes. Where several answer categories are supplied, and respondents are asked to tick the appropriate answer box, will the order in which the categories are listed influence the response? For example, if a question

asks for the respondent's religion, is the first category ticked the most frequently simply because it is listed first? Or is the last category ticked frequently because the respondents didn't like the other options? Different versions of the questionnaire can be tested, with each version showing a different ordering of the categories, and the number of times each category is ticked compared for the different versions. Different sizes and colours of the questionnaires, and varying levels of difficulty of the wording of the questions and guidance notes can also be tested, to determine which produce the best responses. It is astonishing what a difference in response can be achieved by presenting colourful, attractively designed questionnaires with a well-thought out layout.

By now it may have become apparent to the reader that there is, and always will be, a conflict of demands on questionnaire design. Users of census data will want a list of questions which will be as comprehensive as possible, whereas respondents will naturally prefer a short, simple, and easily completed form. Persons involved with checking and processing the data will want the response boxes to be aligned where they can be easily and quickly checked and coded, whereas respondents will want each answer box to be as close as possible to the corresponding question. The print size of the questionnaire must be sufficiently large so that the questions can be easily read, even by those with poor eyesight, and it also important to have an attractive layout of the questionnaire. However, if two questionnaire sheets are employed to improve the layout by increasing the answer space, grouping related questions together, and increasing the coverage of guidance notes, then the increased volume and weight of the questionnaire will create difficulties for those persons involved in checking and coding the data, and if the data is to captured using electronics, then the cost of machine reading time will be doubled.

Data Capture

The census questionnaire should also be designed with data capture in mind. Since a census operation is inherently a very expensive exercise in terms of manpower and technical equipment, any overall savings which can be effected by automation of the data capture, editing, imputation and dissemination phases should be carefully considered. Such cost savings can be achieved by developing a questionnaire which facilitates the data capture stage, and it is common practice in many countries to ask the respondents to tick the appropriate answer box in response to questions. This method not only reduces the time required to convert the responses into machine-readable format, prior to the data editing and subsequent stages, but should also please the respondent by reducing the time required to complete the questionnaire. The U.S. Bureau of the Census has elected to fully automate the entire data processing operation of the 1980 and 1990 Censuses. A brief description of the procedures used is given later in this chapter.

Summary

To summarise, in order to promote a good response, the questionnaire must be readable, simple and easily completed. If the responses given are of poor quality, then none of the methods of **editing**, **imputation**, which is the replacement of missing data by data inferred or derived from elsewhere in the questionnaire or by some other means, and **statistical manipulation** techniques which are now available can fully compensate for the poor quality data. Hence the level of non-response must be kept as low as possible by ensuring that the concepts, layout, appearance, wording and response options of questions are kept as simple as possible, and it may also prove beneficial to introduce the format of the questionnaire to the public in advance of the census, so that the public know what to expect, and how to complete the questionnaires.

New Zealand History

As mentioned in Chapter 1, different schedules were employed to enumerate the European and Maori populations prior to 1951. The enumerators recorded the information on the Maoris in special books supplied to the Native Department. From 1951 onwards, no special Maori schedules have been produced, and standard questionnaires have been used for all members of the New Zealand population. These questionnaires have been completed by the respondents.

In the early censuses, a single large schedule was employed for each household. In the 1926 Census, a personal schedule, intended for use in hotels, camps, ships and other communal living centres, was introduced. Three types of schedules were used: dwelling, family, and personal. Individuals were permitted to seal their personal schedule in an envelope which even the enumerator collecting the schedules was not permitted to open. Because of the convenience and privacy it offered, it became so popular that it was universally employed as a personal schedule and from the 1945 Census onwards, only dwelling and personal schedules were employed. Another major change in the history of New Zealand census questionnaire design was the employment of separate, disposable guidance notes in the 1981 and 1986 Censuses. Separate personal questionnaires were supplied for each member of the household, rather than the previous booklet of forms per household.

To make the questionnaire easy to complete, and also to assist the census staff involved with editing and coding the questionnaires, the majority of questions in the Personal and Dwelling Questionnaires for the 1986 New Zealand Census were completed by ticking the appropriate answer boxes. Apart from the name and address of respondent, the Personal Questionnaire contained 27 questions, 20 of which were completed in this manner. Of the remaining 7 questions, 3 required a numerical answer (the number of years at current address, date of birth,

and number of hours worked the previous week). The remaining 4 questions related to the respondent's present occupation, and could be answered with brief statements. A 19% reduction in the costs of the New Zealand 1986 Census over that of 1981 was achieved, and much of this saving can be attributed to the new questionnaire design and the usage of **computer assisted coding (CAC)**, which is described in Chapter 7.

Apart from the name of the occupier, the address, and the distance from the nearest Post Office if the dwelling was in a rural locality, the Dwelling Questionnaire contained 9 questions. Six of these were to be completed by ticking the appropriate boxes, 2 required numerical answers, and the remaining question required the listing of particular details of those persons who were temporarily way from the dwelling on Census night, but who usually lived at the dwelling. The details requested were the name, sex, age, marital status, relationship to the occupier, and the address at Census night. This information was then used to check that persons temporarily absent from their usual place of residence were enumerated in the census, and did complete a Personal Questionnaire.

The New Zealand Personal and Dwelling Census Questionnaires contain four types of questions: mandatory, standard, cyclic and ad hoc (A Guide to the Content of Questionnaires for the Census of Population and Dwellings of New Zealand, 1982). **Mandatory questions** are those which must be asked in every census, according to the Statistics Act 1975. The mandatory questions cover the following subjects: name, address, sex, and ethnic origin of every occupant of the dwelling and particulars of the dwelling as to location, number of rooms, ownership, and number of occupants on Census Night. **Standard questions** are those which the Department of Statistics regards as being necessary for inclusion in every census, because they are considered to be of national value. Examples of topics covered by standard questions are occupation, employment status, number of children born, and rent paid.

Cyclic questions are those questions which the Department of Statistics regards as being unnecessary for inclusion in every census, but which need to be asked at regular intervals, such as every 10 years. **Ad hoc questions** are generally included in a census questionnaire on a one-time basis, in order to provide national and sub-national statistics for some particular purpose. These questions usually usually emanate from sources outside the Department of Statistics, either as a result of an invitation by the department, or by an organisation's own initiative.

Appendix 4.1 gives a brief history of questions asked at each Census of Population and Dwellings, and Appendix 2.3 lists relevant sections of the Statistics Act 1975. Specimens of the Personal Questionnaire and the Dwelling Questionnaire used in the 1980 Census are contained in Appendix 4.2.

Two examples of changed classifications in New Zealand censuses are as follows:

For the 1971 New Zealand Census, persons of part-European and part-Pacific Island Polynesian ancestry were classified as Pacific Island Polynesian, even if they were of less than half Pacific Island Polynesian descent. For the 1981 Census, when a person had descended from two or more groups, he or she was allocated to the predominant group, this being the component group which has the highest proportion of the respective component groups. In the cases where no group predominated, the allocation was determined according to the following priority order: New Zealand Maori, Pacific Island Polynesian, other ethnic groups (excluding European), and European. Thus, while New Zealand Maoris were not affected by the change in the classification procedure, on the basis of the 1981 Census classification of ethnic origin, the 1971 Census statistics provided an overcount of the number of Pacific Island Polynesians and an undercount of the number of Europeans.

The second example concerns the definition of 'separated'. For pre-1981 censuses, the question on marital separation only included those persons who were 'legally separated' from a spouse. The 1981 definition of 'separated' was augmented to include all those persons 'permanently separated' from a spouse. Consequently, the numbers of separated persons doubled between the 1976 and 1981 censuses, masking any underlying trend in marital separations.

The Department of Statistics has adopted the policy of inviting submissions for topics or questions to be included, amended or deleted from the questionnaires for each census. The contents of the questionnaires for the New Zealand census are tightly constrained by the legal requirements as stipulated in the Statistics Act (1975), by the demands of users of census data, and by the requirement that the questionnaires be kept to a manageable length.

Before being included on the shortlist of questions which will receive further consideration, the questions must satisfy the following ten criteria (A Guide to the Content of Questionnaires for the Census of Population and Dwellings New Zealand, 1982): geographic coverage, universality, nature of information, census versus survey, simplicity, classifiability, confidentiality, justification, importance, and public interest.

The questions must be applicable to all geographic areas of New Zealand, they must apply to a reasonably large proportion of persons or dwellings, and the census must be the only practicable way of obtaining the desired information. The information sought must not be controversial, of a prying nature, sensitive, subjective, nor specialised. Questions asked must not seek opinions nor attitudes. The data collected must not vary from season to season, nor date rapidly. The census must be the best method of collecting the desired information, and a question will not be accepted for inclusion in a census if statistics of sufficient accuracy could be obtained from a sample survey. Unless the census is the only means of obtaining

the preliminary information, the census cannot be used to facilitate an exploratory investigation into a particular topic, prior to more detailed research being undertaken.

Because the census questionnaires are completed by the respondents, they must be easily understood, and should not involve excessive memory recall or thought. Research has repeatedly shown that people's memories are very fickle. Questions which require a search of documents to obtain the correct answers should not be included in a census questionnaire, since persons are enumerated at their Census-night addresses, and documents and other records are often not readily accessible. Multi-part questions are also undesirable, as they are not likely to achieve a good response.

All the responses to the questionnaires are coded prior to analysis, and to facilitate the coding stage, the responses should be of a simple numerical form, or be answerable by ticking the appropriate pre-coded box(es). Written responses are time-consuming for both the respondent and the coding staff and are also prone to error, since clerical translation to numerical form requires reference to code-lists.

To ensure the confidentiality of responses, questions are not included which will require subsequent identification with respondents' names or addresses or the release of identified data outside the Department of Statistics. No information contained in a census questionnaire will be made available by the department to any organisation, including any other government department, or to any person (other than an employee of the department) in any form that would allow identification with the person who supplied the information, or with that person's dwelling.

Before a question will be considered for inclusion in the census, it must be shown that there is a need for the information required, and the usage to which it will be put must be justified. For every question included in a census, the

Department of Statistics provides the public with a brief explanation of why the information is required, and for what specific purpose it will be used.

The suggested questions must be sufficiently important to warrant a legal penalty for nonresponse. No census questions can be answered on a voluntary or optional basis as, by law, all relevant census questions must be answered.

Finally, applicants must be able to show that it is in the public interest to obtain the information.

Australia

The Australian census employs a householder's schedule to collect the personal details of all persons present in the household on census night, as well as particulars on the dwelling. Information on the type of dwelling structure, the material of the outer walls and the reason why the dwelling unit is unoccupied, if applicable, is collected by the enumerator. Any persons in the household who do not wish to have the Householder record their personal details may complete a Personal Form and enclose it in a sealed envelope. The provision of separate Personal Forms when requested allows for the collection of the required information while still protecting the privacy of individuals. It is a prudent provision, for the 1981 Census included such questions as *"Has the person been married more than once?" "How old was each person when they left school?" "What is the gross income (including pensions and/or allowances) that each person usually receives each week from all sources?"* and *"For each woman, how many babies has she ever had?"*

India

As mentioned in Chapter 3, because of the low level of literacy in India, the schedules are completed by the enumerators. An

Enterprise List was employed to collect information on activities or enterprises conducted in each village or town, and the 1981 Census requested a description, classification and information on the nature of each operation, the type of ownership, the social group of the owner, the type of power or fuel used for the activity and the number of persons usually working daily. The other schedules completed by enumerators in the 1981 Census were the Household Schedule, which was in two parts, as was the Individual Slip. The Universal Slip was used to canvass all areas, while the Sample Slip was limited to a selected 20% sample of the enumeration blocks in the state.

The first part of the Household Schedule included questions on the religion of the head of the household, whether he was a member of a scheduled caste or tribe, the name of his caste or tribe, the language mainly spoken in the household, the predominant construction material of the house (whether the walls were mainly of grass, leaves or bamboo reeds; or mud; or brick; or wood; or metal sheets; or stone or concrete cement; or ekra; or other type), the number of married couples usually living in the household, and two questions on the facilities available in the household: what was the source of drinking water and whether the supply was within or outside the premises. The second part of the Household Schedule requested a listing of the members of the household and specification of sex, age, marital status and the relationship to the head of the household.

Data collected in the Universal Individual Slip included membership of a scheduled caste or scheduled tribe, mother tongue, specification of 2 other languages known, the name of the caste or tribe, and whether literate or illiterate.

The Sample Individual Slip asked such questions as the reasons for migration from the last place of residence, ever-married women were asked their age at time of marriage and the number of children surviving at present as well as the number of children ever born alive. Currently married women were

asked whether any child was born alive during the last 12 months.

United States of America

For the 1980 Census, the US Bureau of the Census employed two versions of a Householder's Schedule. The short schedule was supplied to all households, whereas the longer schedule was only supplied to a $16^{2}/_{3}$% sample of housing units in larger areas and a 50% sample in governmental units estimated to have under 2,500 people. The universal schedule consisted of the following questions, which were to be asked of all members of a household: relationship to the householder, sex, age and month and year of birth, marital status, and ethnic origin. The listed categories for ethnic origin were as follows: *White; Black or Negro: Japanese; Chinese; Filipino; Korean; Vietnamese; Indian(Amer.); Asian Indian; Hawaiian; Guamanian; Samoan; Eskimo; Aleut; Other (to be specified)* and a further question asked whether of Spanish or Hispanic origin or descent, and the listed categories were: *Non-Spanish/Hispanic; Mexican, Mexican-American, Chicano; Puerto Rican; Cuban; Other Spanish/Hispanic.* The Spanish/Hispanic section of the American population has always proved difficult to enumerate, and the latter question was included in order to obtain an estimate of the coverage obtained in the Census.

Also included in the universal schedule were questions on the number of living quarters (occupied and vacant) at that address, whether entry to living quarters was directly from the outside or through a common or public hall or through living quarters occupied by another household, whether living quarters had complete plumbing facilities (hot and cold water, flush toilet and bathtub or shower), whether the apartment or house was part of a condominium, and if a one-family house, whether the property was at least 10 acres and whether any part of it was used as a commercial establishment or medical office. If neither case applied and the house or condominium was owned

by the family, an estimate of the value of the property was requested.

The longer schedule, which was only supplied to sample households, included the following highly personal questions: *"What is the highest grade (or year) of regular school this person has ever attended?" "Did this person finish the highest grade (or year) attended?" "If this person is a female, how many babies has she ever had, not counting still births?" "Has this person been married more than once?" "(If married more than once) Did the first marriage end because of the death of the husband (or wife)?" "What was this person's total income in 1979?"* and a further question requiring a detailed breakdown of the sources of income and specification of the annual amount into the following categories: *Wages, salary, bonuses, tips; Own nonfarm business, partnership, or professional practice; Own farm; Interest, dividends, royalties or net rental income; Social Security or Railroad Retirement; Supplemental Security, Aid to Families with Dependent Children, or other public assistance or public welfare payments; Unemployment compensation, veteran's payments, pensions, alimony or child support, or any other sources of income received regularly.*

Automated Data Capture in the 1980 and 1990 US Censuses of Population and Housing

In order to reduce the time lag between the collection and publication of data for the 1980 and 1990 American censuses, the US Census Bureau has elected to automate the coding, editing and imputation phases.

Most questions on the census questionnaires have been designed so that the respondent can answer by filling in the circle next to the appropriate multiple-choice answer. The remaining questions require written answers, which are then numerically coded. For the 1980 Census, after the district offices had completed the enumeration, edit and follow-up phases of the operation, the census questionnaires were processed at one of

three processing centres. The handwritten entries were manually coded, and then high-speed cameras were used to film the questionnaires. The raw film was then processed into rolls of microfilm, which were then read onto computer tape using a **F**ilm **O**ptical **S**ensing **D**evice for **I**nput to **C**omputer. This method of converting microfilm into machine-readable format is known as **FOSDIC**, and the overall process of data capture is referred to as **FACT**, which stands for FOSDIC and **A**utomated **C**amera **T**echnology.

For the 1990 Census, the Bureau has elected to process the questionnaires concurrently, and will establish 11 processing centres, where FOSDIC will again be used for data capture, with supplementary keying of some of the handwritten data and specially designed software will be used to assign the appropriate computer-readable form. Incomplete handwritten responses or those which are uncodable by computer will be handled by clerical "referral" units. There will be two distinct types of district offices; for district offices in certain high-density areas, the processing centres will receive the questionnaires, perform automated check-in using laser sorters, and immediately convert the questionnaires to computer-readable form, thereby performing an automated edit of the questionnaires. The remaining district offices will use wands attached to microcomputers to perform automated check-in, conduct clerical edits for completeness and then send on the questionnaires to a processing centre for data conversion. This reduction in the amount of manual work should not only achieve savings in terms of time and money, but also improved quality of data.

Chapter 5

PILOT TESTING

Introduction

A full-scale census could produce poor results unless all phases of the operation have been rigorously tested. Reviews of previous publicity campaigns and post-censal evaluation studies may suggest new methods, and several small-scale trials of proposed enumeration and coding procedures can be conducted. It is possible to test out more than one version of a questionnaire, to determine which one produces better results. For example, the layout of the questions on the page could be varied, as well as the actual wording of questions. Usually, a sample of the population is selected, and these persons are requested to complete the questionnaires. Once the questionnaires have been completed, they can be used to test the coding procedures. It is also possible to conduct in-depth interviews with the respondents, to ascertain their reactions to each question.

Terms such as **pilot test** and **pretest** have no agreed meaning, and are often used interchangeably. We will adhere to the following interpretation of their terms adopted by the New Zealand Department of Statistics:

Pilot testing is the overall process of developing the census questionnaires and the field operations, such as delivery, collection and initial checking of the questionnaires.

Pretesting is the process of developing specific aspects of the census, such as particular questions asked or procedures used.

Field testing is a general term used to cover the overall exercise of testing concepts, questionnaire wording and layout, procedures, etc. of a survey or census.

Dress rehearsal is the final field test(s) with the questionnaire and the processing of the data thus collected to test the coding, editing, tabulation and publication procedures

Pilot Testing

A glossary of sampling and census terms is included at the back of this thesis, but whatever sampling method is selected for the survey or pilot test, care must be taken to ensure that the population actually sampled is representative of the target population. Often people will decline to answer questions because they believe that their responses may not please the interviewer, and it is essential that time is spent encouraging such persons to participate in the survey. Such nonrespondents would typically have given different responses to those supplied by respondents, and if they cannot be convinced to participate in pilot tests, then the conclusions drawn from the pilot tests will not apply to the target population.

Internationally, most statistical organisations now pilot test census questionnaires as a matter of course. In addition to testing the actual questionnaires, a number of other areas such as field procedures, instructions for field staff (enumerators and sub-enumerators), and the development of coding for the questionnaire data can be investigated using pilot tests. Obviously, it is essential to ensure that the planned theoretical approach to the whole census operation does in fact work. Not only must every stage of the administrative cycle be proven to work when tested by itself, but all stages, when fitted together, must mesh and work according to plan.

During the questionnaire development process, questions which may prove difficult for particular groups of the community,

such as the elderly, high school students, or various minority ethnic groups, are usually tested informally or in unstructured interviews, and the sample of respondents will generally be taken from these groups. Such a sample is **'purposive'**, as it is designed with the emphasis on the range of the relevant characteristics of the population being included, rather than the distribution matching that of the wider population. That is to say, rather than obtaining a true cross-section of the population for the sample, the emphasis is on ensuring that the particular community groups which may have difficulty with the questions are well represented in the sample. Pilot tests conducted for New Zealand censuses have included samples from Papakura City, which is south of Auckland and has a high concentration of Polynesian people, and Dunedin, which has a high proportion of elderly persons. It is hoped that using such a heavily biased sample will ensure that the questions are adequately tested, and that any problems with the questions will be identified for further study.

One version of a formal pilot test, sometimes referred to as **frame of reference probing**, involves the construction of a sample of households, members of which complete the questionnaires, and are then interviewed by trained staff who will probe the respondent's answers in an effort to yield more detailed information about how or why each respondent answered each particular question. Respondents are asked additional questions to determine their understanding of the intent of specific questions, and the meaning of their replies to those questions. The probing questions can be of an ad hoc nature determined by the interviewer at the time of the interview, or structured follow-up questions which would have been determined before the interviews. In the former case, it is usually preferable for researchers or questionnaire designers to conduct the interviews, as they can gain first-hand knowledge of how their questions were interpreted and they should know how far to pursue the reasons why a particular response was given.

Frame of reference probing is especially useful in highlighting questions which are ambiguous. Because such questions could be interpreted in more than one way, they would yield misleading statistical analyses, and maximum advantage should be taken of the opportunity to eliminate as much ambiguity as possible before the actual survey or census. During frame of reference probing, the interviewer will often be asked to explain questions which are unclear or ambiguous to the respondent. Interviewers must bear in mind that it is the respondent's interpretation of the question that is being sought, and that the interviewer's interpretation of the type of answer required would undoubtedly increase the chance of **interviewer bias**, which is the influence of the interviewer on the answer supplied by the respondent.

The need for extensive pilot testing may be reduced in some areas, and valuable information for future question design provided, by employing reverse record links (used to access records such as dates of birth and addresses from previous censuses) and links with other sources, such as other surveys, birth and death record forms and migration records. The proposed questionnaires can be assessed in terms of the accuracy of the responses: Did the respondents recall the events? Were the events reported with reasonable accuracy? Was there any antagonism to reporting the events? Which topics were reported the most accurately? Which topics were poorly reported? What is the optimum reporting period for asking respondents to recall events?

The two basic approaches to record checks are called **reverse record checks** and **forward record checks**. For the reverse record check, a sample of persons with the desired characteristics or experiences is drawn from administrative records and as many of these as possible are interviewed and their responses to the questionnaire are compared to data on similar topics in the administrative records. The forward record check is conducted by selecting a sample of completed survey questionnaires, matching as many as possible of these

questionnaires to administrative records and comparing the answers to similar topics.

The reverse record check has two important advantages when used in conjunction with the development of questionnaires. The first advantage is reduced cost. Forward tracing can prove to be expensive and time consuming. If a full-scale survey of particular characteristics or behaviour is being considered, the reverse record check provides a sample of persons possessing these characteristics or exhibiting the desired behaviour at a reasonable cost, particularly when the variable of interest occurs so rarely in the population that screening the population would be prohibitively expensive. The second advantage is that the sample of persons known to possess the desired attributes can provide clues as to the proper way to phrase questions, or even to test whether the desired information can be usefully collected by a sample survey.

The reverse record check method has the disadvantage that important aspects of the topic in question may not be covered by administrative records. For instance, if a reverse record check were employed to investigate the accuracy of reporting sources of income, only sources which were in the administrative record system could be included in the record check. Any income sources sampled from the administrative records which were not identified by a respondent could be easily identified, but no conclusions could be drawn about income sources reported to interviewers which were not in the administrative sample. In such a case, a forward record check may prove to be of more value.

Further checks on the quality of data can be made by comparing such items as the reported age with the reported year of birth, and the reported temporary visitors with reported absentees.

America uses reverse record checks with its current Population Survey, and links census records to income tax, health and

social security records. Canada uses reverse record checks with its Labour Force Survey.

Pretesting

Many methods can be employed to test questionnaires, ranging from informal pretests to formal pilot tests. Informal pretests involve a relatively small number of interviews held in the kind of setting chosen for the final survey, such as at home or at work, and as such are not designed for rigorous statistical evaluation. They may involve 'trying out' questions on colleagues and friends to eliminate the most obvious errors in questionnaire layout and instructions, to narrow the range of alternative question wordings to be used in formal testing, and to determine whether the questionnaire is too long or whether any questions cause refusal problems (because respondents are unwilling or unable to answer the question).

Informal pretests offer a relatively cheap and simple method of reducing nonsampling error; according to newspaper reports, senior officials of the U.S. Census Bureau have even tried out census schedules on their children. However, since informal tests are normally restricted to a few interviews, and produce subjective information from interviewers and respondents, the inferences drawn from the results may not apply to the population as a whole.

Whether informal or formal pretests are employed, it is essential that the interviewers minimise the risk of interviewer bias by reading survey questions exactly as they are written, by standardised and non-directive probing of answers which do not meet question objectives, by recording verbatim the respondents' answers (and not interpreting what they thought were the *intended* answers), and by not presenting information about themselves or commenting on the respondents' answers in ways which could indicate a preference for some versions of answers over others.

Early pretests conducted by overseas agencies tended to be full schedules, but experience has shown that these limited the amount of probing that could be done on each question, and slowed down processing of the responses. It is preferable to restrict the number of questions probed with any one respondent, and interview a larger number of persons in order to collect sufficient data, rather than risk the respondents' tolerance level by probing more questions with fewer respondents. The total length of the interview should be restricted to a short time span; probably no more than 15-30 minutes.

Unless it is possible to establish the validity or reliability of questions using some other method, each topic selected for possible inclusion in the census should be thoroughly pretested, and a report prepared, covering each question which has been pretested. It is essential that the pretest effectively answers the following questions: Was the question interpreted and answered reliably? What was the respondent reaction to the topic? In other words, was the question acceptable? Did it engender antagonism, annoyance, hostility, or offence? Did different question wordings, layouts, or levels of questioning elicit different information such as varying responses? Was it ambiguous - could it be interpreted in more than one way? Was it always answered? Did the question involve memory recall or recourse to personal papers? Any questions which require respondents to supply information which relates to prior events or earlier matters are less likely to be answered accurately. The respondent will often provide the first response which comes to mind, without bothering to verify its accuracy, or may even completely omit the question.

Pretesting the questionnaires and the accompanying explanatory notes before the actual census allows an assessment of the public acceptance of each question. Careful analysis of the results of such pretesting, and the necessary amendments to the questionnaires, should achieve maximum response in the actual census. It is essential that everyone completes a

census questionnaire, and if the design of the questionnaire or of the questions themselves bemuse or antagonise the respondents involved in a pretest of the questionnaires, then pretesting will provide the opportunity to examine the cause of the problems, and to conduct further pretests as necessary.

Responses to a questionnaire can be affected by the presence and order of the included questions. If only one pretest is planned, the entire questionnaire should be covered in each of the interviews. A series of tests will provide the opportunity to devote one or more of the informal tests to potentially troublesome sections of the questionnaire, using purposive samples if required. Differing versions of the question wordings or order can also be tested. However, such a series of tests must conclude with the entire questionnaire being tested to observe how the sections work together.

Pretests can be used to determine whether questionnaires and accompanying instructions need to be made available in languages other than English (or the official language of the country in question), to determine the reliabity of census information when questionnaires are completed by persons without a good grasp of English, to identify which language groups require special assistance, and to specify what form of help they need.

Another method of pretesting is by group interviews, using a family environment, a work environment, a classroom or a community group setting. The group could focus on specific topics, or range more generally over the census topics. Group testing could be exploited to observe how much influence a group setting has on the interpretation and response to the questions. For instance, if a child is answering a question, and asks an adult what the question means, is he/she told the response to give, or is the question explained, and the child allowed to supply his/her own answer? How much effect does peer pressure have on the responses? Is the opinion of one or

two of the more vocal members of the group adopted by other group members?

After the interviews, information can be gleaned from the interviewers through discussions or written evaluations. Because discussions, usually referred to as **interviewer debriefings**, involve verbal exchanges among a group of interviewers and a discussion leader (preferably a survey designer), qualitative rather than quantitative feedback can be gained on problems in the structure or wording of a questionnaire. Debriefing conducted during the developmental stages, may result in revised question wording and response categories, identification of sensitive questions, improved flow of the questionnaire, and estimation of the respondents' ability to answer the survey questions. A number of versions of a question can be tested if sessions are held frequently and changes in the questionnaire are implemented throughout the testing period.

Debriefing conducted at the end of a survey will assist in the evaluation of the performance of the questionnaire during the survey, the analysis of the results, and can also yield recommendations of changes in future surveys. While debriefings will not allow specification of the extent of problems detected in the questionnaire, they do have the potential of uncovering problems that were not anticipated by survey designers.

Information about the effect the attitudes and behaviours of both respondents and interviewers may have on the responses can be obtained, either after a particular stage of a survey has been completed, or after the entire survey has been completed. Such an approach, often referred to as **structured post-interview evaluation** or **rating**, can be used to improve a questionnaire draft, to improve future surveys, or to yield information about the kinds of errors that may have been introduced during the data collection process. The evaluation can be in the form of self-administered questionnaires or

personal interviews. The former method is more commonly used, as it is cheaper and usually more practical. Evaluation questionnaires can be used to obtain assessment of the extent of cooperation of the respondents, assessment of the questionnaire, or even self-assessment from the interviewers. Interviewers can be asked to evaluate each separate interview, or to make overall assessments of their respondents as a whole. Because interviewer bias can not only affect the quality of data provided by the respondents, but also the frequency of responses to the questions, it is important that potential sources of bias be thoroughly investigated. Attitudes of the interviewers about the objectives and value of the survey and the attitudes of the interviewers towards the respondents will often be transmitted to (or inferred by) the respondents, as will any inhibitions the interviewers may have about any questions in the survey. Such factors can influence not only the quality of data received, but also the frequency of nonresponse.

It is aiso possibie to use structured evaluations by interviewers in conjunction with similar evaluations by respondents, enabling comparison of the perceptions of the interviewers and the respondents. Another variation is to use evaluations in conjunction with interviewer debriefing, thus obtaining quantitative responses to every question for every interviewer, as well as providing the opportunity for interviewers to disclose problems which had not been anticipated by the survey designers.

Field Testing

The effect of paying enumerators "lump sums" rather than allowing extra expenses for difficult follow-ups can be examined in field tests, as can the effect of the sub-enumerator's personality on the rejection rate (nonresponse rate). Field coverage, procedures employed, training methods, staff manuals and other aspects of quality control should also be tested. In the United Kingdom, Sub-enumerators provide the

details about the type of dwelling. Field tests could be employed to determine whether it is more economic in terms of data quality to have trained staff, rather than respondents, providing such information.

Further points of interest are:

(i) Should separate dwelling and personal questionnaires continue, or should a single household form be used? The New Zealand Personal Questionnaire was originally introduced in 1926 as a schedule to be used in non-private dwellings. However, it proved to be so popular that it replaced the Family Schedule, and it is unlikely that New Zealanders would willingly accept the re-introduction of a Householder's Questionnaire.

(ii) What paper colour is most suitable for questionnaires? What colour ink should be used? As was discussed in Chapter 4, questionnaires which have an attractive layout and are printed on coloured paper will elicit a better response. To assist those members of the population with poor eyesight, and to make the questionnaires easy to read for everyone, the wording of the questions must stand out clearly on the pages. When considering the colours of the paper and ink, it is also important to select colours which are suitable for the method of data capture which will be used when coding and editing the responses. Will Optical Character Recognition be used? What colour combinations of paper and ink produce the best results for this method?

(iii) Should the census questionnaires be posted back, rather than being collected by subenumerators? The US Bureau of the Census uses a mail-out/mail-back system for the distribution and collection of their census questionnaires in all but the traditionally "hard-to-enumerate" areas. Undoubtedly, this will have achieved a reduction in the proportion of the budget spent on enumerators' wages, but

will require a more intensive follow-up programme for those households which do not post back their questionnaires. The effect of such a programme will vary from country to country, according to factors such as the size of the population, the mobility of the population, the proportion of the population in the 15-35 year age group, the number of different languages and dialects spoken and the success of the public relations programme. One cannot assume that what works well in one country will automatically be successful elsewhere.

(iv) Would follow-on surveys be possible or acceptable to the public? Would such surveys involve violation of confidentiality or have some other adverse effect on the response rate to future censuses? Because the number of questions and the length of the census questionnaire must be limited to avoid adversely affecting the public's response to the census, it is not always possible to collect as much information as was originally desired. In such cases, it would be nice, from the point of view of those interested in the responses to particular questions, to follow-up the census with additional questions. However, the public are assured that all responses will be treated confidentially, and targeting persons who supply particular response(s) for follow-up surveys may arouse antagonism and distrust.

(v) Should regular surveys be introduced in addition to, or instead of the census? The frequency of the census could be reduced if regular surveys are conducted during the inter-censal periods. However, to prevent over-exposure of the public to such surveys, it would be necessary to limit their number and frequency. As experience has shown that censuses achieve a higher response rate than surveys, it would be necessary to ensure that such surveys incorporated intensive follow-up campaigns in order to reduce the proportion of nonresponses.

Pretests could be performed specifically on these aspects, or they could be incorporated into other pretests or the dress rehearsal(s).

Dress Rehearsal

Once all the pretests are completed, the decisions are made as to which topics will be included in the dress rehearsal, and which of the pretested options will be used. The dress rehearsal is in effect a final pilot test, used mainly to solve remaining areas of doubt. It is important that all question wordings in the dress rehearsal have been adequately field tested, and only very minor cosmetic changes should be made to census questions as a result of the dress rehearsal.

The dress rehearsal is intended to test out the questionnaire(s) as a whole, to ensure that all forms employed are of an acceptable length and structure, the instructions are understood, and that the order of questions do not lead to misunderstandings. It could also be used to test the field operations, and to permit the editing system for the full census to be developed (or 'fine-tuned' if it is already well advanced). A limited number of variations could be used as part of the dress rehearsal, such as posting back the questionnaires, placing instructions in different locations on the questionnaire or an attached sheet, the provision of a check-list at the end of the questionnaire, and so on. All questions should have already been exhaustively tested, and only those which have been successfully through the pretesting procedure should be included in the dress rehearsal. Only very minor changes should be permitted from the dress rehearsal to the full census.

New Zealand History

The first public pilot test of proposed questions for a New Zealand census was conducted in May 1975, in 3 areas:

Hamilton, Feilding and Dunedin (New Zealand Census of Population and Dwellings 1971 Internal Migration). The size of the pilot test was small because the decision to conduct a pilot test was made late in the planning cycle. Fifty households were sampled in Hamilton, 25 in Fielding and 50 in Dunedin. The respondents were asked to complete the proposed Census Questionnaires as well as a Comment Sheet, which contained the following questions:

(a) Did the wording of any questions confuse you?

(b) Did the layout of any of the questions confuse you?

(c) Did you have any trouble in following the notes or instructions to any questions?

(d) Overall, how easy (or hard) did you find the questionnaire to follow?

(e) If you had the choice, would you object to having to answer any of the Pilot-test questions in the Census proper?

(f) General comments (if any).

The respondents were asked to supply comments for questions (a) to (c), and to select one category from the following list for question (d): *'very hard', 'hard', 'not too hard or not too easy' , 'easy'* . Question (e) required a listing of the question numbers.

Field staff collecting the questionnaires had been instructed not to obtain any missing information on the Questionnaires, but to ascertain the reason for each omission and to record it on the Questionnaire. This procedure ensured that while the field staff had no influence on the answers supplied by the respondents, they provided additional information on the suitability of each question.

For the analysis of the pilot test, any questions which had a relatively high non-response rate were deemed to have been wrongly interpreted by the respondents as 'Not Applicable' or were found to be objectionable or difficult to follow and/or understand. The analysis revealed that questions which were split into several parts were far more likely to be partially or completely omitted, questions on nationality, how New Zealand citizenship was acquired and employment intentions were very

poorly answered, and that the notes for guidance (which were printed on the back of the pilot test questionnaires) were rarely followed.

As a result of the pilot test, multi-part questions in the actual 1976 Census were kept to a minimum, questions on nationality, New Zealand citizenship and employment intentions were completely omitted, and the notes for guidance were printed alongside the questions.

The pilot test effectively only tested the responses to the questions. The comment sheet sought the overall response to the Questionnaire, but the detail collected was very shallow. In the author's opinion, the pilot test would have been more effective if in-depth probing of the responses to selected questions was conducted by trained interviewers. While asking fewer questions of each respondent in the interviews would mean that the number of respondents interviewed would have to be increased, it would permit more time to be devoted to examining the reason for the responses given. In the actual pilot-test, the field staff were only permitted to obtain brief reasons for non-response, and while this prevented interviewer bias, it also excluded the opportunity to follow-up on the reasons why each particular response was given. It is also likely that the respondents would have been poorly motivated by the prospect of supplying written comments in addition to completing a full questionnaire.

In preparation for the 1981 Census of Population and Dwellings, a pilot test was conducted in September 1979. Five geographical areas were chosen under the following criteria:

1. At least one area from each of the northern North Island, the southern North Island, and South Island regions were to be included.

2. In order to reduce the possibility of adverse publicity being given to the pilot test, centres of population with strong media coverage were to be avoided.

3. In order to reduce non-participation by households selected for the pilot test, centres of population which were heavily surveyed by commercial market relations firms were to be avoided.

4. The 3 centres included in the 1975 Pilot Test were to be avoided (Hamilton, Feilding and Dunedin).

The selected areas were Papakura City (south of Auckland), Whakatane Borough (in the Bay of Plenty), Wanganui City (south-west of central North Island), Hutt Valley (north of Wellington) and Invercargill City (in southern South Island). The pilot test was limited to 600 households, with proportional allocation of households. In other words, the number of households selected in each centre was based on the comparison of the number of houses and flats in that centre with the number of houses and flats in all five centres, at the time of the 1976 Census of Population and Dwellings. This method of sampling is referred to as **sampling proportional to size.**

The resultant allocation of households to each centre was as follows: 160 to Papakura, 40 to Whakatane, 80 to Wanganui, 200 to Hutt Valley, and 120 to Invercargill. Systematic random sampling was employed to select the addresses to be visited, and the sampling frame consisted of the occupied and unoccupied dwellings, dwellings in the course of erection, and vacant sections enumerated in the 1976 Census of Population and Dwellings. If a selected address no longer contained a private dwelling, or if the household at the address could not, or would not, participate in the pilot test, a neighbouring private dwelling was substituted for the originally selected address.

As for the 1975 Pilot Test, respondents are asked to complete Personal and Dwelling Questionnaires and Comment Forms. The respondents were asked to supply descriptions of problems or difficulties experienced in completing the Questionnaires, and to comment on the format, wording, colour and usage of arrows. A question on 'Health Disabilities' which was being considered for inclusion in the 1981 Census Personal Questionnaire was included to assess the public reaction, and respondents were also asked for their attitudes to postal return of the Questionnaires in the Census proper. The following categories were supplied, and each respondent was asked to nominate one of them:

Very Happy, Happy, Neither Happy nor Unhappy, Unhappy, Very Unhappy

Questions pertaining to the Dwelling Questionnaire were also included on the Comment Form. The first sought to ascertain the degree of difficulty experienced by the household in deciding who was the "occupier" of the dwelling" for Census purposes (in other words, who would complete the Dwelling Questionnaire in the Census proper) and how difficult it was to decide whether the dwelling was private or non-private. Again, five categories were supplied, and each respondent was asked to indicate which one was applicable. The listed categories were: *Very Easy, Easy, Neither Hard nor Easy, Hard, Very Hard* . Details were also sought on furniture and home appliances supplied by the landlord to households in rented dwellings. The respondents were asked to indicate which of the following was provided by the landlord: *any tables, any chairs, any beds, refrigerator, clothes washing machine.*

No attempt was made by field staff to obtain any missing information on the Questionnaires, but they did record on the Questionnaires any verbal comments offered by the respondents on any aspects that had caused difficulty.

Unfortunately, the results of the pilot-test were not used to determine the final format of the Questionnaires used in the Census proper. For the 29 questions included in the Personal

Questionnaire, 2 were as pilot-tested, 9 were slightly modified, but only 1 of these was as a consequence of the pilot-test experience; 16 were significantly modified, but only 2 of these were as a consequence of the pilot-test experience; and 2 questions were not included in the pilot-test, although 1 was pilot-tested in 1975.

Of the 18 questions in the Dwelling Questionnaire, 2 were as pilot-tested, 9 were slightly modified, but only 4 of these were as a consequence of the pilot test experience; 7 were significantly modified, but only 1 of these was as a consequence of the pilot-test experience.

However, one direct result of the 1979 Pilot Test was that the physical size of the Questionnaires was deemed to be too large, both for the respondents and the interviewers. As a consequence, the Notes for Guidance were transferred to a separate disposable document which was slotted into each Questionnaire.

A series of tests were conducted during the development process of the 1986 Census Questionnaire (New Zealand Census of Population and Dwellings Census '86 Questionnaire Content and Submissions, 1985). A small-scale preliminary exercise was conducted in September 1982 to test the feasibility of using an Optical Mark Reader for data capture, thus eliminating the need to transcribe data to coding sheets. Four pretests, two skirmish tests and two pilot tests were conducted during the period from August 1983 to November 1984.

The pretests focused on a limited number of question topics (usually less than five), and involved approximately 200-300 households. To ensure consistency of interviewer training, they were generally conducted in one centre. The pretests usually involved the testing of alternative formats of the questions under study, and three of the tests involved a follow-up interview with selected respondents, to gain information on each respondent's understanding of the terms used, the general

impressions and difficulties experienced by respondents, and to determine the reasons for inconsistent responses.

The small scale skirmish tests were conducted in the field to develop the best wording or layout for one or two questions, and involved experimenting with alternative versions of a question.

The full range of questions were included in two pilot tests involving approximately 1500 households in several centres. The second pilot test was the final test in the series and was regarded as a dress rehearsal for the Census questionnaires and enumeration procedures. The timetable for the testing programme and the questionnaire topics tested is given in Appendix 5.1.

In addition, a pilot test of the questionnaires was conducted to test the new data processing system using Computer Assisted Coding which was developed by the Department of Statistics for the 1986 Census. In the 1986 Personal Questionnaire, apart from the name and address of the respondent, 23 of the 27 questions were answered by ticking or entering an appropriate number in the answer boxes provided. It was hoped that these responses could be processed within 6 months, and be published (without editing) as provisional data. In actual fact, the publication of the Provisional Local Authority Population and Dwelling Statistics bulletin was cancelled because the final Local Authority Population and Dwelling Statistics for the 1986 Census were published several months earlier than scheduled.

Australia

Three small pretests, each covering approximately 500 households, were conducted to evaluate topics which had met all the selection criteria. The first pretest was conducted in Canberra on 25 May 1978, and covered the following topics:

Age; marriage duration; religion; childcare; languages (proficiency in English, first language and languages used last week).

The second pretest, conducted in Sydney on 29 June 1989, tested the following topics:

Source of income; pensions; retirement provisions/schemes; leisure; holidays; trips taken (150 km or more and overseas); household relationship; age; marriage duration.

The following topics were covered in the third pretest in Melbourne on 27 September 1978:

Labour force; major activity; voluntary work; holidays; childcare; income; family structure; fertility; unemployed; languages (first language and languages used last week); racial origin; source of income.

Following the delivery and collection of the trial schedules, follow-up interviews were conducted to ascertain the accuracy of the responses and the reasons for nonresponse.

Four field tests were conducted in 1979 to evaluate the layout of the 1981 census schedule. Unfortunately, the first three tests were conducted without knowledge of which topics census would be included in the census proper, as the government decision on the census content was not announced until November 1979. The wording and layout of the questions, the wording and layout of the response categories, the design and utility of the instructions, and the order of the questions within the schedule were evaluated using twelve schedule designs. The reasons for nonresponse and inaccuracies were determined from follow-up interviews. Interestingly, the field tests showed that the data quality was affected by the differing designs of the questionnaires and instructions.

A dress rehearsal was conducted 12 months before the census proper. The dress rehearsal schedules were used to test the

processing systems and to provide processing rates to facilitate estimation of the staffing requirements for the census. In order to simulate recruitment and training conditions, temporary group leaders, collectors and processing staff were recruited from three local government areas, and schedules were delivered to 17,000 households.

India

Draft versions of the population record (the second part of the Household Schedule) and the universal slip (which was the first part of the Individual Slip and consisted of demographic, social, cultural and economic questions) were tested in a pilot study in 9 states over the period 12-21 June 1978. The sample consisted of 20 rural and 10 urban units from each state, and the schedules and the instructions were printed in English. The first pretest was held in September and October of 1978, and involved all states and unions except Lakshadweep, Mizoram, Dadra and Nagar Haveli, and Pondicherry. The population record and both parts of the Individual Slip (the universal slip and the sample slip which comprised of questions on migration and fertility) were used. Both household sampling and area sampling were employed. The schedules were again printed in English, and census staff performed the enumeration duties.

The second pretest was conducted in the second half of June 1979 and, in preparation for the census proper, all the schedules and instructions were translated into regional languages, and schoolteachers were employed as enumerators. Ten units (5 rural and 5 urban) were selected.

As a result of the pilot tests, it was found that household sampling was not feasible, and for the census proper, a 20% sample of enumeration blocks were systematically selected from 12 states for the sample slip, while 19 states and union territories were canvassed entirely with the sample slip.

Pilot Testing in the United States of America

During the pilot-testing phase, several experimental programmes examining alternative approaches to the 1980 U.S. Census were run and, excluding the alternative questionnaire experiment which employed a national sample, each experiment was implemented in only a fraction of the district offices. The alternative questionnaire experiment tested the effect of questionnaire design on mail-return and item-completion rates. The other programmes examined alternatives to the delivery of questionnaires by the Postal Service, the cost-effectiveness of following up non-responding households by telephone rather than a personal visit, enumerator training methods, coding, imputation procedures, quality controls, the publicity programme, and examination of various sources of error in coverage and content. For the 1980 Census proper, 90% of the population were canvassed by the mail-out/mail-back system and the remaining 10% by the traditional door-to-door visit by enumerators. (The corresponding figures for the 1970 Census were 60% and 40%, respectively.)

The US Census Bureau used intensive reinterview design in the 1976 and 1986 National Content Test (NCT) (Johnson and Woltman, 1987) to estimate the response biases of alternative question versions that had been proposed for the 1980 and 1990 Census. In this design, a probability sample of households is selected, and mail-out/mail-back Questionnaires with alternative question versions are randomly assigned to these households. The returned Questionnaires are regarded as the initial interviews. Two months later, an intensive reinterview is conducted on a subsample of the original households, using the same questions. The idea is to compare the responses given for each question at the initial interviews and reinterviews.

The average difference between the finite population parameter and its estimate is estimated by the total survey mean square error (MSE). The total survey MSE is decomposed into sampling

error and response or measurement error components. **Sampling error** is error that results from using a probability sample rather than a complete enumeration to estimate a parameter of a finite population. **Nonsampling error** is all other sources of error, and includes data collection error, coding errors and processing errors.

The MSE is obtained by expanding $(Y_{ut} - y_u)^2$ and taking expectations, first with respect to the measurement trials (averaging over t), and second with respect to the sample design (averaging over all possible realisations of u, incorporating any assumptions about nonresponse and incomplete coverage), and by assuming that the sampling distributions are uncorrelated with the measurements or response deviations.

The decomposition of the MSE is as follows:

(1) sampling error variance;

(2) sampling bias (including biases due to nonresponse and imperfections in the sampling frame or the list of sample units used to collect the sample);

(3) sample response variance (variability in identical, independent repeated measurements of the same sample unit);

(4) correlated response variance (variability caused by correlation among the response errors of the different sample units such as correlation caused by assigning the same interviewer to more than one household);

(5) response bias (a constant difference between true and recorded values that recurs in repeated identical independent measurements of the same sample unit).

To estimate the sample response variance, 2 independent replications of the same measurement procedure is generally required (as opposed to using an intensive reinterview as one of the measurements). However, the sampling and simple response variance components of the MSE both diminish in magnitude as the total sample size increases. In contrast, the response bias and correlated response variance do not, and are likely to dominate the MSE in large samples. To estimate the correlated response variance, the interviewers are randomly assigned to specified households.

The approach currently used by the Census Bureau assumes that each person in the population has a true state that corresponds to one of the respective categories of a question. The general linear probability model

$$Y_{ut} = y_u + B_u + d_{ut}$$

is employed, where the subscript u denotes the members of the sample for one of the experimental question versions and the subscript t refers to the t^{th} in a series of hypothetical trials or repeated measurements of the same respondent using the same measurement procedure. It is assumed that the repeated measurements are independent, the response errors are independent in repeated measurements of the same units, and that

$$E[d_{ut}|u] = 0.$$

In the above model, Y_{ut} represents the t^{th} response given by the u^{th} respondent, and y_u is the corresponding true state. The response error for the t^{th} measurement of the u^{th} unit,

$$Y_{ut} - y_u$$

is defined as the sum of the response bias component B_u and the random response error component d_{ut}.

It is assumed that bias from the initial interview does not contaminate the reinterview standard of unbiasedness. Moreover, since the initial measurements were mail responses, the correlated response variance is assumed to be zero, that is,

$$E[\, d_{ut}\, ,\, d_{vt} \mid u,v\,] = 0$$

and thus interviewer effects are assumed to be the main source of correlated response variance.

Because of the desire to minimise the time required for respondents to complete the Questionnaires and for ease of coding, most census or survey questions ask for a response which falls into one of several categories. For instance, respondents are often asked to tick the box corresponding to a selected option. In such cases, the information collected is discrete data, and if the question has a 'Yes/No' response, then the response Y_{ut} (the t^{th} measurement of the u^{th} respondent) and y_u, the true value of the u^{th} respondent, are both binomial random variables. In other words,

$Y_{ut} = 1$ if u^{th} respondent answers "Yes" in the t^{th} measurement trial,

$Y_{ut} = 0$ if u^{th} respondent answers "No" in the t^{th} measurement trial

and

$y_u = 1$ if the correct answer is "Yes",

$y_u = 1$ if the correct answer is "No".

The responses to each question in the initial interview and reinterview are then cross-classified in the following table:

	Initial Interview			
	Quest. A		*Quest. B*	
Reinterview	Yes	No	Yes	No
Yes	a_A	b_A	a_B	b_B
No	c_A	d_A	c_B	d_B

where $a_A + b_A + c_A + d_A = n_A$,

$a_B + b_B + c_B + d_B = n_B$,

n_A = size of sample A = the number of respondents answering Questionnaire version A

n_B = size of sample B = the number of respondents answering Questionnaire version B.

Hence, in the first two cells of the table, a_A is the number of respondents in sample A who answered "Yes" in both Questionnaire version A and the reinterview, and b_A is the number of respondents in sample A who answered "No" in Questionnaire version A and "Yes" in the reinterview.

The proportion of the respondents in the population with the characteristic who will be misclassified in the initial interview, the **false negative rate**, is estimated by

$$\theta' = \frac{b}{a+b}$$

and the proportion of the respondents in the population with the characteristic who will be misclassified in the initial interview, the **false positive rate**, is estimated by

$$\phi' = \frac{c}{c+d}.$$

In other words, the number of respondents in sample A who answered "Yes" in the reinterview is $a_A + b_A$. Of these respondents, the proportion the number of respondents who answered "No" in Questionnaire version A, is b_A. Hence the estimate of the false negative rate for Questionnaire version A is

$$\frac{b_A}{a_A + b_A}.$$

Similarly, the number of respondents in sample A who answered "No" in the reinterview is $c_A + d_A$. Of these respondents, the proportion the number of respondents who answered "Yes" in Questionnaire version A, is c_A. Hence the estimate of the false positive rate for Questionnaire version A is

$$\frac{c_A}{c_A + d_A}.$$

Assuming that the intensive reinterview measurement is free of response bias, the *net difference rate* (NDR) is an unbiased estimator of the overall response bias of the experimental question version, that is,

$$NDR = \frac{a+c}{n} - \frac{a+b}{n} = \frac{c-b}{n}$$

The overall response bias is the rate of misclassifications in the initial interview. Although biases due to nonresponse and incomplete coverage are likely to be present, if they are constant across panels and repeated measurements, the relative response biases of alternative question versions can be assessed by comparing the NDR's.

When interviews and reinterviews measure independent replications of the *same* measurement procedure, the proportion of inconsistent reports divided by 2, $\frac{b+c}{2n}$, is an unbiased estimator of the overall or average response variance of the common measurement procedure, but when the reinterview is regarded as an improved measurement procedure, $\frac{b+c}{2n}$ underestimates the response variance of the initial measurement procedure.

To test for total difference between the Questionnaire versions, a χ^2 test of row-column independence (with 3 d.f.) is applied to the 2x4 array

$$T = \begin{matrix} a_A & b_A & c_A & d_A \\ a_B & b_B & c_B & d_B \end{matrix}$$

Because of the design, the hypothesis of independence should not be rejected.

To test whether the proportions of respondents answering "Yes" and "No" in the reinterviews are the same, a χ^2 test (with 1 d.f.) is applied to the 2x2 array

$$C1 = \begin{matrix} a_A + b_A & c_A + d_A \\ a_B + b_B & c_B + d_B \end{matrix}$$

To test whether the false negative rates for the Questionnaire versions are the same, a χ^2 test (with 1 d.f.) is applied to the 2x2 array

$$C2 = \begin{matrix} a_A & b_A \\ a_B & b_B \end{matrix}$$

and to test whether the false positive rates for the Questionnaire versions are the same, a χ^2 test (with 1 d.f.) is applied to the 2x2 array

$$C3 = \begin{matrix} c_A & d_A \\ c_B & d_B \end{matrix}$$

The expected values used in the χ^2 test are obtained from the following table:

	Initial Interview			
	Panel A		*Panel B*	
Reinterview	Yes	No	Yes	No
Yes	$n_A p_R(1-\theta_A)$	$n_A p_R \theta_A$	$n_B p_R(1-\theta_B)$	$n_B p_R \theta_B$
No	$n_A q_R \phi_A$	$n_A q_R(1-\phi_A)\phi$	$n_B q_R \phi_B$	$n_B q_R(1-\phi_B)$

where the probability of a false negative,

θ_U = $Pr(Y_{ut} = 0 \mid y_u = 1)$

= Pr(response = "No" when true state = "Yes"),

the probability of a false positive,

ϕ_U = $Pr(Y_{ut} = 1 \mid y_u = 0)$

= Pr(response = "Yes" when true state = "No"),

p_R = proportion responding "Yes" in reinterview,

q_R = $1 - p_R$

= proportion responding "No" in reinterview,

and the response bias component,

B_U = $-\theta_U$ if $y_u = 1$

= ϕ_U if $y_u = 0$.

As noted by Johnson and Woltman (1987), the assumption of independence of interview bias and reinterview bias may not be valid, because respondents may remember their initial responses to the questions. Moreover, inferred differences between interview and reinterview might reflect actual changes in the population which occurred between the measurements, rather than any difference in the measurement procedures.

Linear measurement models of the form

$$Y_{ut} = y_u + B_u + d_{ut}$$

were originally developed for continuous data, and are likely to be invalid when the mean of Y is close to 0 or close to 1. In such cases, linear logistic models might give a better approximation since taking the logarithms of the observed values will reduce the skewness of the data.

The assumption that each person in the population has a true state that corresponds to one of the respective categories of a question is quite valid for options such as Male versus Female, or Employed versus Unemployed, but is not necessarily valid for true states which could be anywhere between two boundaries. For example, a question on disabilities may give the options "prevented from working" and "not prevented from working", whereas a respondent's true state may be somewhere in a continuum between "not at all limited in working" to "totally prevented from working". Moreover, the existence of true states, which allow response biases to be obtained, may be questionable when the phenomenon being measured seems subjective or sensitive.

Chapter 6

QUALITY CONTROL: COVERAGE

Introduction

The term **census coverage** refers to the response rate achieved in the census, and **content error** is a measure of the inaccuracy of information supplied by the respondents, whether deliberately or unintentionally. The **nonresponse bias** is the difference between the true data value and its estimate, which is obtained from census (or survey) data. As mentioned in Chapter 3, some members of the population may be missed during a census operation, while other persons may be counted more than once. These events are defined as **underenumeration** and **overenumeration** (or **multiple enumeration**), respectively.

Net Undercoverage

Overenumeration generally occurs when members of the population change their addresses at a time very close to the census date and are counted at both their former residences and the new residences, or when persons own holiday homes and are enumerated at both residences.

Underenumeration arises from two sources:

(i) Not all of the members of a household are counted. This generally occurs when individuals deliberately avoid being enumerated, or when adult members of a household do not understand that information is required on all members of the household, including babies and children. The US Bureau of the Census (1975b, p.5) estimated that nearly

half of all people missed in the 1970 Census lived in households where others were enumerated.

(ii) Entire households are missed. Apartments or holiday homes which are in or adjacent to houses, or attached to shops and offices are not always identified as separate dwellings. Housing sections are frequently split into two, and new houses built behind the existing houses. Huts, caravans and tents in caravan parks, camping grounds and temporary camps will not always be detected, particularly in the latter case, as there will often be no official record of temporary camps. Where a master address register is maintained for mail-out/mail-back enumeration, some dwellings may be incorrectly classified as vacant, while other dwellings may be omitted from the address register because census staff are not aware of their existence. Where door-to-door enumeration is used, field maps may not show new roads and changes of road names. Errors may have occurred in the definition of the boundaries of the Enumeration Districts (or Sub-districts), resulting in gaps between the Sub-districts, and failure to canvass all households.

Pre-census checks by field staff on whether private dwellings are listed on the master address register and whether they are correctly classified as "occupied" or "vacant" will reduce the undercoverage rate of the census. However, if all members of some households are determined to avoid enumeration, it will not be possible to detect all incorrect classifications, no matter how exhaustive the efforts are to achieve maximum coverage.

The US Bureau of the Census has used a mailout/mailback system for the last two censuses and has constructed the Master Address Register (MAR), which is the list of addresses to which the questionnaires are sent. Despite intensive efforts by the Bureau to ensure that the MAR was as complete as possible, research by the Bureau (US Bureau

of the Census 1973a) and by the General Accounting Office (1980) revealed that omissions from the MAR were concentrated in central cities, particularly in older buildings where there were several households at one address, and in rural areas, presumably where residences have no mailboxes and cannot be seen from the road. During the 1970 US Census, a sample of households originally enumerated as vacant was rechecked, and it was found that 11.4% of them were occupied at Census time.

Because the term "census coverage" does not clearly indicate whether both underenumeration and overenumeration have been taken into account, the terms **net undercoverage rate** or **net census error** are used. The net undercoverage rate (census error) is defined to be the difference between the gross undercoverage rate and the rate of erroneous enumerations. Generally, the frequency of overenumeration is very small relative to underenumeration, and is also easier to detect.

Various methods of estimating the net undercoverage are discussed below, but it should be mentioned here that, because of the inherent problems with each of these procedures, the methods can only produce rough estimates of the net census error.

Once the degree of net undercoverage has been estimated, it has to be decided whether to use this information to amend the population count or to merely report the actual population count. Because government funding and representation are generally allocated proportionally to the size of population groups, this is an important decision which is made all the more difficult by the fact that the estimates of net undercoverage are themselves prone to error. In America, several local authorities have challenged the official censal counts in court, producing unofficial counts obtained from alternative population lists. Such litigation is conducted at great expense to both parties and highlight the problems

involved in estimating the net census error and in deciding how to use the estimates to adjust the data.

If the decision is made to adjust the raw head count, the extent of the adjustments must then be determined. Experience has shown that the degree of undercount varies between racial groups. For example, in the USA, the black population and the Hispanic population have historically proved difficult to enumerate, resulting in relatively high levels of undercount. In 1970, 1.9% of the total white population were underenumerated, whereas the rate for the total black population was 7.7%. Results of evaluation studies (US Bureau of the Census 1982b, attachments 5 and 15) indicated that the 1980 omission rates for various racial categories were: non-Hispanic whites 4.8%; non-Hispanic blacks 10.9%; Hispanics 11.7%. Thus, if no adjustments for undercoverage were made, or if the population counts for all races were uniformly adjusted, then population groups with higher levels of underenumeration would be disadvantaged.

To make matters worse, in addition to coverage differentials for the various races, coverage differentials are generally observed within each racial group with respect to geographic areas, sexes, and age groups within each sex. For this reason, it is generally not sufficient to adjust the census data for undercount by race and sex at a national level only, since it is not valid to assume that the undercount for a particular race in local areas is consistent with the estimated national undercount. The Hispanic racial category used by the US Census Bureau includes several widely different cultures and populations, each with their own living patterns. According to Bailar (1985), the proportion of Hispanics living in the southwest cities is much lower than that in a north-eastern state such as New York. Moreover, the Hispanics in the southwest cities are predominantly Mexican, whereas those in New York are predominantly Puerto Rican. Thus a national adjustment for undercount for any population group could not be uniformly applied across all regions.

O'Brien (1984) reported that an investigation into the relative coverage in the 1980 Census of Puerto Rico by the US Census Bureau produced an estimated coverage of 94.05% for persons aged 30 years and over, whereas the coverage for persons aged under 30 years was 90.85% and, in particular, coverage for the 20-29 year age group was only 88.02%. The undercoverage estimates were obtained by matching the 1980 Puerto Rico Labour Force Study with the 1980 Census. It should be noted here that the coverage estimate of 88.02% for the 20-29 year age group means that 11.98% of this age group could not be matched by the study into the census based on the available information. It does not imply that the undercoverage rate for this age group was 11.98%. O'Brien states that a similar coverage age group differential was also experienced in the 1980 US Census.

Although matching census data with survey data is not considered to produce reliable estimates of undercoverage, these estimates do indicate a marked coverage differential between the age groups. It is highly probable that the 20-29 year age group is more mobile than the other age groups, and hence will prove more difficult to match. The usage of alternative population lists, such as survey data, to assess census undercoverage rates is discussed more fully in the following section on post-censal coverage evaluations.

Because of the coverage differential between geographic areas, adjustments for undercount by variables such as age and ethnicity should only be executed if reliable estimates of the undercount for small areas could be obtained. The lack of historical data for many minority groups and the lack of internal migration data compound the other problems associated with obtaining reliable estimates of undercount, and mean that there are no known criteria which can be applied to verify the accuracy of such estimates. Local-area population estimates will depend on the models chosen for the estimation. Different models, each just as justifiable as the others, can produce significant swings in the population estimates.

The public relations aspect of the census must also be considered. Censuses have always enjoyed higher response rates than smaller-scale surveys. It is unclear whether this is because of the extensive publicity campaigns mounted when a census is imminent, or whether the simple notion of a periodic head count has a motivating effect on the population. If the latter is true, then a decision to adjust the census data may result in a poorer response rate from the public, who may reason that it is no longer important to furnish the requested information, since the authorities should be able to obtain it using estimation procedures.

Whichever option, if any, for adjusting the population counts is finally selected, an estimate of the under-enumeration rate, and information on how this estimate was obtained, and whether the population counts have been adjusted (and by what method) should also be furnished.

Post-censal Coverage Evaluation

In order to assess the success of a census, post-censal checks on the coverage achieved and the content error incurred are commonly undertaken. However, it is not sufficient to estimate the coverage achieved for the total population. As mentioned earlier, experience has shown that the response rates will vary markedly between different ethnic groups, and these must also be assessed as carefully as possible. Moreover, the national total nonresponse rates for racial groups cannot be applied universally to smaller geographic areas if the racial groups are composite groups, as the composition of such groups will vary from area to area. The total nonresponse rates also vary between different age groups and sexes, with males in the 15-35 years age group proving to be the most difficult to enumerate.

Most householder or dwelling census questionnaires require a listing of temporary absentees and visitors, and this

information can be used to check for undercoverage. For instance, the listing of temporary absentees can be used to attempt record linkage to questionnaires that supply the same names and usual place of residence. If no match can be found, then either undercoverage or content error has occurred.

An estimate of the coverage achieved by the census for various racial groups or for the total population can be obtained by comparing the population count with that predicted by updating the previous census count with data on births, deaths and migration registered during the intercensal period. This is known as **demographic analysis**, as it uses the **demographic equation**

expected population count = previous census count + births - deaths + immigration - emigration

The difference between the census count and its estimate (or predicted value) at the census date, obtained by adjusting previous census data or by some other means, is known as the **error of closure.**

Serious limitations of demographic analysis are that it can only be applied successfully at the national level, it only works for those groups for which there are historical statistical series, and it depends on comprehensive and accurate birth, death and migration records. As mentioned earlier, lack of internal migration data prevents estimates being made of local (or state) populations, and the populations of minority groups cannot be estimated because of the lack of historical data. Also, estimates need to be made of the numbers of foreigners who have entered a country illegally, referred to as illegal aliens in the USA, as these persons will not have notified the authorities of their existence.

Estimates obtained using demographic analysis will vary according to the assumptions made. It is common practice to

produce a range of estimates, using varying migration rates and birth and death rates. This is not as strange as it may sound, as several countries experience varying levels of illegal immigration. Persons who have entered the country illegally will generally continue to avoid authorities, and will not register any births or deaths which occur in their families. The practical difficulty in assessing the true levels of births, deaths and migration is common to all methods of estimating census coverage.

It is of interest to note that one such set of estimates using demographic analysis produced an estimated error of closure of 6,200,000 (3.3%) for the 1960 US Census, 6,100,000 (2.9%) for the 1970 US Census and 3,200,000 (1.4%) for the 1980 U.S. Census. These figures were obtained from the Statistical Abstract of the United States of America which was published by the US Census Bureau. Another set of estimates using demographic analysis, quoted by Bailar (1985) is 2.7% for 1960, 2.2% for 1970 and -2.4% for 1980. In other words, a sudden switch was experienced from the pattern of estimated population counts which were higher than the official census count to a lower estimated population count.

One possible explanation for such a massive swing is that millions of illegal immigrants were enumerated for the first time in the 1980 Census. It is suspected that there was a massive influx of illegal immigrants between the 1970 and 1980 censuses. This influx, combined with an outreach program initiated by the US Census Bureau for the 1980 Census, to actively encourage participation by reluctant respondents, would account for the gross discrepancy between the anticipated and actual population counts. Warren and Passel (1983, 1984) estimated that about two million undocumented aliens were counted in the 1980 US Census.

Another problem which must be addressed when using births, deaths and external migration data to update previous census data is that of ethnic "category jumping". This occurs when

members of a population identify themselves in one racial category in one census, and then change to another category in the following census. This effect may be caused by an increased awareness of racial identity, or by changes in the official definitions of the categories. As there are no detailed records tracing individuals from one census to another, patterns of inter-ethnic movement are very difficult to assess.

Undoubtedly, one of the best methods of evaluating census coverage is to compare the census data with data which has been collected independently of the census, but during much the same period of time. In particular, some sections of the population have proved difficult to enumerate; fear of authorities, fear of deportation, lack of understanding of the official language all contribute to the nonresponse rate in censuses. For this reason, some alternative sources of data are necessary to provide comparisons with census data. The alternative population lists should be both representative and focused on the "hard-to-count", and they must have been produced independently of the census, either from surveys or administrative data. However, matching census counts with alternative population lists will only produce an accurate estimate of net census error if **all** of the following conditions hold:

(i) the alternative population lists have no erroneous inclusions (or have been corrected or deleted prior to estimation);

(ii) the matching procedure must be virtually perfect; and

(iii) the two systems must be statistically independent or have a valid basis for estimating the effects of correlations.

Several potential sources of alternative population lists are surveys, administrative data collected by social welfare and medical agencies and lists of taxpayers, drivers, voters and income tax exemptions. However, the issue of confidentiality

would have to be resolved before such lists could be used. It should again be stressed that this method of coverage evaluation will only give reasonable estimates if the alternative population lists are not affected by the same causes of nonresponse experienced in the census enumeration.

In September 1980, New York City filed a law suit against the US Census Bureau. New York City had compiled an alternative population list from 10 local lists and contested the official census count. The Census Bureau was ordered to compare the list to the census and determine:

(a) the number of persons on the list who were counted in New York City in the 1980 Census;
(b) the number of persons on the list who were not living in New York City on 1 April, 1980; and
(c) the number of persons on the list remaining.

The Census Bureau submitted its report in November 1982, in which 8.1% of the New York population were classified as census omissions.

Medical insurance schemes have been in existence in America for some considerable time, and have become established to the point where the medical insurance files can be used to provide a picture of the age distribution of the population which, since it is obtained independently of the census, can be used as a comparison. Special lists of the hard-to-count can be obtained from sources such as rosters of welfare recipients, central city schoolchildren, people sending money orders overseas or patients admitted to public hospitals.

Administrative data has the advantage of providing a sampling frame which is independent of any census or survey. Because any sample drawn from administrative records is not based on household interviews, it is unlikely that the pattern of omissions which occurred in a census or survey will be reproduced in the sample. This advantage can be used to good

effect when investigating groups with traditionally poor census coverage, as sampling can be easily controlled on variables such as race and income, permitting the over-sampling of any "hard to enumerate" groups. However, tracing and matching can prove to be expensive and time consuming, and different techniques should be thoroughly investigated before finally deciding which techniques to use.

Administrative data will generally only furnish a headcount of the population, since the information has generally been collected on a continuous time frame, rather than on one reference night, as for the census. Moreover, it generally does not cover the same topics as the census, or at a sufficient depth. Hence, while administrative data may be used to produce an estimate of net coverage error, it usually cannot be used to provide an assessment of the content error of questionnaires.

On the other hand, there is some administrative data which is highly correlated with census data. Such data could be used to increase the accuracy of estimates such as income and employment at the subnational level by employing the geographic totals of the administrative data. This post-stratification of survey estimates would reduce the variance of the estimates and would also yield small area estimates. Administrative records could also be used to employ multiframe designs, thus improving the coverage of groups not represented adequately in area sampling frames, enabling oversampling of potentially rare groups of special interest in the population, and reducing the sample variance of estimates yielded from the survey.

Data from surveys should also be treated with caution, as experience has shown that groups that had the highest level of undercoverage in censuses were even more poorly covered in surveys, and that this improvement in coverage is especially great among the hard-to-enumerate, low visibility groups. Hansen (1985) attributed this phenomenon to the widespread public understanding and publicity for the census, and the

special procedures used for hard-to-count groups. Matching of census questionnaires with those from surveys has also proved to be difficult, particularly in countries experiencing high mobility of the population. Fay (1985) estimated that approximately 20% of the USA population moves in a single year. The time frame of the census and the construction of the alternative population lists is also important. If a survey is conducted at a time close to that of the census enumeration, then it is highly likely that the persons omitted from the census count will also be missed in a survey. In other words, there will be a high correlation between the two omission rates.

One method of alleviating this problem is to conduct the post-enumeration survey (PES) several months after the census. The longer the time period between the census and the survey, the more independent the PES will be. (For the 1980 U.S. Census, the Census Bureau conducted a two-part PES, with one survey conducted just after the census and a second survey conducted four months later.) However, this benefit of increased independence is offset to some extent by the longer delay in producing the final census estimates and also by the increased difficulty in matching, especially for persons who have moved since the census.

A possible alternative to a PES is a **Reverse Record Check (RRC)**. This is an evaluation programme in which a sample of the population is drawn from a frame created several years prior to the census, traced forward to the time of the census, and matched to the census. As for the PES, the proportion of the sample which is unmatched provides an estimate of the proportion of the population which was missed in the census; in other words, an estimate of the undercoverage rate.

Because an RRC selects the sample to be matched to the census at a greater distance in time from the census than the PES sample, in theory, an RRC will overcome all or part of the bias introduced by a statistically correlated PES. However, even if this assumption is correct, it will be offset to some extent by

the bias introduced by the inability to successfully trace some fraction of the sample to census day.

To create an independent list to match against the census, a sample is usually drawn from the previous census, and supplemented by a sample of birth records, immigration records and a sample of people missed in the last census (as determined by a PES or previous RRC). The advantage of using a previous census as a sample frame is further increased by including a sample of missed persons from an earlier PES or RRC. Furthermore, since the data for a RRC has been collected prior to the census, it is ready for initial matching as soon as the census data is available, whereas a PES must normally be conducted several months after the census in order to avoid overlapping with and interfering with the census itself.

However, although the sampling frame for the RRC tends to be more independent and more complete than can be achieved by a PES, and the RRC facilitates immediate matching, as mentioned above, these advantages are balanced and may be overwhelmed by the problems of tracing, and the final sample that can be traced may be neither complete nor independent.

The U.S. Bureau of the Census used the RRC technique in 1960, and Statistics Canada has been using the RRC since 1961. Both organisations used the technique of **Retrospective Tracing** (tracing after the current census was completed), and experienced the following nonresponse rates (Hogan 1983):

U.S., 1960		Canada,1976	
Source Not Located	Percent	Source Not Located	Percent
Total	12.2	Total	4.8
Census	9.0	Census	3.1
Missed	16.8	Missed	9.6
Births	14.4	Births	7.6
1950 Registered Aliens	0	Immigrants	10.6

It should be noted here that since Canada conducts censuses every 5 years, the Canadian tracing period is half that of the American censuses. Despite the fact that Canada has used the RRC technique for four censuses, the nonresponse rate of 4.8% is still undesirably high.

The U.S. Census Bureau used a sample of births between the 1950 and 1960 Censuses, a sample of immigrants (registered aliens), a sample of the population identified as missing in the 1950 Census Post-Enumeration Survey, and a sample of the population included in the 1950 Census. Despite intensive efforts to follow up all the samples, the RRC was not sufficiently successful to provide acceptable measures of undercoverage (Hansen, 1985).

The U.S. Census Bureau also used a RRC for Medicare recipients to evaluate new procedures for the 1970 Census. However, as in earlier censuses, while the coverage evaluation using area sampling and matching was relatively successful in identifying coverage of dwelling units and households, the measurement of undercoverage was inadequate (Hansen, 1985).

As was briefly mentioned in Chapter 5, the problems of tracing may be substantially reduced by the technique of **Forward Tracing**. Forward tracing also uses a sample from a previous census, a sample from immigration records, and a sample of missed people from the census, but unlike the RRC, the tracing begins at the beginning of the period.

Forward tracing can be approached using any of the following techniques:

1. Tracing without any personal contact.
2. Tracing with contact at the beginning, but no further personal contact.
3. Tracing with periodic contact, year-by-year if necessary.

While personal contact will almost certainly improve the tracing, it may introduce a conditioning bias, and will

undoubtedly raise the costs. As usual, there is a trade-off between increased efficiency and increased costs. The cost of Forward Tracing may be offset by improved coverage and valuable information on mobility patterns of the population.

If a respondent moves after the tracing period has begun, and has not notified census staff, potential sources that can be used to locate respondents are relatives or neighbours of the respondent or other contact persons known to the census staff, the Post Office, telephone books, directory assistance, city or suburban directories, utility companies, and government or local authority administrative records.

Fay and Cowan (1983) reported that major studies conducted by the US Bureau of the Census to evaluate undercoverage using direct survey methods have not been as successful as the traditional method of demographic analysis. Demographic analysis has provided systematically higher estimates of undercount, and the estimates have been more internally consistent. The two major causes for the poorer performance by direct survey methods are correlation bias and the problems of matching respondents between the census and surveys, caused to a large extent by the high mobility of the population. Experience has shown that persons missed in the census are also likely to be disproportionally underrepresented in the sample surveys. This means that the two population counts, obtained from the surveys and census, are not independent of each other. In other words, there is a correlation between the data obtained using surveys and census. This correlation bias prevents surveys from being a good yardstick against which to measure census coverage.

Post-Censal Content Error Evaluations

Many countries check the content error of questionnaires after a census has been conducted by checking for internal consistency of responses within the questionnaires and by linking the

questionnaires with records of the same respondents from earlier censuses, surveys, or administrative data such as birth records, to check for consistency of the information supplied.

Since it is usually mandatory for all persons to complete census questionnaires, it is possible to use the census as a benchmark, and examine non-response bias in surveys, which are usually voluntary. If the respondents are asked to complete the census questionnaires, record linkage can also be used to compare the response rate to that obtained from interviewer-conducted surveys. However, care needs to be taken that such record linkage does not violate the guarantee of privacy given to respondents. If the record linkage is only used to improve the quality of census data, then it might be argued that no violation has occurred, since information on individuals will not be published.

Another potential problem is that the response rate in the census may be affected if the public believe that the information can be readily obtained by linkage to other sources. Successful linkage depends on all the censuses and surveys having good response rates and accurate information, since the identifiers used to obtain the linkage, such as name and address, must be present and accurate. Hence the introduction of record linkage, while attempting to improve the census data, may in fact have an adverse affect on the quality of information supplied. Moreover, when discrepancies between responses are identified, a decision has to be made as to which response is to be accepted as valid, and which is to be corrected.

Other areas commonly investigated in post-censal evaluations to check content error are **age-heaping** and **underenumeration of babies**. Age-heaping refers to the phenomena of respondents, particularly older respondents, reporting their ages as ending in particular digits, usually 5 or 0. For instance, it is quite common for a person of 64 years of age to report their age as 65 years, or a person of 51 or 52 years of age to report their age as 50 years. By examining bar

graphs of the population distributions, it is easy to see which ages are the most popular. It is common practice to examine the separate distributions for each sex, in order to detect any difference in age preferences between the sexes. Computer programmes will readily produce bar charts of the frequencies of the ages for the total population by sex, for racial groups by sex, and for any other specified population group by sex. It has been found that for digits other than 5 or 0, even numbers are more popular than odd numbers, but the outstanding feature of reported ages are the peaks produced at digits ending in either 5 or 0.

Experience has shown that, whereas respondents will supply data on children who have attained one year of age, the corresponding information for children aged less than one year is frequently omitted, causing **underenumeration of babies.** It is unclear whether this underenumeration occurs because respondents do not realise that information is required for every member of a household, including babies, or whether the cause is rooted in superstition. Because of the uncertainty of the cause of cot deaths, many parents do not feel confident that their babies will survive into childhood until they have attained the age of 12 months, and it is quite possible that some parents may feel that to supply particulars on children aged less than 12 months would be "tempting fate".

It is possible to estimate the rate of underenumeration of babies by comparing the enumerated number of children aged less than one year of age with that expected from birth records covering the last twelve months. However, this in itself will be an underestimate, as not all births will have been recorded.

United States of America

For the 1980 US Census, 95.5% of housing units were contacted by mail, as opposed to approximately 60% for the 1970 Census.

People in the remaining areas were enumerated using the conventional door-to-door canvassing by enumerators.

The US Census Bureau uses several methods of evaluating the extent of census omissions. Appendix 6.1 lists the published coverage evaluation studies for the 1970 Census and Appendix 6.2 lists the special procedures employed to improve the 1980 Census coverage and post-censal evaluations. The 1980 Coverage Improvement Program consisted of fourteen distinct operations that were designed to improve coverage either by encouraging public cooperation or by improving the enumeration procedures. Details of the main pre-census and post-censal quality control measures for the 1980 Census are as follows:

Pre-Census Coverage Improvement Procedures and Edit Checks

Mail Areas

The mail-census areas were further categorised into **TAR** (Tape Address Register) and **Prelist** Areas, according to the method used in preparing the address lists. The mailing lists in TAR areas were prepared independently by commercial firms and then updated by a combination of Post Office review and precanvass by census enumerators. In Prelist areas, the mailing lists were generated by enumerator canvass.

(i) TAR Areas

The Census address registers were compiled in late 1979 and early 1980. The U.S. Postal Service conducted an edit of the TAR mailing lists, known as the Advance Post Office Check (**APOC**). During this operation, addresses were added, deleted or amended, as required. The updated addresses were then assigned geographic codes, such as tract and block numbers. The coded addresses were then assigned into Enumeration Districts (EDs), and a Master

Address Register (MAR) was then prepared for each ED. In addition, a Precanvass Address Register (PAR) was produced for each ED, containing a listing of each basic address and the associated number of housing units in the basic address that had been assigned to the ED.

An independent precanvass field operation was then conducted by census enumerators to edit the PARs. In order to check the quality of each enumerator's work, a sample of housing units was deliberately suppressed from the PARs. If an insufficient number of suppressed units were not reinstated by an enumerator, the ED was recanvassed. Two additional postal reviews, the Casing and the Time-of-Delivery checks, were then conducted to further update the MAR.

The APOC resulted in an estimated 5.5% addresses being added to the Census address lists (Thomas et al,1984), the Precanvass operation added 5.0% (Fan et al, 1984), and the Casing and Time-of-Delivery checks contributed an estimated 3.4% additional addresses (Thomas et al, 1984). Of the addresses added as a result of the post office checks, the proportions which were found to be occupied and vacant in urban areas were both 5% higher then those in rural areas. The higher rate of additions in urban areas could indicate that the census coverage may have been less effective in rural areas. However, there may be no grounds for assuming that the rate of additions for the rural and urban areas should have been the same, since housing units are often part of multi-unit structures in urban areas, and the urban drift would result in far more families living in dwellings such as caravans, sheds, lean-tos, huts and even cars. Undoubtedly, it would not be uncommon for some housing units to be unofficially partitioned into two or more separate dwellings to house families.

It is interesting to compare the percentages of improved coverage for the various racial categories added by the Precanvass operation to the corresponding percentages of the weighted population. The improved coverage percentages for Whites and Blacks were 85.0% and 9.3%, and the corresponding percentages for the weighted population were 83.1% and 12.2% (Fan et al, 1984). The improved coverage percentages for Hispanics versus Non-Hispanics were 5.4% and 94.6%, respectively. The corresponding percentages for the weighted population were 6.2% and 93.8%.

As can be seen from the above data, the percentages of Blacks and Hispanics added as a result of the Precanvass operation is disproportionately lower than the percentages for the rest of the population. This differential improved coverage is even more marked when the data for Centralised District Offices (DOs in cities with 1,000,000 or more population) is examined. The improved coverage percentages for Whites and Blacks were 63.0% and 25.5%, whereas the corresponding percentages for the weighted population were 57.7% and 31.3% (Fan et al, 1984). The improved coverage proportions for Hispanics versus Non-Hispanics were 16.9% and 83.1%, respectively, whereas the corresponding percentages for the weighted population were 18.9% and 81.1%.

One reason offered by Fan for the disproportionate number of Whites added to the Census counts as a result of the Precanvass operation is that most housing units added to the TAR lists were single units, and the proportion of Whites in single units was higher than that of Blacks and Hispanics. Undercoverage of minority groups has been of concern to the US Bureau of the Census, particularly since it believes that a flood of illegal immigrants has entered the country since the 1970 Census. In an effort to reduce the differential undercoverage for the 1980 Census, a Non-Household Sources (NHHS) Programme was instigated. The

NHHS was conducted after the census, and is discussed in more detail in the following section on post-censal coverage improvement operations.

(ii) Prelist Areas

In contrast to the precensus coverage improvement operations for the TAR areas, a major postal review (Advance Post Office Check) of the Prelist areas was cancelled, no precanvass check was conducted, and the two postal reviews that were conducted were thought to contain duplicate enumerations due to additions resulting from incorrectly geocoded postal areas. Consequently, the Prelist Recanvass Operation was introduced during the final stages of the 1980 Census to improve the coverage in the Prelist Areas. The operation was conducted in a total of 137 district offices, 134 of which were decentralised offices and 3 were "two-procedure" offices where both conventional and decentralised procedures were used for operations. Some offices were selected because they were located in the more rural parts of the Prelist Areas which had suffered relatively severe coverage problems in previous censuses. Other district offices in Prelist Areas which were not initially selected for recanvass were recanvassed as time permitted. In some district offices, only selected enumeration districts within the district offices were recanvassed. The procedure consisted of enumerators systematically canvassing their districts to establish the every housing unit had a corresponding listing in the Master Address Register. Units which were not listed were added to the register and the occupants of the housing units were interviewed. A recanvass was then conducted by another enumerator to verify the listings. Unverified units or listings were resolved by the respective district offices. The Prelist Recanvass Operation added 0.8% to the count of housing units.

After the census questionnaires were received back from the respondents and edited, any households whose questionnaires failed the full edit check were contacted by telephone or a personal visit. A separate quality control operation was conducted for the telephone follow-up procedure. The major coverage-improvement check was made by having enumerators check addresses which had been deleted from the address registers in earlier operations, and then employing different enumerators in a second follow-up to verify whether the units were actually vacant or should have been deleted.

A local review programme was conducted by providing local government officials with the opportunity to independently review field counts before they were finalised and while census district officers were still open and thus able to check reported discrepancies. The geographically coded address lists were made available to local authorities. Any challenges had to be supported by detailed evidence before the address registers were changed. Counts of building permits, administration records such as tax or utility records, or aerial photography were accepted as sufficient evidence to warrant editing the field counts.

In addition to the precensus address counts, the preliminary population and housing counts for local areas were furnished to local governments. Again, queries about counts raised at the enumeration district or meshblock level required detailed supportive evidence before the counts were amended. At approximately the same time, reviews were also conducted by bureau staff in the district offices and at the Census Bureau headquarters. These local reviews often pinpointed major problems such as clusters of missed housing units, geographic misallocations and incorrect geographic boundaries.

Automation of the Master Address Register

For the 1980 Census, the initial address control file was computerised, but the district offices only received hard copy (paper lists) of the addresses, and changes to these registers, such as additions, deletions, corrections or moving addresses from one enumeration district to another, were made manually. The returned questionnaires were also sorted and checked manually. Addresses from which no questionnaires had been returned were then visited by enumerators. $7.5 million had been budgeted for the sorting and checking phase of the 1980 Census operation, but the actual cost was $19 million (Bounpane and Jones 1988).

Using an automated address file, changes can be keyed in, effecting automatic updating of the file, and the questionnaires can be checked-in and sorted by computer, using bar-code technology and multiple-pocket laser sorters. For cases where the lasers are not successful in reading the bar codes, or are not available, hand-held wands which are attached to computers can be used to scan the bar codes. However, the costs of the two methods must be weighed against their efficiency and post-census utility. Whereas it is estimated that one laser sorter can process about 11 times as many questionnaires as one wand station, it could cost about 50 times more, and require more maintenance and more skilled personnel. Moreover, the microcomputers in the wand stations would have more utility after the census. However, more extensive usage of wands would require more wand stations and production clerks.

Test censuses conducted in 1985 and 1986 successfully implemented an automated address control file and automated check-in. For the 1990 Census, some collection offices will use wand readers, whilst others will employ laser sorters for the check-in phase.

Conventional Areas

For the conventional areas (areas which were covered by door-to-door canvassing as opposed to mail-out/mail-back areas), a coverage check was made on the enumerators' work. Prior to the enumeration, crew leaders made a listing of 24 addresses in each enumerator's district. After the census was taken, these listings were matched to the listings of housing units made by the enumerators. If no addresses were missed, then the enumerator's work was assessed as being of good quality. If one address was missed, the quality was assessed as acceptable, but the missing address was added to the address register. If more than one address was missed, then the enumerator's district was recanvassed.

A postenumeration check for the conventional areas was conducted by the U.S. Postal Service. For every housing unit visited by an enumerator, an address card had been completed. The Postal Service reviewed these address cards, and listed addresses to which mail was delivered but for which no card had been received, and cards received for nonexistent addresses. As a result, an estimated 0.068% of all enumerated households were added to the household count (Thomas et al, 1984).

As the crew leaders had already checked each enumerator's work, no full questionnaire-edit operation was carried out as in the mail-census offices. However, a sample of questionnaires for each enumeration district was reviewed for completeness by office clerks, and the questionnaires were edited as necessary to identify those with missing information that should be included in the follow-up operation. The office clerks also conducted a sample tolerance check to investigate whether each enumerator had correctly employed the sampling pattern by comparing the actual population in the enumeration district to an estimate based on the number of people enumerated on the long-form questionnaires for the enumeration district.

In the follow-up operation, the enumerators telephoned or visited the housing units whose questionnaires failed the edit operation or who refused to complete questionnaires. Enumeration districts that failed the sample tolerance check were "resampled" by transcribing some long forms to short forms and additional long-form information was collected where necessary. Enumeration districts that failed the coverage test were recanvassed.

A local review programme was conducted for the conventional areas. As for the mail-census areas, local government officials were permitted to independently review and challenge the precensus address counts and the preliminary housing and population counts.

Unfortunately, operational problems with the 1980 local review contributed towards only 32% (Lueck et al, 1984) of the government jurisdictions participating in the review. Problems cited by local review staff were: changes in the program; timing; lack of proper planning; and lack of training of personnel to implement the program and to liaise with the local officials. Of the local government authorities which did challenge the field counts, 39% submitted sufficient evidence of inaccurate counts.

US Post-Censal Coverage Improvement Procedures

Post-censal procedures used by the Bureau to evaluate the coverage achieved include demographic analysis, measurement of the omission and overenumeration of various age groups, matching Census data with Current Population Survey data, estimation of the net coverage errors for housing units which received their census questionnaires in the mail, and evaluation of special procedures which had been designed to improve census coverage. As mentioned earlier, the National Vacancy Check detected a misclassification rate of housing units of 11.4%. As a result, about 0.5% of the population was added to

the population counts, using an imputation process which was based on the housing unit and population omission rates estimated from the survey. These rates were also used to randomly convert vacant housing units to occupied housing units.

The Current Population Survey (CPS) is a major survey which has been conducted by the Census Bureau since 1940. It covers the civilian noninstitutional population, and involves sampling 66,000 households every month. As well as facilitating evaluation of omissions of persons within households in the census, and why such omissions occurred, it provides up-to-date labour force information and socioeconomic data. This data is particularly valuable, since the American censuses are conducted every 10 years.

The Post Enumeration Programme (**PEP**) conducted by the Census Bureau is based on a case-by-case matching, and hence is adaptable to any demographic group and to any level of geography. It also differs from demographic analysis in that it does not require a separate estimate of the illegal immigrant population. Following the 1980 Census, the PEP was conducted in two parts: the P-Sample and the E-Sample. The P-Sample was designed to measure gross undercoverage in the census, and 150,000 households surveyed in April and August CPS's were matched against the census. However, duplicate enumerations and matching problems caused by geocoding errors and imputed data generate **overestimates** of the undercount. A P-Sample pretest, called the Unresolved Cases Study Pretest, involved following-up on 56 unresolved matches. The results of the follow-up were as follows (Cowan et al, 1984): 14 cases were matched, 24 were not matched and 18 remained unresolved.

The E-Sample was designed to measure the rates of geocoding error, definitionally incorrect enumerations (caused by incorrect imputation) and duplicate enumerations. A sample of 110,000 households was selected from the census for reinterview, and resulted in an estimated rate of erroneous enumerations of 3.4% for the total population (Cowan et al,

1984). The components of the estimated erroneous enumeration rate were as follows: geocoding errors 1.0%, definitionally incorrect enumerations 1.6% and duplicate enumerations 0.8%.

The results from the CPS-census matching operation were combined with the results from the reinterview sample to provide PEP population estimates, which were compared with the census counts. The proportion of the PES cases not found is taken as the estimate of the proportion of the national population missed by the census (that is, the **gross undercoverage rate**). The net undercoverage rate is estimated at the national level by specified age, sex, race and Spanish-origin categories. However, the accuracy of these estimates is limited by problems in matching the geographic codes allocated for the census and the CPS or reinterviews.

Evaluation of the coverage achieved for the 1980 Census using demographic analysis and Post-Enumeration Programme estimates had limited success. It is of interest to note that the undercount estimate for the 1980 US Census obtained by the Post-Enumeration Programme estimates ranged between 1.0% *overcount* and 2.1% *undercount*, whereas the DA (demographic analysis) estimates of the *undercount* ranged between 1.0% and 2.2% (US Bureau of the Census 1988). As anticipated, matching problems contributed towards the lack of success for the PEP and inaccurate estimates of migration rates caused problems with the DA procedures.

For the 1990 Census, the Bureau intends to have the geographic support system, named TIGER (Topologically Integrated Geographic Encoding and Referencing system) automated. The TIGER system will be used to select a probability sample of census blocks, thus eliminating the necessity for assigning geographic codes to the addresses. This method should ensure that the search for potential matches will be conducted on questionnaires in the right areas, thus reducing the number of unresolved cases in the PES programme.

For the 1980 Census, a Non-Household Sources (NHHS) Programme was instigated in an effort to reduce the differential undercoverage of minority populations. It was conducted only in certain areas with large minority populations and was designed to enumerate persons who had been missed in households for which a census questionnaire had been received; that is, the aim was to reduce undercoverage within households. Census questionnaires were matched against alternative population lists obtained from the Department of Motor Vehicles, the U.S. Immigration and Naturalisation Service and the 1979 New York City Public Assistance file.

Keeley and Thompson (1984) reported that usage of the NHHS lists yielded an improved coverage rate of approximately 1.9% for the total population. Although this rate was well below the anticipated rate of 10%, the NHHS Programme did improve the coverage of Black and Spanish races. The reported rates are as follows: Non-Spanish and Non-Black 23.9%; Spanish 33.7%; Black 28.1%; and Unknown Race/Origin 14.3%.

After the 1970 Census a large-scale survey, called the National Vacancy Check, was conducted to evaluate the proportion of occupied housing units which had been misclassified as vacant. The estimate obtained from the survey was 11.4% and as a result, approximately 0.5% of the population was added to the 1970 population counts by imputation. The imputation process used was based on housing unit and population underenumeration rates estimated from the survey. The estimated misclassification rate was also used to randomly convert vacant housing units to occupied housing units.

The Bureau of the Census decided to eliminate the above imputation procedure from the 1980 Census, by conducting follow-up operations on a 10% sample of housing units which had been classified as vacant or nonexistent. After the first follow-up, housing units which were found to be occupied were matched to the Master Address Register and the persons in the households were processed into the census counts. The second

follow-up consisted of a further check of all housing units which were still classified as vacant or nonexistent. Different enumerators were used for the two stages, and the classifications compared. Any units with conflicting classifications were matched to the Master Address Register and the census counts were amended.

As a result of this coverage operation, 8% of occupied units and 0.8% of the total population were added to the census counts. An additional 2% of misclassified nonexistent units were found to be vacant, and 6% of housing units which were classified as vacant were found to be nonexistent.

It is of interest to note that the proportion of Black and Hispanic persons added to the census counts as a result of the Misclassified Occupied operation was higher than that for the general census population. This finding indicates that the operation should be seriously considered for inclusion in future census operations, as it will assist in reducing the differential undercoverage.

A further coverage study conducted by the Census Bureau for both the 1970 and 1980 censuses consisted of checking a sample of the responses to the following question, which was identified as H4 on the census questionnaires:

How many living quarters, occupied and vacant, are at this address?

Not surprisingly, this study was known as the H4-Edit, and it was hoped that comparison of the answers given by respondents with the number of units in the census records would detect missing housing units in small multi-unit structures. For the 1980 study, a systematic 0.1% sample was taken of the Enumeration Districts (ED's), resulting in 284 ED's, of which 260 were in centralised or decentralised offices. The remaining 24 ED's were conventional ED's, and which did not retain the questionnaires. All census questionnaires in the 260 ED's which had failed the H4 Edit were used to produce a tally of adds for an address, and this tally was then compared with the

Master Address Register. Only 30 of these sample ED's had housing units added to the housing unit count as a result of the study, resulting in 0.1% of housing units being added to the 1980 count. The corresponding rate of additions to the 1970 housing unit count was 0.2%.

US Questionnaire Content Evaluations

The two main content evaluations conducted by the Census Bureau were a content reinterview study and a utility-cost record-check evaluation. The content study involved reinterviewing approximately 12,000 households which had been supplied with the long form of the questionnaire. The reinterviews focussed on items which were new or had been substantially changed for the 1980 Census, and included more extensive, probing questions to measure the accuracy of the responses on the submitted questionnaires. For the utility-cost record-check evaluation, a sample of respondents were supplied with information on the average monthly gas and electricity costs for the previous 12 months, and a further sample were not given any utility-cost information. The census responses for the two samples were then compared to determine whether the furnished utility-cost information resulted in improvements in the census data.

New Zealand

New Zealand is one of the few countries in the world which does not regularly conduct some form of post-enumeration survey to evaluate census coverage and content errors. However, errors of closure are calculated, and those for the New Zealand censuses from 1956 to 1981 are listed in Appendix 6.3. These figures indicate consistent under-enumeration of the New Zealand population in recent censuses, except for the 1981 Census, which had a higher population count than anticipated. The Department of Statistics was aware that the 1976 Census

count was considerably lower than its estimate, indicating that a significant undercount had occurred. However, no adjustments could be made to the published data until the data from the 1981 Census confirmed that the undercount was significantly large. This unusually large undercount may have been largely due to the census following shortly after an announcement that the Government intended to deport illegal overstayers in New Zealand.

Using the 1976 Census count as a base for the expected population total in 1981 produced a high positive error of closure, indicating that some sections of the population which were not covered in the 1971 Census were enumerated in 1981. When the expected population total for 1981 was recalculated using 1971 Census data, the error of closure changed from +0.38 to -0.23. The negative sign is consistent with what would be expected, since underenumeration is far more likely to occur than over-enumeration in a census. The published figures for the 1976 Census were amended, using the data from the 1971 and 1981 Censuses, adjusted for the births, deaths and external migration.

While New Zealand's isolation from other countries should guarantee that the external migration statistics are unaffected by persons entering or leaving at other than official ports, any errors in the migration data will compound the inaccuracy of the error of closure statistic. Some alternative source of data for estimating the error of closure is needed, such as post-enumeration surveys.

As New Zealand has not established facilities to regularly conduct several surveys on a national level, one of the few options for consistency checks on data in a sub-national area is to compare the census counts with alternative population lists which have been obtained independently of the census. For instance, school rolls for particular areas can be compared with the enumerated children of primary, intermediate, and high school age. Any anomalies which are highlighted by this method

should then be carefully examined in an effort to explain the discrepancies, and any necessary adjustments made before publication of the data.

The electoral roll is often used as a frame from which to select samples, but there are several reasons why it is not suitable to use as an alternative population list. It only contains data on individuals who have attained the legal age of voting (18 years); the onus is on the individual to ensure that his or her name is on the roll, and publicity campaigns to encourage people to check that they are included on the roll and that the recorded details are correct are only conducted prior to each election. Since New Zealanders exhibit a high mobility pattern similar to that of the USA, the roll will be incomplete and contain many inaccuracies at any given point in time. Moreover, the time frame for updating the roll is different to that for the census, since New Zealand elections are usually conducted once every three years, whereas the censuses occur on a five-yearly cycle.

Local authorities could also be permitted to review census counts before the census operation were completed. Comparison of Census data with their own estimates of the size of the local population, based on data such as the number of building permits issued and electoral rolls should reveal some of the anomalies, and if sufficient evidence were to be produced of any errors in the enumeration phase, the counts could be amended.

The Department of Statistics has carried out investigations of the levels of underenumeration of babies, age-heaping and other apparent inconsistencies on an ad hoc basis. The first attempt to assess the accuracy of reported ages was conducted in 1921 by selecting a sample census questionnaires and attempting to match the reported ages with the appropriate birth registration records. Although constraints on time and staff resources and problems experienced with the matching procedure limited the sample to 2,219 questionnaires, the study did reveal that, in line with trends experienced overseas, the frequency of

incorrectly stated ages rose substantially with the increasing age of respondents (New Zealand Census and Statistics Office 1925). The study also showed that the majority of respondents misreported their ages by only one year. Unfortunately, the published table does not reveal whether there was any systematic pattern in the ages which were misreported by one year, but it is quite possible that the New Zealand respondents may have tended to favour the even numbers when reporting their ages. An unpublished paper from the Department of Statistics mentions that the 1921 study included a comparison of age misreporting by sex, which indicated that, while females were only marginally less accurate than males in age-reporting, they did show a decided preference for understating their ages.

The next investigation of reported ages was conducted following the 1971 Census. The study was conducted by selecting a sample of 1,339 questionnaires from the 1971 Census and comparing the reported ages with those reported on the corresponding 1966 Census questionnaires. This investigation provided a consistency check rather than an assessment of accuracy of age-reporting, since the censal age data was not verified with data which was obtained independently of the censuses. The study revealed (Department of Statistics unpublished paper) an inconsistency rate of 15.5%; that is, assuming that the ages reported in the 1966 Census were correct, 15.5% of the matched 1971 questionnaires contained misreported ages. This estimate of content error in age-reporting was virtually the same as that indicated by the 1921 study.

Following the 1976 Census, the Department attempted to assess the upper limit for the underenumeration of babies aged 0-52 days, and to determine the actual percentage of babies aged 0-52 days who were not listed as absentees on the Dwelling Questionnaires. A brief description of this study is given in Appendix 6.4, but note that the study excluded ex-nuptial births, and of the nuptial births, only those families which were still residing at the address supplied on the Birth Registration Form

(or whose questionnaires were located amongst those collected from maternity hospitals or annexes) were covered. Moreover, because the 1976 Census exhibited such a large undercount of the total population, the figures obtained in the study cannot be regarded as typical of the level of underenumeration of babies that normally occurs in population censuses in New Zealand.

Khawaja (1982) applied Myers' blended test to single-year-of-age data from the 1961-1981 New Zealand Censuses, which are conducted every five years, and found that there was a pronounced shift in digital preference from one Census to the next, indicating that the same age cohorts have been mis-reporting their ages. Khawaja also found that while the overall age preference was around 2.0% for the Total Population, the preference for the Maori population was almost twice as high, and that the error indices were fractionally higher for males than for females in all censuses excluding 1961.

The 1981 Census data was linked with the Household Survey (HHS) to examine non-response bias in the HHS; with the Social Indicators Survey to compare responses to a self-completed questionnaire with an interviewer-conducted survey; with birth, death and migration records to determine the misclassification of ethnicity; with the Census of Agriculture to determine content error of data on address and occupation; with birth registration forms to check for underenumeration of babies (see above); with 1971 Census records to check for consistency of responses; and internal linkage between temporary absentees and visitors to check for underenumeration (Unpublished paper, Department of Statistics).

Australia

During the computer editing phase of the 1981 Australian Census unacceptable combinations of data (such as a person aged 9 being reported as married) were detected, and if they could not be corrected automatically, they were entered in an

"edit report" for correction at a later date. For the cases of nonresponse for age, sex, marital status or employment status, the response was imputed whenever possible by using other information on the census schedule. When this could not be done, the response in question was randomly imputed. For households which could not be contacted or which refused to supply data, dummy households were imputed, usually based on information supplied by the census collector.

Following the 1981 Census, the Australian Bureau of the Census conducted a PES to estimate the undercoverage of persons and dwellings. A multistage area random sample of approximately 35,000 private dwellings was employed, and all usual residents of the selected dwellings were included in the survey coverage unless they were absent from their usual residences for more than 1 month. Visitors were excluded unless they were absent from their usual residence for more than 1 month, there was no-one at their usual residence during the survey period, or they usually lived in a nonprivate dwelling.

The PES began three weeks after Census day, and highly trained staff intensively interviewed approximately 95,000 persons. The data obtained from the interviews was then matched to the census questionnaires to determine the proportion of persons in the PES who were not enumerated in the Census and the proportion of multiple enumerations in the Census. Any census questionnaires that were received by mail after the PES interviews had commenced were treated as missing for the PES purposes, although the data from the questionnaires were included in the final statistics. Matching these questionnaires with the PES data would have introduced bias into the undercoverage estimate, since it is likely that the PES may have prompted respondents to return previously incomplete questionnaires which would otherwise not have been returned. The net undercoverage rate in the PES sample was calculated to be approximately 1.9%, and it was then used as an estimate of the Census undercoverage rate for the total population.

The PES interviews provided data on names, sex, ages and marital status, country of birth, usual place of residence, location at census night, addresses before and after census night and any other addresses which may have been included on a census questionnaire. The PES provided estimates of the undercoverage rates by age and sex for Australia, and by age and sex for each state and territory. The corresponding undercoverage rates for smaller areas were estimated by a mathematical procedure called iterative proportional fitting.

India

For the 1981 Indian Census, the editing was conducted in three stages. In the first stage, the individual slips were edited before the basic data was prepared for tabulation by village and urban block. In the second stage, detailed editing was performed manually before the sample schedules were coded. The third stage of editing was done at the input preparation stage on direct entry systems and on the computer by editing the tape.

An exploratory statistical quality control programme was restricted to the coding of responses to the economic questions on the individual slips. Samples were selected from 8 metropolitan areas, rural and urban areas in 5 districts and 3 union territories and the coded responses were checked for accuracy. The primary purpose of the operation was to screen out errors which had been made when coding the data so that only a tolerable proportion of errors remained in the data which had been passed for further processing. It was hoped that the quality control programme would provided insight into the causes of coding errors, and indicate measures which could be adopted to reduce such errors in future censuses.

The coverage error and content error of the 1981 Census of India was assessed in a PES which involved a sample of 4,000 blocks covering 15 states and the Union Territory of Delhi. Each state

was divided into 3 substrata: rural, noncity urban and city; and proportional sampling was used within each strata. All houses in the sample blocks were re-enumerated, matched and reconciled with census records to estimate the coverage error. Omission or duplication of households was classified as a Type I error, and omission or duplication of an individual was classified as a Type II error. The content error of the individual slips, the population record and migration and fertility data were estimated by sampling 10% of the houses (about 15 houses) in each block. Following the PES survey, any unmatched records were reconciled by revisiting the field.

Because of the high level of illiteracy in India and the diverse languages and religions, the census questionnaires are completed by enumerators who interview the respondents, using the local dialect. Consequently, the way the questions are phrased by each enumerator and the way the question is interpreted by each respondent will influence the response given. Moreover, the information sought may not be known. For instance, the respondent may not know his age according to the Julian Calendar, and may give it according to some other calendar. Or the householder may not know the correct ages of every member of the household; this would be particularly true in the case of extended families.

A census evaluation study (CES) was also undertaken to quantify the extent of omission or duplication of younger-aged children and to determine the accuracy of age-reporting of children in the census. 50 SRS (Sample Registration System) units in rural areas and 25 SRS units in urban areas in each state were canvassed with the CES schedule. Approximately 1,200 SRS units were involved in the CES. The birth records the from Sample Registration System were then adjusted for deaths and migration, and used to produce a list of children in the SRS units who had survived to the data of the census. This alternative population list was then compared with the census counts for the corresponding age groups.

Chapter 7

QUALITY CONTROL: CODING AND EDITING

Introduction

Because any published data which is identified as incorrect will bring into question the validity of the entire data set, it is essential that errors in the data are identified and corrected wherever possible. Quality control procedures must be instigated at each stage of the census operation. In addition to the pretests, pilot tests and pre-censal field checks which have already been discussed in detail, the natural order of quality control phases of the census operation is as follows:

Initial editing for completeness of the questionnaire and patently inconsistent information. Initial editing is conducted by field staff if questionnaires are collected or by receiving offices if a mailout/mailback system used.

Coding and editing. The procedure of coding is conducted solely to facilitate the data entry phase of the operation. If the data editing phase is to conducted manually, then it may be performed prior to the coding phase. However, most statistical agencies have automated the editing (and imputation) phases, and hence conduct the coding phase first. Data items on each questionnaire are coded, checked for coding errors, then screened to detect inconsistent information. The latter phase is referred to as **editing**. Should the coding phase also be automated, the coding and editing phases can be combined into a single operation. Automated coding is referred to as **Computer Assisted Coding (CAC)**.

Data item imputation checks (if required). If the decision is made to "manufacture" values for missing or

inconsistent data items, some quality control procedure should be conducted to ensure that these imputed values are "reasonable". Such a procedure is actually another editing procedure.

Post-censal coverage and content error evaluations. An overall assessment of the underenumeration and overenumeration rates, and the quality of the data collected.

Total nonresponse imputation checks (if required) to ensure the data values imputed for the whole questionnaire are "reasonable".

Final check on table presentation and data confidentiality.

Generally, these phases overlap, and it is common practice to reduce the time frame from enumeration to publication of census data by beginning each quality control phase as soon as is practicable. However, because several of the procedures are iterative operations, it is not always possible to begin a subsequent quality control procedure until the previous procedure has been completed.

Random checks on enumerators' work and linkage to questionnaires from previous censuses or surveys can be used to identify erroneous or fictitious responses or enumerator bias. Linkage to related questions within the questionnaire itself may also identify discrepancies in responses. However, editing errors will occur, and some data items will be rejected as invalid when they were in fact true. Editing errors are difficult to detect, as the data items selected by the screening process will be unusual values, that is to say, they will not follow the usual pattern of responses. Such editing errors can only be minimised by requesting verification of the data from the particular respondents, or linking the census questionnaires with questionnaires from previous censuses or other surveys

and checking that the information from the various sources is consistent.

Coding procedures for data which have been accepted as valid will not always work as anticipated, and coding errors will occur. As is discussed in more detail below, the most efficient way to reduce the amount of coding errors incurred is to independently check coded questionnaires. This is a procedure known as "double coding".

There are several methods of estimating the coverage achieved and the quality of the information collected in a census. These are known as post-censal checks, since they can only be undertaken after all processing of the census data has been completed. The most commonly used post-censal checks include comparison of census counts with estimated population counts, separate surveys and administrative data. The amount of overenumeration (multiple enumeration) can be assessed by checking that information on each member of the population has only been supplied from one source, whether it be a personal questionnaire or a householder's schedule. Underenumeration rates and content error can be estimated by matching census questionnaires with questionnaires from earlier censuses or survey questionnaires. Many of the problems involved with post-censal coverage and content error checks are discussed in the final section of this chapter, but it should be noted here that there is no known method of accurately estimating the success of a census, since the exact count of the population of the population will always be unknown. Any population estimates, used as a benchmark against which to compare the census counts, are themselves prone to undercounting and/or invalid assumptions.

Data Coding

The purpose of coding data is to convert the responses to questions into a format which can be rapidly assimilated into a

large data file, prior to processing and analysing the data. When coding is performed manually, the response to each question is compared with a coding reference list or manual to determine the appropriate code for that question, and the code is then transcribed onto a coding sheet or onto the questionnaire itself.

During the coding phase, three types of errors may arise: **factual errors**, **interpretation errors**, and **writing errors**. Factual errors are usually caused by lack of concentration on the part of the coder. The frequency of such errors can be reduced by careful selection, training and motivation of coding staff, and ensuring that coders take regular, frequent, short breaks. Interpretation errors occur when coding schemes are inadequately detailed. For example, the responses to questions on employment are often difficult to categorise, since many borderline cases will arise, requiring a comprehensive coding scheme. Writing errors include transcription and transposition errors, and occur when the handwriting of a respondent is not clearly legible or the coder is not paying sufficient attention. Transcription errors occur when one or more digits are misread or miscopied, and transposition errors occur when one or more digits are interchanged.

During any coding phase, it is necessary to ensure that double coding is performed to detect as many coding errors as possible. The two basic methods of double coding are **clerical review** and **blind double coding**. During a clerical review, coders scrutinise pre-coded questionnaires, checking for valid codes and legible digits. Particular care is necessary to identify codes which, although plausible, are not correct, as such errors will not be detected during subsequent computerised checks. For blind coding, each questionnaire is independently coded by two coders, and the two completed documents are then compared by two persons reading them to each other, reconciling inconsistencies by discussion. While any form of double coding is time consuming and expensive, it

is essential that errors introduced during the coding phase are kept to a minimum. Salmond (1981) quoted an experiment conducted by the Management Services and Research Unit of the New Zealand Department of Health, in which 563 replies to a questionnaire were manually coded, and a random sample of 300 of the questionnaires were then clerically reviewed. The data from the questionnaires were then checked again using computerised editing procedures. The frequency of undetected errors in the single-coded questionnaires was found to be more than five times that of the clerically reviewed questionnaires.

Unfortunately, the more detailed the coding scheme, the longer the coding phase will take. As accurate, comprehensive and detailed data is of paramount importance, the only viable method of reducing the time required for the coding phase is automation. Appendix 7.1 describes the data preparation and coding procedures developed for New Zealand censuses, including the Computer Assisted Coding process which was used for the 1986 Census.

Computer Assisted Coding (CAC)

Because the coding phase of any operation is a repetitive, monotonous and time-consuming process, many statistical agencies have automated their coding procedures. Computer Assisted Coding is a procedure which automates as much as possible of the clerical reference to a list of descriptions and codes, and the subsequent allocation of a code to the data. CAC is particularly useful when coding responses which frequently do not fall into natural categories, such as *Address, Occupation* and *Ethnic Origin.*

Using CAC, the coder enters the supplied description through a VDU (Visual Display Unit). If there is only one possible code for the supplied description, it will be displayed on the screen and stored within the computer for later use. Where alternative codes exist, the coder is asked to either select from

a range of codes displayed on the VDU or to enter further descriptions to reduce the number of alternatives.

Because CAC does not require the coding staff to refer to coding manuals, provided that coding staff are regularly given short and frequent breaks from their work, then the rate at which the data is processed should improve dramatically, accompanied by a marked decrease in the incidence of factual errors. The rate of interpretation errors should also decrease, since the necessary information for allocation of the appropriate code is obtained by the coder answering prompts which are displayed on the VDU. Transcription and transposition errors will still occur, but their frequency will also be reduced, since provided the information is entered correctly, automatic coding ensures that only valid codes are accepted.

The rate of unmatched codes occurring through incorrect spelling can be reduced by applying algorithms to the supplied description and to the reference list. Because CAC can be self-correcting if information entered is misspelt, particular accuracy is not required when typing in data such as residential areas or street names. Such an automatic "spelling corrector" will mean that the amount of re-entering of descriptions is kept to a minimum, and provided that the algorithm requests verification of the assumed "correct" spelling by the person entered the description, no bias will be introduced by the application of the algorithm. Where the range of codes is small, the entire set can be displayed, along with the descriptions, and then the appropriate code selected.

A further advantage of CAC is that, should the coding scheme be found to lack sufficient detail, a modified algorithm can be implemented immediately it is incorporated into the computer system. Should it be necessary to alter codes, any such changes will be immediately implemented by all coders. In some cases, it may be viable to have the computer program amend the codes already entered; in other situations, it may prove more practical to re-process the relevant questionnaires.

Whatever the decision, because revised codes are immediately implemented using CAC, the amount of recoding will be kept to a minimum. When coding is performed manually, time delays are inevitable between the decision to alter a code, and the production and distribution of copies of the revisions to the coders. Further time will be lost while the coding staff adapt to the revised system. When the code in question has been memorised by the coders, considerable effort is required to ensure that the revised code is used. Failure to distribute copies of the amendments to all the coding staff, or failure on the part of some coders to assimilate the amendments into the coding scheme will entail ongoing coding errors and additional time delays before the amended codes are universally implemented.

To further reduce the frequency of coding errors, a prompt which requests the coder to verify that the selected code is correct could be incorporated into the computer program. However, since it is still possible for coding errors to occur when CAC is implemented, all questionnaires should be blind double coded. The computer can then be used to compare the independently assigned codes, and to tag any conflicting codes for further investigation.

The automation of the coding phase can also be expanded to include the editing and imputation phases. For example, if some questions on the questionnaire are related to other questions, information entered for the current response can be checked to ensure that it is consistent with the responses to earlier, related questions. It is also possible to link the questionnaire currently being processed with corresponding questionnaires from other surveys or previous censuses to supply missing information, to validate responses, or to refine the subsequent coding of occupation. Such linkage often proves to be more economic than matching of questionnaires at a later date. For example, when coding a respondent's place of employment, the details of a respondent's employing organisation can provide the information required, and Personal

Questionnaires for farmers and farmworkers can be linked to the farms held on the Agricultural Register.

Special Coding

Special coding may be necessary to ensure the accuracy of key variables. For example, sex and age may be used as reference variables during the editing phase, and hence must be as free from error as possible. Two special coding methods frequently employed are **redundancy** and **check digits**. The questionnaire can be designed to include questions which, although phrased differently, ask for the same information as previous questions. The purpose of such redundant questions is to check the accuracy of the responses and also to check the accuracy of the coding. For example, one question may request the respondent's date of birth, while a later question requests the respondent's age. Discrepancies can be resolved by correcting any obviously incorrect codes, and where the supplied age conflicts with the date of birth, the usual procedure is to recalculate the age from the coded date of birth at the computer-assisted editing stage. However, including redundant questions will reduce the amount of information yielded if a limit is placed on the total number of questions asked or on the length of the questionnaire, or could risk affecting the quality of the data collected if the questionnaires are made unnecessarily long by the inclusion of redundant questions.

Check digits are usually employed to verify the accuracy of identification data for when studies involve following individuals over a period of time. As mentioned later in this chapter, several of the post-censal coverage evaluation procedures involve matching census data with data collected from samples selected prior to or subsequent to the census, or matching census data with alternative population lists. The check digit is usually a single digit which is added onto the beginning or end of an identification number already present in

data records. Because the identification number will be used to trace individual respondents over a period of time, it must be both unique and constant over time. For this reason, variables such as age, height or weight would not be suitable as identifiers, as they are neither unique nor constant.

As an illustration, suppose we decide to derive the identification number for each respondent from their full name, sex and date of birth. The respondent's name and sex could then be numerically coded and combined with the date of birth to produce the desired number. Let us further suppose that the respondent's full name is Jill Lucy Brown, and that she was born on 22nd March, 1948. As the surname is generally used as the first part of an identifier, we will write it first, followed by the Christian or first names. Using the simple coding procedure of replacing each letter in the name by its corresponding position in the alphabet, and coding the sex classification as '1' for 'male' and '2' for 'female' yields

B	R	O	W	N		J	I	L	L		L	U	C	Y	Female	22/03/48
02	18	15	23	14		10	09	12	12		12	21	03	25	2	220348

which will produce the identification number

0218152314-10091212-12210325-2-220348.

The above example is merely intended to provide a simple illustration of coding, and there are undoubtedly many better ways of obtaining an identification number. In practice, such a large number would be reduced to a more manageable length, while still retaining its uniqueness; the second Christian name would not normally be included, and either the first Christian name or the sex would be redundant for the purposes of producing an identification number.

The check digit is calculated by multiplying each digit in the identification number by a number (weight), and then adding the products together. The resultant sum can then be reduced to

one digit using a modulus (the remainder left after dividing the sum by the specified number). This check digit must be unique for any identification number. The check digit is recalculated during the computer-assisted editing phase and compared with the original value, and any discrepancies due to coding or data entry errors resolved.

Clare Salmond (1981) cited the following examples which produce the same check digit for both the correct and incorrect numbers when modulus 11 is used:

Weights	5	4	3	2	1	Weighted Sum
Correct number	5	6	0	0	6	55
Incorrect number		5	6	0	6	44

Since 11 divides evenly into both 55 and 44, the check digit for both numbers is 0, and hence the error would not be detected.

Weights	7	6	5	4	3	2	1	Weighted Sum
Correct number			5	6	0	0	6	55
Incorrect number		5	6	0	0	0	6	66
	5	6	0	0	0	0	6	77
				5	6	0	6	44
					5	6	6	33

Again, 11 divides evenly into all of the products, producing a check digit of 0, so none of the incorrect numbers would be detected using the above scheme.

Data Editing

Editing can be defined as the detection and identification of errors in a data set, whereas **imputation** is the "correction" of such errors by substitution of values thought to approximate the missing values. The role of the edit process is to alter

erroneous fields, and not to alter valid ones. A record must be accepted, even if it fails several statistical edits, if information from ancillary sources or from the record itself supports its validity.

As was mentioned in the above section on data coding, the questionnaire can include redundant questions to check both the accuracy of the responses and to check the accuracy of the coding. However, the presence of redundant questions will either increase the respondent load or prevent the inclusion of other questions in the questionnaire. For this reason, redundant questions, if used, should be restricted to those necessary to validate particularly important questions.

Edits which should be employed as a matter of course are **range edits** and **consistency edits**. Range edits are performed by determining the admissible set of values for each particular variable, and identifying any response which is not in the admissible set. By excluding blanks from the admissible set, range edits will detect missing data as well as unanticipated values. At this stage, it is possible to stop the editing process and investigate the response in question, but the common practice is to tag the response and continue with the editing process.

After all the individual tagged values have been investigated and accepted or deleted or imputed, a further range edit should be performed, followed by checks for internal consistency. The accuracy of the information supplied by each individual can be checked to some extent by ensuring that the responses to some questions do not conflict with other responses.

Consistency edits usually require extensive knowledge of the subject matter of questions, as these edits must specify sets of values for specified combinations of variables which are jointly unacceptable. For example, it is highly unlikely that an 11-year old child would have had any experience of tertiary

education, was head of a household, or was married, divorced or separated.

After the consistency edits are completed, the tagged variables are investigated. Transcribing or key-punching errors which were not detected in the double coding are corrected, but the remaining impossible or unlikely values must be carefully investigated. Unless it is possible to reinterview the respondent without introducing any bias, impossible values are deleted and either coded as "missing" or imputed. Assuming further information cannot be obtained, Salmond (1981) advocated subjectively separating the unlikely values into those close to, although outside of, the tolerance limits (the set of acceptable values); those which are statistically unusual; and those which are highly improbable. Values identified as highly improbable will be deleted and either coded as "missing" or imputed, while the marginally out-of-tolerance values will be retained unless a very good external reason for suspicion is found. The statistically unusual observations are then subjected to statistical edits before deciding to accept, delete or impute each value. Again, after the necessary corrections to the data have been made, a further consistency edit should be performed.

Frequently, the editing and imputation of data are combined into one phase. Fellegi and Holt (1976) wrote the seminal article on automatic editing and imputation. Although their paper deals primarily with categorical (qualitative) data, some of the theoretical results also apply to quantitative data. **Quantitative data** are measurements, such as heights, weights and lengths. In contrast, **qualitative data** are data which cannot be measured, but which can be categorised. Survey and census data are usually qualitative, and responses are frequently assigned into pre-determined categories. For example, the sex of a respondent will be either male or female and the marital status of a respondent will be one of the following: never married, married, married but permanently separated, widowed, divorced. Two examples of quantitative

data obtained from censuses are the number of children born to a female respondent, and the usual number of hours worked per week. Fellegi and Holt refer to edits which only involve qualitative (coded) data as **logical edits**. Edits involving quantitative data are called **quantitative arithmetic edits**.

Fellegi and Holt recommend that the set of edit rules specified by subject matter experts should be simple and unrelated, thus permitting the addition of further edit rules or the deletion of some of the specified edits. The specified consistency edits are referred to as **explicit edits**, and they frequently logically imply additional **implicit edits**. For example, suppose that a questionnaire lists the categories of the variable *Marital Status* as *Single, Married, Divorced, Widowed or Separated* and the categories of the variable *Relationship to Head of Household* are listed as *Spouse, Daughter, Son or Other.* Furthermore, suppose that 2 explicit edit rules are that the following values for variables cannot hold simultaneously:

Age = 0-14 and *Marital Status = Married, Divorced, Widowed or Separated*

Marital Status = Single, Divorced, Widowed or Separated and *Relationship to Head of Household = Spouse.*

Then any value within the range *0-14* for the variable *Age* implies that the variable *Marital Status* can only contain the value *Single.* Similarly, any value in the set *Single, Divorced, Widowed or Separated* for the variable *Marital Status* implies that the value for the variable *Relationship to Head of Household* cannot be *Spouse.*

However, since any value within the range *0-14* for the variable *Age* implies that the variable *Marital Status* can only contain the value *Single*, and the value *Single* for the variable *Marital Status* implies that the value for the variable *Relationship to Head of Household* cannot be *Spouse* the following implied edit rule is also present:

The following values for variables cannot hold simultaneously:

Age = 0-14 and *Relationship to Head of Household = Spouse.*

While the initially specified edit rules are all that are necessary to identify the records which pass or fail the edits, Fellegi and Holt recommend that the logically implied edit rules should also be taken into account when systematically investigating variables to determine which (if any) variables should be changed to "correct" a data record. The procedure of generating implied edits will also identify any internal inconsistencies in the explicit edits.

When a record fails one or more edits, the editing procedure must then select the variables which, between them, appear in all the failed edits. Fellegi and Holt recommend that the data in each record should be made to satisfy all edits by changing the fewest possible items of data. Adopting this procedure will thus preserve as much as possible the frequency structure of the values of the variables. Furthermore, if the imputation rules are derived from the corresponding edit rules, there will be no necessity to re-edit the data after imputation.

The imputation method derived by Fellegi and Holt uses the distribution of current data to impute missing variables, and is often referred to as the **hot deck** method, since it does not use historical data or data from other sources. An incomplete record is matched with other complete responses, and the variables driving the match are the same ones used to edit those fields.

The impact of the contribution of Fellegi and Holt to the field of Statistics is attested to by the widespread adoption of hot deck methods by such agencies as Statistics Canada, the US Bureau of the Census, the US Social Security Administration, the US Internal Revenue Service and the New Zealand Department of Statistics.

Statistical Edits

As a first step, stem-and-leaf diagrams or histograms should be constructed for each variable to informally check their distributions. In particular, the distributions should be examined for skewness or bimodality. In addition, basic statistics, including the **mean**, **median**, **range** (or **interquartile range**), **standard deviation** and a skewness statistic, should also be calculated. Definitions of these statistics are included in the glossary at the end of the thesis.

Comparison of the mean and the median will indicate whether the data is skewed. If the data is very skewed, then neither the mean nor the standard deviation should be used in any statistical analysis. In such cases, the median and the interquartile range will be more appropriate statistics to use. In particular, data on age and income will always have non-normal distributions.

If a variable appears to have an approximately normal distribution then, according to the empirical rule, approximately 68% of the observations will lie within one standard deviation of the mean, approximately 95% of the observations will lie within two standard deviations of the mean, and almost all of the observations will lie within three standard deviations of the mean. This rule can then be applied to identify possible outliers.

If an outlying observation proves to be valid and is merely a result of the random variability of the data, it should always be retained in the data. Retention of accepted outliers may mean that statistical criteria which are based on the underlying assumption of normal distributions must be interpreted with caution or, alternatively, the employment of nonparametric tests. Regardless of the normality assumption, it is generally recommended that a low significance level, such as 1%, be used.

One test criterion for a single apparent outlier is the ratio of the difference of the "outlier" and the sample mean to the sample standard deviation, where both include the outlying observation. Grubbs (1950, 1969) gave critical values for this ratio, described a test involving a sample standard deviation omitting the outlier, and developed procedures for testing more than one outlier at one end of an ordered sample.

Dixon (1950, 1953) gave an alternative set of tests based entirely on ratios of differences between observations. The criterion used depended on the sample size. Reed et al (1971) gave a modification of one of Dixon's criteria which is independent of a set of critical values:

> Reject the largest observation if the distance between it and the next largest observation is more than one third of the range.

Remembering that a "rule of thumb", often referred to as "The Empirical Rule", is that for most mound-shaped real-life distributions,

> 68% of all observations lie within 1 standard deviation of the (arithmetic)mean,
> 95% of all observations lie within 2 standard deviations of the mean, and
> almost all observations lie within 3 standard deviations of the mean.

Depending on the sample size and the shape of the distribution, the standard deviation can be roughly estimated to be anywhere between $\frac{\text{range}}{4}$ and $\frac{\text{range}}{6}$.

Using the upper limit, 2 standard deviations would be approximately $\frac{\text{range}}{3}$. Since the empirical rule tells us that 95% of all observations lie within 2 standard deviations of the mean, we can be sure that only unusually high or unusually low observations will be further away than $\frac{\text{range}}{3}$ from the next largest (or smallest) observation.

Such a rule is undoubtedly only a rough measure, but it is very simple, and can be used as a screening device to determine whether a more formal investigation for outliers is required.

Once observations have been identified as statistical outliers, there are no set rules as to how they should be treated. The following options are available:

1. All of the outliers could be deleted from the data set. This would reduce the skewness of the data, and could result in the edited data set having a more more normal distribution.

2. A sensitivity analysis could be performed to examine the effect of removing each outlier in turn from the data set. If the temporary removal of a single outlier results in the distribution bearing a closer resemblance to a normal distribution, then the decision may be made to delete that outlier from the data set.

3. All outliers could be retained. In such cases, if the data set is very skewed, then the median should be used as a measure of location instead of the mean, and the interquartile range would be the better measure of the spread of the data set. Any statistical tests on highly skewed data should only include **nonparametric** tests; tests which do not involve the mean or standard deviation.

Whatever the final decision, outliers should always be reported, accompanied with information on the extent to which outliers have been included in, or deleted from, the data set. Any analyses of the data should be accompanied by explicit statements of any assumptions made, such as normality of the data.

Potential Future Developments in Data Entry and Editing Procedures

With the rapidly increasing usage of computers as a medium for storing, manipulating and analysing data, it is now possible to seriously consider the prospect of field staff using hand-held terminals to enter census data, or by the respondents themselves. Alternatively, since many countries already have established computer networks, it may be possible for persons with access to personal computers to link directly into a computer program which interactively requested the information and recorded the responses.

A programme could be supplied which prompted the users (whether field staff or respondents) as to the next required response, and the programme could also perform an initial editing of the data. Any data value entered which was not within a previously defined range could trigger a prompt which would request verification of the response. If the response was subsequently confirmed as correct, then it could be provisionally accepted, but tagged for further investigation, such as statistical editing and potential imputation.

It would also be possible to extend this editing phase to a more comprehensive editing operation which included checks for consistency between the current response and the responses to previous questions. For instance, if the response for *Age* was a value less than 16 years, then an entry of *Separated, Married* or *Divorced* would be rejected for the variable *Marital Status*. Such an operation would undoubtedly minimise the amount of imputation required, since many queries could be resolved "on the spot". However, while this would an be exciting prospect for those involved in the Quality Control phase of the census operation, care would need to be taken that exposure of the respondents to such a detailed editing operation did not affect the attitude of the respondents to the census. Some respondents would resent having their responses challenged,

while others would be daunted by the prospect of using a hand-held computer terminal.

Chapter 8

QUALITY CONTROL: IMPUTATION

Introduction

In earlier chapters, it was suggested that the response rate to a census can be markedly improved by carefully selected questions, good questionnaire design, thorough pilot testing, stringent enumeration procedures, extensive publicity campaigns and continuous quality control checks on the performances of the enumerators. However, despite all efforts of the census staff, a 100% response rate and 100% accurate data will never be achieved.

Post-censal coverage and content error evaluation programmes will generally provide some estimate of the accuracy of the data and the coverage achieved. Because the census data is used for such important purposes as determining the sites of new industries, allocation of government funds and resources, revision of electoral boundaries, revision of the number of political seats and estimation of future population numbers, it is common practice in many countries to adjust the census counts for undercoverage, and to impute data for missing or erroneous information.

In this chapter, we discuss various methods of data imputation. We will briefly discuss problems inherent in any imputation procedure, and then cover some of the imputation procedures used for incomplete data records and for detected total nonresponse. Imputation for item nonresponse is often referred to as **data item imputation**, whereas imputation for total nonresponse is referred to as **total imputation**.

Treatment of Missing or Inconsistent Data

The process of detecting and identifying errors in the data, referred to as editing, is a noncontroversial practice. However, there is widespread controversy as to how missing or patently erroneous data should be treated. In the United States of America, follow-up on Questionnaires which have been rejected during the editing phase is conducted over the telephone, and households which have not returned their Questionnaires are visited by the Enumerators.

Ideally, the correct information should be obtained from the respondent. However, there may be too many cases of partial nonresponse to maintain an effective follow-up programme, the respondent may not know or may not wish to provide the required information, or time and financial restrictions may prevent any follow-up operations. Another factor to be borne in mind is that further contact with the respondent may generate hostility, particularly if the respondent feels that he is being hounded by census staff, or that his privacy has been violated. In such a situation, not only will any responses to further questioning be doubtful, but the respondent will be antagonistic towards future censuses.

If reinterviewing the respondent is not considered to be possible or advisable, the response in question could be recorded as a missing value, the entire data record could be deleted, or a value for the response could be 'manufactured' (imputed). For example, several other records with similar characteristics could be selected and the average value of the relevant field of all of these records calculated and inserted into the missing field of the record in question.

Deleting entire records from data sets because one or more fields are inconsistent, invalid or missing incurs substantial loss of information, as data contained in the valid fields is discarded. Moreover, deletion of entire records will compound the underenumeration which will undoubtedly already exist.

Recategorising inconsistent or invalid fields as "Not Supplied" or "Unknown" while leaving the remaining data fields intact will often create problems, as many statistical analyses require complete data sets. Should users of census data elect to impute values for missing data fields, they will not have all the information available to census staff, and the delayed imputation phase may not be as successful as imputations which are effected during the data processing phase.

Even if partially complete data sets are acceptable, the analysis will probably yield conclusions different to those which would have been obtained had the full data set been available. Values of data supplied by respondents will often be quite different to those that would have been supplied by nonrespondents, had they been persuaded to cooperate. Analyses obtained from incomplete data sets should be treated with caution, as the estimates obtained will almost certainly be biased.

Imputation may be preferable if it reduces the response bias and preserves the relationships between variables as much as possible. Examination of other data records with characteristics similar to those contained in the valid fields of incomplete records may well provide data items which could be substituted into the fields in question. If, for example, age or racial distribution is to be used to determine the current and future needs of educational and medical facilities, then imputation of data items may be preferable to using an incomplete data set as a base for planning for such facilities, as experience has shown that both minority populations and young adults have suffered higher rates of underenumeration than other population groups.

However, imputation procedures also have their drawbacks. The same criticisms of using data supplied by respondents to estimate missing or erroneous data apply to imputation. Because the true data values are unknown, imputation errors are usually difficult to detect. Of course, there is no way of

knowing whether the overall picture so obtained will be an accurate picture of the population. What is more, estimates obtained using imputation will be less reliable than if they were based on a complete set of real data, and the usual variance estimates are inadequate because they do not include the unknown error due to imputation.

Statisticians will probably always differ over the issue of whether or not to impute data, and what method of imputation should be used. Opponents of imputation argue that, as there is no way of testing whether an imputed value is what the respondent would have supplied had he or she correctly completed all the questions, then no attempt should be made to fabricate a response.

If the questionnaires are collected by field staff, an initial check for completeness will reduce the amount of editing and imputation required. No matter how sophisticated the procedures for editing and imputation, there can be no guarantee that the data produced is correct. Hence the manual collection of questionnaires may prove to be relatively cost-effective, since it will undoubtedly encourage respondents to complete the questionnaires as fully as possible.

Problems Inherent in Any Imputation Procedure

Follow-up surveys and other evaluation studies have shown that the distribution of data which would have been supplied by nonrespondents, in terms of questions which were partially answered or completely omitted, generally differs from the distribution obtained from fully completed questionnaires. Data obtained from census nonrespondents by means of personal visits from enumerators and follow-up surveys often has a different mean and variance, since it has more extreme values than that obtained from those persons who submitted complete questionnaires. For example, many people over the age of 35 years, particularly women, are reluctant to disclose their age;

persons on social welfare benefits may be reluctant to report any casual or part-time employment in case such information will result in the loss or reduction of their benefits, and others who have unofficial or secondary jobs or investments may be reluctant to pay a higher tax rate which would be applied should their true total income be revealed.

Extensive research has been conducted into nonresponse procedures. Unfortunately, there is no known method of imputing for nonresponse which yields estimates which accurately assess the true data values. It is important that any imputation procedure used yields estimates which are consistent (that is, the method doesn't produce unexpected values) and are as close to the true values as possible.

One common method of assessing the performance of an imputation procedure is to take a complete data set in which all the values are known, and to remove some of the data items and treat them as missing values. The imputation procedure is then conducted, and the values obtained are then compared with the original data values. Frequently, several alternative imputation procedures are assessed using this method, and the procedure which yields the most consistent results and the minimum nonresponse bias, should such a procedure exist, is then selected.

Various methods of deciding which values to delete have been used, including **Monte Carlo** methods and adopting known nonresponse patterns in data. Monte Carlo methods are based on the assumption that the numbers were obtained genuinely at random although, in practice, "pseudo-random" numbers are selected, since the selection of random numbers is too slow for digital computers. Since the latter method of adopting known nonresponse patterns uses a realistic data mechanism to delete data values, it can produce a better assessment of how an imputation method would perform when applied to the data set under consideration.

Another popular method of assessing imputation procedures is to compare, for each variable, the distribution of the values supplied by respondents (stated values) with the distribution of stated values plus imputed values. However, these distribution of nonrespondents may be different to those of the respondents. Moreover, the higher the amount of imputations performed, the greater the effect on the distribution of stated values plus imputed values.

According to Pritzker, Ogus and Hansen (1965), provided that the nonresponse rate is less than 5%, most nonresponse imputation procedures will provide acceptable results. However, as the nonresponse rates for most surveys usually exceed 20%, the method of nonresponse imputation may have a marked effect on the survey estimates. Unfortunately, even surveys which include intensive follow-up procedures to reduce nonresponse still experience high nonresponse rates. As mentioned in the discussion on censal coverage, the undercoverage rates experienced in censuses not only vary between the racial groups, but they also vary between geographic locations within racial groups. Moreover, the undercoverage rates are frequently far greater than 5%. Hence it is important that, whenever possible, imputation procedures which reduce or minimise the nonresponse bias are used.

Data Item Imputation

A cautionary note about the usage of imputation is that nonresponse stems both from random events, such as a simple mistake in overlooking the question or the mood of the respondent at the time, and from causes such as inhibitions affecting reporting and difficulties experienced by respondents in answering the questions. Respondents who are reluctant to divulge income information may refuse to supply the particulars, or they may lie and distort the information in order to evade reporting the true values. The former response is immediately identifiable as nonresponse, but the distortion of

information can only be inferred by checking the response with responses for other items, or by some aspects of the interviewing situation, such as the presence of some other person who can edit the information supplied. Unless such distortions are detected, the supplied information will be accepted as valid, and included as potential donors.

Perhaps the best approach to imputation is to consider a variety of "reasonable" models, to impute the values accordingly, and to observe how the results of the data analyses change under the varying models. The data analyses will invariably contain some summary statistics, including at the very least the mean and standard deviation, and also the median if the distribution is skewed - as will be the case for data on ages and incomes.

Comparison of these statistics, and examination of the variation incurred by employing the various models, will assist in the final selection of the appropriate model to use for imputation of the missing data items. Such an approach is known as **sensitivity analysis**, and it will provide some comfort to the data analysts if the analysis proves to be **robust**; that is, the choice of the model has little effect on the results of the imputation and any subsequent analysis of the data set. Depending on the method of imputation employed, it may be necessary to re-edit the data once the imputation phase has been completed to ensure that the imputed values do not conflict with other data items.

Whichever model is finally selected, any publication of the data and the results of any analyses should always be accompanied by a description of the method of imputation employed, the frequency of imputation, and any other relevant information. Failure to provide such qualifying information on the anticipated error margins of the data would be misleading.

The least controversial method of data item imputation is to incorporate questions into the census schedule which produce redundant information. Data obtained from the completed

questions on the questionnaire can then be used to infer the appropriate response for the missing data item(s). For example, if information on the respondent's age is considered to be important, two questions could be asked: the first requesting the respondent's age, and the second requesting the date of birth. Should the data supplied for such questions conflict, the usual procedure is to use the supplied date of birth to impute the respondent's age. Most people have memorised their dates of birth, and are more likely to misreport their age than their date of birth.

However, incorporating redundant information in a questionnaire entails either the questionnaire containing more questions, or restricting the length of the questionnaire by replacing one or more questions with the redundant questions. While the latter option entails sacrificing extra information for the purposes of quality control, both options may have a negative effect on the data collected, should the respondents realise that redundant information is being sought.

Alternatively, it may be possible to link the questionnaires to those from surveys or administration files which incorporated background variables recorded for both respondents and nonrespondents that were highly correlated with variables that were likely to be missing. For example, when data items on personal income are missing, linkage to data from surveys, administrative files held by government departments, or salary and wage files held by respondents' employers would provide much of the required information. However, the matching of cases will be both incomplete and subject to error, and the methods of collecting the data may differ; for instance, different definitions could have been used, or the questions asked may produce slightly different information. Moreover, there will be a significant difference in response between census and survey data, and the comparison values must be adjusted to account for that response before evaluating such imputation procedures. Another possible source for the information would be a linkage to the respondents' bank

accounts, but it is highly unlikely that the public would accept such a concept, and the antagonism generated by such a suggestion would undoubtedly adversely affect the quality of data supplied by the respondents.

Alternative approaches for data item imputation include:

1. equal-weights models
2. regression models
3. logarithmic models
4. ratio models
3. cold-deck models
4. hot-deck models.

Irrespective of the particular method selected, the first step in any imputation procedure is to check for any obvious errors. For example, a common experience in many censuses and surveys is for some data values to be misreported or miscoded by a multiple of one thousand. For example, some respondents may report a variable as a poundage, rather than as the requested tonnage. Other examples of common mistakes are weights reported in pounds rather than in kilograms, lengths reported in imperial measurements rather than metric, and temperatures recorded in Fahrenheit rather than in Centigrade. In such cases, the recorded response can be corrected by subject-matter specialists, without recourse to an imputation procedure.

Equal-weights models correspond in important aspects to a procedure of ignoring missing values and defining the sample size for a specific value of a variable as equal to the number of times that value was observed. The weight initially assigned to each missing value is redistributed equally over the sample elements in which that value was actually observed. In other words, equal-weights models effectively assume that, for each data item, the missing values have the same distribution as that of the data supplied by respondents. Values for missing

data items are imputed using the distribution of the same data items supplied by the respondents.

Regression models, as the name implies, use regression methods to estimate the missing data item from the other data items. However, such methods can only be used to impute quantitative variables. The data variables which are used for the basis of predicting the unknown variable are called the **explanatory variables** or **independent variables**. The predicted response is calculated as a weighted average of the explanatory variables. A variety of models can be used, ranging from the simplest model involving only one explanatory variable to the fuller models which incorporate several explanatory variables and full interactions between these variables. If a selected model does not include interactions, then it is assumed that there are no correlations between the explanatory variables; in other words, there are no linear relationships between the explanatory variables.

If regression models are used to predict values for missing data items, then several competing models may be fitted to a complete data set which was obtained from respondents and which has values for all the independent variables required for the prediction. By examining the differences between the predicted values and the actual data values, the best model can then be selected. However, such a method assumes that the respondents and the nonrespondents have the same regression relationships; in other words, there is no underlying difference in the distributions of respondents and nonrespondents.

Regression models usually impute a mean of the predictive distribution, which is conditional on the independent variables in the model. Even if the assumptions of the model are true, the distribution of the imputed values will have a smaller variance than the distribution of the true values, causing larger deviations between the predicted values and the (unknown) actual values. In an effort to reduce these deviations, random errors can be added to the predictive means. These errors can

be either random normal variates or residuals which are randomly selected from the model, referred to as empirical residuals. If it cannot be assumed that the residuals are normally distributed, then empirical residuals should be used.

However, while such relatively complex regression models can be used to improve the distributional properties of a single dependent variable, no regression models have been found which cater for cases where more than one data item is missing from a data record. Such cases require multi-variate methods, which are presently most successfully resolved using hot deck methods.

Logarithmic models are frequently used for data sets which are highly skewed, since taking the logarithm of data values will reduce the variation between the data values. Data on incomes is invariably skewed, and usually has the bulk of respondents in the low-medium income categories, with relatively few persons receiving abnormally high incomes. Transforming the data by taking the logarithm of the income values will reduce the skewness of the data set.

Logarithmic models are often incorporated into regression models. However, a drawback of using logarithmic regression models is that when the reverse transformation is used to relate the results of an analysis back in terms of the original data, large increases in the variation of the predicted values may occur. Little and Samuhel (1983) reported an approximate 15% increase in their predicted values when modelling the logarithm of wages and salary using simple linear regression. Another problem is that the relative importance of the predictor variables can be distorted by the excessive weighting of low income observations.

Ratio models. When considering economic data, a widely used criterion requires the ratio of two responses lie between prescribed bounds. The upper and lower bounds are determined by historical observation, subject-matter expertise, and, when

feasible, by a sample of responses. Ratio edits can incorporate data from an earlier time frame such as previous censuses or surveys, or from external files such as administrative records. However, data obtained from earlier time frames may need to be adjusted for inflation. If the questionnaire contains related questions, then one response may produce the upper and/or lower bounds for the second response. For instance, the New Zealand Census Personal Questionnaire contains a question on Social Security Benefits received by the respondent. The response to this question can then be used to place bounds on the following question on Income from Social Security Benefits.

Gordon Sande (1979) employed linear edits for the imputation of economic data. He used linear programming to determine a feasible region for each variable; that is, upper and lower bounds were determined for possible data values, and a record which fell into this region was accepted as a potential donor record. Sande used variables involved in the edits of a given variable (or set of variables) for matching with a donor record.

Ratio models, like logarithmic models, can be incorporated into regression models. Little and Samuhel (1983) used linear regression to model the wage rate per hour by dividing the usual predictor variable of wages and salaries by the product of the weeks worked and average hours worked per week, that is

$$\text{wage rate} = \frac{\text{wages and salaries}}{\text{weeks worked} \times \text{average hours/week}} .$$

Estimation of the wage rate, rather than the wages and salaries should reduce the skewness of the data, since the wage rate takes into account the average number of hours worked each week.

Ratio regression models may predict some negative values, which would be nonsensical if variables such as age or income were being predicted, and in such cases, the values should be set to zero. These models would also be susceptible to distortion of the predicted variable if the variable involved in

the ratios were inaccurately reported. For example, the wage rate modeled by Little and Samuhel (1984) relied on the accurate reporting of data on weeks worked and the average hours worked per week.

Cold-deck and **hot-deck models** provide a relatively low-cost and simple method of imputing missing values in a data matrix, and can provide a clear methodology for handling the multivariate problem by matching as many variables as possible with a donor record. They are discussed in more detail below.

Cold-Deck Imputation

Cold-deck imputation entails using values from some prior distribution to substitute for missing responses. Usually, the distribution used is derived from as similar a population as possible. A common approach is to match census data from previous censuses, from surveys taken from the same population, or from administrative files. The responses are jointly classified on one or more variables, the aim being to minimise the variance of the potential imputed values. This is achieved by defining the cross-categories in such a way that the values for a variable will be as similar as possible and values of different variables will be as dissimilar as possible.

On detection of an incomplete data record, values of other variables in that record are used to match with a record from a prior distribution. A search could be made for the first record from the prior distribution found to match the incomplete record, and the value contained in the relevant variable imputed into the incomplete data record, or the search could continue until all matching records in the prior distribution were identified. If several records from the prior distribution matched, then one of several values could be imputed into the missing data field. The value in question could be randomly selected, or the imputation could be generated from a regression equation or logarithmic or ratio models to impute a

single value from the several potential donor values. Depending on the design of the imputation procedure, a donor record could be considered as a potential donor for later imputations or it could be excluded from the pool of potential donor records.

However, matching a prior distribution with census data will be both incomplete and subject to error because of employment of different definitions, different coverage of questions and different levels of detail. Moreover, there will be a significant difference in response between administrative or survey data and census data, and the comparison values must be adjusted to account for that response before evaluating such imputation procedures.

Another criticism of this method is that the data obtained may be outdated, but perhaps the biggest problem is the lack of clear methodology for handling the multivariate problem; it is often not sufficient to consider imputing the data item (variable) in question without taking into account the effect the other data items will have on the variable. For instance, imputation of marital status will necessarily take into account age, and possibly also race and income.

Hot Deck Imputation

Basically, a hot deck method is a record-matching technique in which an incomplete record is compared with a complete record having similar characteristics. The missing value of the variable in the incomplete record is then imputed from the value which appears in the corresponding variable in the complete record. Some variations of the hot deck method are:

(1) The sequential hot deck method, in which the immediately preceding complete record is used as a donor to impute missing values to the incomplete record.

(2) The random hot deck method, in which all the complete and incomplete records are pooled randomly to make a hot deck. Missing fields in an incomplete record are imputed by selecting a donor at random from the complete records present in the hot deck. To prevent particular respondents receiving abnormally large weights, the restriction that each complete record may only be used once as a donor can be applied. In situation where incomplete records in a particular stratum outnumber the complete records, secondary imputation can be used.

(3) The distance hot deck method, in which the nearest complete record, which is not necessarily the immediately preceding record, is used as the donor. In the case where a missing value is equidistant from two equally eligible donors, the mean of the two donor values is imputed in the missing field. The distinction between the sequential method and the distance method is that for the latter, forward and backward searches are made for a suitable donor record, whereas the sequential hot deck method is restricted to a backward search.

(4) As for the cold deck method, a regression equation can be employed to calculate a value from eligible donor records.

(5) Usage of a moving average of values in a field to substitute for a missing value. This procedure would prevent extreme values form being duplicated, and would therefore slightly reduce the variances of the estimates.

(6) Usage of a grand (overall) average of values in a field to substitute for a missing data item.

Sometimes the order of the respondents has an effect on the responses, and in such cases, the sequential or distance methods would be preferable to the random hot deck method. They are also more convenient procedures for computer processing. As an example, suppose that questionnaires were

processed according to the location of the respondents on Census night. Direct comparison of information supplied for *'Full Address on Census Night'* and *'Usual Residential Address'* can be used to determine which persons in a household were in their normal place of residence. This information, in combination with *'Full Name'*, *'Age'* and *'Relationship to Occupier or to Person in Charge of Dwelling on Census Night'* can then be used to impute data items for any members of that household who did not supply their addresses. Similarly, values for *'Religious Denomination'* and *'Ethnic Origin'* can frequently be imputed using sequential or distance methods.

To improve the precision of hot deck methods, both the incomplete data records and the donor records can be categorised according to the amount of information which is available for the variables which are to be used for matching. Whenever possible, the nonrespondent (incomplete) record is matched with a donor record which has the same characteristics for the matching variables and the same level of detail for those characteristics. If no match can be found, a search is made for a match using weaker criteria; in other words, less detail is required for the matched characteristics.

Hot deck methods may be further classified into two categories for the imputation of the data:

(1) **Sequential imputation**, where a particular value for a variable of a record is imputed, taking into account the valid variable values in the record, then the next missing variable is imputed, taking into account the valid values, including the one which has just been imputed, and so on, until all the missing values have been imputed.

(2) **Joint imputation**, where missing values of variables are imputed simultaneously. This method takes into account information in the data record which may be correlated with the value of the variable being imputed. Joint imputation ensures that the incidence of combinations of

values will appear in the same proportions as in the population, thus preserving the joint distribution of the variables. Fellegi and Holt (1976) recommended that it be used whenever possible, with sequential imputation used as a default option; that is, used only when joint imputation cannot be achieved.

As mentioned in Chapter 7 in the discussion of data editing, if the rules for imputation are generated from the rules used in the editing process, then a further edit following the imputation phase to ensure consistency of responses will be unnecessary.

A Simple Example of Hot Deck Imputation

As an example of hot deck imputation, consider a data record containing the following characteristics:

Christian Name: Elizabeth *Sex* : Male *Age* : 12
Marital Status : Married *Relationship to Head of Household* : Wife
Education : Primary

The first step on any editing or imputation process is to check for coding or data entry errors. If we assume that the data has already undergone these screening processes, then we must decide which information in the data record conflicts with other information, and which data items are to be imputed.

In this data record, the data values for *Christian Name* and *Sex* conflict, as do the data values for *Sex* and *Relationship to Head of Household*, *Age* and *Marital Status*, and *Age* and *Relationship to Head of Household*. Since the response to *Christian Name* required a written answer, it is reasonable to assume that this information is correct, in which case the value for *Sex* is incorrect, and the data value *Female* will be imputed. As mentioned in Chapter 4, in order to reduce the time required for the completion of the questionnaire (whether by the respondent or by an enumerator) and to speed up the coding process,

wherever possible, questions on a questionnaire are answered by marking the appropriate pre-coded answer boxes. While this method undoubtedly accelerates the enumeration and coding phases, it increases the potential for content error, since it is easy for the wrong answer box to be marked. To minimise this risk, many data-collecting agencies ensure that the answer boxes are placed as closely as possible to the options available for each question.

Had the above data record also contained information on *Date of Birth*, this could have been compared with the value '12' for *Age*. Since most people can supply their date of birth without any memory recall or calculation, when data supplied for the variables *Date of Birth* and *Age*, it is standard practice to take the information supplied for *Date of Birth* as correct, and to use this data to impute a value for *Age*. However, since in this example we do not know the respondent's date of birth, we must decide whether or not we accept the data supplied for *Age* on the basis of the data values supplied for the other variables.

Since the data for the fourth and fifth variables, *Marital Status=Married* and *Relationship to Head of Household=Wife*, conflict with that supplied for *Age*, it seems reasonable to impute the value for *Age*. Note that we could have assumed that the response *Age* : 12 was correct, and imputed values for the variables *Marital Status* , *Relationship to Head of Household* and *Sex*. However, this assumption would entail imputing values for three variables, rather than two. Moreover, it seems more probable that the response for *Age* was wrong, rather than both responses for *Marital Status* and *Relationship to Head of Household.*

At this point, it becomes clear why imputation procedures normally use a sequential hot deck method or a distance hot deck method. Such procedures allow questionnaires from a single household to be processed together, facilitating the imputation of missing or invalid data items. In this example, comparison of data for *Surname*, *Age*, *Marital Status* and

Relationship to Head of Household from other questionnaires received from that particular household could be used to determine which variables in the above data record contained incorrect information, and would suggest an appropriate range of data values for incorrect data items. For example, if examination of related questionnaires confirmed that the information *Marital Status=Married* and *Relationship to Head of Household=Wife* was correct, then it would be reasonable to impute a data value for *Age* which was within 10 years of the husband's age.

First consider sequential imputation for this example. If we assume that the fourth and fifth variables are correct, that is, *Marital Status=Married* and *Relationship to Head of Household=Wife*, then the third variable must contain a value no less than the legal age for marriage. If we further assume that the data was collected for a New Zealand census or survey, and note that the legal age for marriage in New Zealand is 16 years of age, provided parental consent has been obtained, we can infer that the respondent must be aged at least 16 years.

The information supplied for the sixth variable, *Education*, may be correct, as although it is standard practice for New Zealand children to attend both Primary and Secondary Schools, it is possible that a child will only receive Primary Education. Clause 59 of the 1914 Education Act stipulated that every child between seven and fourteen years must attend school, and Clause 10 of the 1920 Education Amendment Act extended school age from fourteen years to fifteen years. "School age" is defined in the 1914 Education Act to be any age between the ages of five and fifteen years, and although most children begin their primary education as soon as they reach 5 years of age, they are not legally required to attend school until they are seven years old. On the basis of this information, we could decide to accept the data value *Education=Primary* or we could search all donor records for one with the same characteristics for the remaining variables. Should we find a suitable donor

record, the data value for *Education* could then be imputed into the above data record.

Whatever data value is finally accepted for *Education* for the above example, it will have no effect on the values of the other variables, since it will not assist us with imputing a value for *Age*, and it is not related to the other variables. (We had already determined that the data value for *Age* must be at least 16 years, and this is unaffected by the information contained in the variable *Education*.)

Applying the joint imputation method to this example, and again assuming that the values to be imputed are for the first and second variables, the procedure is to search the potential donor records until one is found which has the characteristic of *Married* for the third variable, a value of at least *16 years* for the second variable and the value *Female* for the first variable. The value contained in the second variable will then be imputed as the "correct" value.

Using either technique, if no match was found for the incomplete data record, the selection procedure could be weakened to accept any donor record which contained the characteristic of *Married, Separated, Widowed or Divorced* and whose remaining characteristics were as specified above.

In this example, using the sequential imputation technique, a second search for a donor record for the appropriate characteristics is unnecessary since, under the assumptions made, there was only one possible value for the first variable. However, the distinction between the two methods is that sequential imputation instigates a separate search for a potential donor record which satisfies the criteria for each field to be imputed , accepting the value which has just been imputed as "valid" when matching characteristics in successive searches. Joint imputation involves a single search for a donor record which simultaneously satisfies all criteria for the fields to be imputed.

Whatever variations of the hot deck method are being considered, their relative efficiency in estimating missing values should be examined under the following criteria:

(1) The quality of missing value estimates produced.
(2) The extent to which the estimates of the population means are biased.
(3) The degree to which the population covariance structure is retained in the imputed samples.

Hot deck methods are typically used with large data sets. When considering the robustness of such methods in small samples, the effects of imputation on sample covariances should be carefully examined.

Hot Deck Method Used to Impute US Tax Data

The U.S. Internal Revenue Service provides estimates of population and subpopulation totals for several hundred financial items from an annual sample of Corporate Tax Returns. The basic sample design is highly stratified, and since 1981, has been modified to include matrix sampling, where items not observed in the subsample are predicted using a hot deck imputation procedure.

Since a relatively small number of very large corporations dominate the estimates, the population under consideration is highly skewed. These very large corporations (identified as Group A returns) are selected with certainty, and stratified matrix sampling is used for the smaller returns (identified as Group B). Group A includes not only the very large corporations, but also corporations of any size for which it is believed these schedules are significant.

After editing the schedules,

Final Other Income = Original Other Income - Changes due to editing

and

Final Receipts = Original Receipts + Changes in Receipts due to editing Other Income Schedule .

Missing information in the subsample is imputed using a hot deck procedure within adjustment cells. A record with items to be imputed is matched to a record, in the same adjustment cell, with complete information. Since the original amount is available on all records, the relative change, rather than the dollar amount of the change is "hot decked", in an effort to reduce the coarseness of the hot deck procedure and to eliminate further corrections to balance the record.

For Final Other Income, the imputed value for the i^{th} record would be obtained by the following expression:

$$(1 - C_d) \times (\text{Original Other Income}_i),$$

where C_d is the ratio

$$\frac{\text{Change on donor record d}}{\text{Original Other Income on record d}}$$

and "d" is the donor record, with complete information, that was matched to record i as part of the hot deck process.

For variables such as Other Income, the lower and upper bounds on the relative changes are, respectively, 0 and 1. However, for a variable such as Receipts, there is no intrinsic bound on the relative size of the change (except that it cannot be negative). The amount being added need not have any relationship to the amount originally in Receipts; and original amount of zero can be changed to a nonzero amount, and even if a small amount is added to a nonzero amount, it can result in a large relative change. Hence significant changes in the microdata may be made, with potentially adverse consequences for estimates of subpopulations.

The Internal Revenue Service (1984) stated that, after preliminary analyses, the hot-deck imputation was not expected to significantly affect the estimates of important population and subpopulation aggregates, and it was hoped that the imputation procedures would not severely distort the distributions within microdata sets. In other words, for most variables and most subpopulations the distributions should not be too distorted.

Hot Deck Method to Impute US CPS Income Data

In the early 1960's, the U.S. Bureau of the Census ignored nonresponse when publishing CPS data on incomes. Although the nonresponse rates were relatively low, the estimates were biased as the respondents were not a random subsample of the sampled individuals. Since 1962, a hot deck procedure of assigning the income of a matched individual to each person who did not report his/her income has been used. Several variables are used to define a match, and the treatment of multiple income items has been modified to preserve their covariance structure.

The imputation scheme for earnings currently used by the U.S. Census Bureau initially classifies nonrespondents into one of eight groups, according to the combination of data values which have been supplied. Data records in Group One have data for 10 variables, and the only omitted information is Earnings. The 10 variables use for matching are Sex, Age, Race, Education, Relationship to Head of Household, Weeks Worked Last Year, Full-Time/Part-Time Status, Occupation-Industry, Class of Worker, and Earnings Recipiency. Data records in Group Eight only have information on Sex, Age and Education. Within each group, data records are further categorised according to the level of detail of the supplied information. Within each particular level of detail of a group, called a cell, nonrespondents are matched with respondents with similar values of the variables available for that group. The

nonrespondent is then assigned the matched respondent's values of the missing items.

If no match on all these items can be made, matches are made at a lower level of detail, by reducing the number of categories in the variables or by omitting some matching variables. The quality of the matches varies considerably, depending on the availability of suitable respondents for matching.

As noted by Lillard, Smith and Welch (1982), the hot deck method of imputation within each level of detail is similar to fitting a fully interactive analysis of variance, and then imputing the predicted mean plus a residual which is obtained from observation, rather than from a formula. Such a residual is referred to as an **empirically based residual**. To illustrate why this is so, let us take the simple case described by Little and Samuhel (1983) where three categorical variables X_1, X_2 and X_3 are used to define a match, and y_{ijkr} is the earnings for respondent r in a cell with $X_1 = i$, $X_2 = j$ and $X_3 = k$.

If a nonrespondent in this cell is matched to respondent m, the hot deck imputed value y_{ijkm} can be decomposed as

$$y_{ijkm} = \bar{y}_{ijk} + r_{ijkm} ,$$

where $\bar{y}_{ijk}$ is the predicted mean value from estimating μ_{ijkr}, the expected earnings in the cell, and

$$r_{ijkm} = y_{ijkm} - \bar{y}_{ijk}$$

is the residual for a respondent chosen randomly from the $(i,j,k)^{th}$ cell.

Recalling that a fully interactive regression model specifies

$$y_{ijkr} = \mu_{ijk} + \varepsilon_{ijkr} ,$$

where μ_{ijk} is the expected earnings in the cell and ε_{ijkr} is a random error, the comparison becomes obvious.

Since the model is fully interactive and interval-scaled variables such as age are grouped into categories, the form of

the equation used to relate the earnings to the predictors will depend on the selection of categories.

Advantages and Disadvantages of Hot Deck Imputation

Hot deck imputation assumes that respondents and nonrespondents have the same distribution within the cell defined by the matching predictors. Moreover, the precision of the imputation may be compromised by omitting detail from the model. Lillard, Smith and Welch give an example which clearly illustrated the effect that lack of detail may have on the precision of an imputed value. When detailed occupational coding was used to impute the mean income for nonreporting white male lawyers in the 1980 CPS, matching produced the imputed value $33,448, whereas when only one-digit coding was used for the match, the imputed value plummeted to $15,594.

Because nonrespondents from rare population subclasses with particularly high or low values of a variable tend to be difficult to match, they are pulled towards the mean of the distribution of the variable by the lack of detail at the level a match is made. Thus, while the overall mean of the imputed distribution will probably be a reasonably close estimate of the mean of the nonrespondent distribution, the variance of the nonrespondent distribution will be underestimated.

Another problem with implementing Hot Deck imputation is that a donor for imputed values may be used more than once, effecting abnormally large weights for particular respondents, and thus reducing nonresponse bias at the expense of increased variance of estimates in repeated sampling. This problem can be avoided by permitting each complete record to be used only once as a donor.

Hot deck methods also assume that values supplied by respondents are free from response bias. In other words, that reported values have not been deliberately distorted in order to

evade reporting the true values. Should such a value be selected as a donor, the effect of the response bias will be further increased.

However, one distinct advantage of hot deck imputation is the relative ease with which multivariate joint distributions are handled. In contrast to regression-based imputation models, the hot deck methods can simultaneously produce imputed values for more than variable, preserving the joint distribution of these variables.

Compromise between Hot Deck and Regression Imputation

The essential difference between the above hot deck method and a method based on a regression model is that the hot deck method matches the incomplete record with a donor record which has corresponding characteristics for the variables selected for matching. As mentioned above, problems occur when there are no respondents in that cell, or the nonrespondents in that cell outnumber the available donor records. In contrast, regression models use independent variables to predict a mean and may also add a residual which is either a normal random variate or an empirical residual.

Consider the case described above where three categorical variables X_1, X_2 and X_3 are used to define a match within a cell, and y_{ijk} is the earnings for respondent r in that cell with $X_1 = i$, $X_2 = j$ and $X_3 = k$. However, the selected residual is no longer required to be in the same cell as the nonrespondent; in other words, it is a randomly selected residual. To emphasise that this residual need not necessarily come from the same cell as the nonrespondent, we will denote it as $r_{i'j'k'm}$.

The imputed value y_{ijkm} can be decomposed as

$$y_{ijkm} = \tilde{y}_{ijk} + r_{ijkm},$$

where

$$\tilde{y}_{ijk} = \tilde{\mu} + \tilde{\alpha}_{1i} + \tilde{\alpha}_{2j} + \tilde{\alpha}_{3k} .$$

where $\tilde{\mu}$, $\tilde{\alpha}_{1i}$, $\tilde{\alpha}_{2j}$ and $\tilde{\alpha}_{3k}$ are estimates of the parameters within that cell. However, since $r_{i'j'k'm}$ is a randomly selected residual which need not be in the same cell as the respondent,

$$r_{i'j'k'm} = y_{i'j'k'm} - \tilde{\mu} - \tilde{\alpha}_{1i'} - \tilde{\alpha}_{2j'} - \tilde{\alpha}_{3k'} .$$

This model adopts advantageous features from the hot deck methods and regression models which do not include interaction terms, and is attributed to Fritz Scheuren in Schieber (1978). Unlike the hot deck method, this method will work when there no respondents in that cell. Moreover, assuming that the respondents and nonrespondents have the same distribution, the variance added by the imputations will be smaller when the donor records are outnumbered by the nonresponses. Should there only be one donor record in a cell, the hot deck method would use the respondent value repeatedly, yielding estimates with inflated standard errors. In contrast, the model-based method would select residuals from the entire respondent file.

This compromise between hot deck imputation and a regression model method of imputation allows the simultaneous inclusion of a large number of covariates in the model, with greatly reduced restrictions on the level of detail compared with those imposed on the hot deck. The predictions of the respondent means are potentially more accurate, since the variance added by the imputations should be smaller, and the assumption of a common distribution of respondents and nonrespondents is weakened.

Total Imputation

Perhaps the biggest problem with using imputation techniques to adjust the census counts is that the census is the unique source of cross-tabulations involving person, family, and household-related variables. Hence, if large-scale imputation is to be affected, the characteristics of the "additional" persons

must also be imputed and, to date, no such imputation techniques are available.

However, should the decision be made to impute for total nonresponse, several options are available, including:

1. item-by-item imputation
2. single overall weight adjustment
3. usage of weighting classes to reduce nonresponse bias
4. synthetic estimation and regression estimation

Item-by-Item Imputation

Each field in every data record could be imputed, using the same techniques as described in the above section on data item imputation. For the 1960 Census, the U.S. Bureau of the Census imputed values for entire questionnaires by substituting the questionnaire responses of the previously-listed responding household for every nonresponding household. Such a procedure can yield rather large variances of the estimated values, particularly if some of the records selected for duplication contain extreme (singularly large) values. An article by Hansen, Hurwitz and Madow (1953) gave a maximum increase in variance of about 12 per cent.

Single Overall Weight Adjustment

The procedure of using a single overall weight adjustment is based on the assumption that the distributions of data from respondents and nonrespondents are the same. The frequency of each possible value for a particular variable for respondents is calculated. The sum of the number of respondents plus the estimated number of nonrespondents is then multiplied by the frequencies for respondents to yield estimates of the frequencies of each possible value for each particular variable. Suppose, for a particular variable, n respondents supply data.

The estimated mean would be the simple mean of the respondents, that is,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_j$$

where x_j represents the value of the variable for the i^{th} respondent. Thus the expected value of $\bar{x}$,

$$E[\bar{x}] = \bar{x}_r$$

where $\bar{x}_r$ represents the mean of all the respondents in the population.

Letting Q represent the nonresponse rate (and thus the response rate

P= 1-Q), the bias of $\bar{x}$ is

$$\begin{aligned} \text{bias } \bar{x} &= E[\bar{x}_r] - \bar{x} \\ &= \bar{x}_r - \bar{x} \\ &= \bar{x}_r - (P\,\bar{x}_r + Q\,\bar{x}_s) \\ &= \bar{x}_r\,(1 - P) - Q\,\bar{x}_s \\ &= Q\,(\bar{x}_r - \bar{x}_s). \end{aligned}$$

In other words, the bias of the estimated mean depends not only on the nonresponse rate, but also on the difference between the mean for respondents and the mean for nonrespondents. Since the characteristics (such as the mean and standard deviation) of nonrespondents often differ from those of the respondents, single weight adjustment estimates can be considerably biased.

Usage of Weighting Classes to Reduce Nonresponse Bias

The nonresponse bias of the above simple adjustment procedure can be reduced by making nonresponse adjustments using weighting classes. The values of one or more survey items are

used to partition the population into classes. For each class, the same procedure as for a single weight adjustment is followed, but since the proportion of respondents and the variable means would vary from class to class, the nonresponse adjustments would vary accordingly.

Let us assume that the population is partitioned into c classes, based on the values of one or more items. Let W_1, W_2,...,W_c represent the proportions of the population members contained in each of these classes, and Q_1, Q_2, ...,Q_c be the nonresponse rates in each class.

$$\bar{x}^* = \sum_{i=1}^{c} W_i \bar{x}_{ir}^*$$

where $\bar{x}^*$ is the weighted estimate of the population mean and $\bar{x}_{ir}$ represents the sample mean among respondents in the i^{th} weighting class.

The bias of $\bar{x}^*$,

$$\text{bias}(\bar{x}^*) = E[\bar{x}^*] - \bar{x}^*$$
$$= \sum_{i=1}^{c} W_i Q_i \ (\bar{x}_{ir} - \bar{x}_{is})$$

where $\bar{x}_{is}$ represents the mean of all those in the population contained in the i^{th} weighting class who would not respond.

If the values $\bar{x}_{ir} - \bar{x}_{is}$ tend to be less than $\bar{x}_r - \bar{x}_s$, the differences between the means of the respondents and the nonrespondents, and the values of the response rates, Q_i, vary from class to class, the nonresponse bias will be reduced by appropriate choices of weighting classes. However, this method requires the identification of the characteristics which will define weighting classes which vary both with respect to response rates and anticipated rates (had full coverage been obtained). Moreover, the characteristics used to define the weighting classes must be available for both the respondents and the nonrespondents. In many cases, this last requirement severely

limits the choices of variables to use to define the weighting classes.

Synthetic Estimation and Regression Estimation

Ericksen and Kadane (1985) proposed usage of the US CPS (or a "megalist" obtained by supplementing the CPS sample with matched samples from such other lists as US drivers-licence records, school records and welfare records), plus housing unit samples from the census, to evaluate the overcoverage and undercoverage. Demographic analysis estimates and any other available information would be used to derive estimates of the correlations between misses in the US census and in the CPS (or the megalist) to obtain estimates of coverage. The coverage estimates would be made separately for each of about 50-100 areas made up of combinations of the supervisors' districts used in the census operation. Multiple regression would then be performed, using known independent variables to derive estimated ratios of population corrected for underenumeration and overenumeration (on the basis of matching studies) to counted population for states, cities, counties and smaller areas. Separate regressions could be performed for differing types of areas, such as central cities, suburban and nonmetropolitan and for different racial groups (in this case, blacks, Hispanics and other whites) by age and sex classes.

Commenting on Ericksen and Kadane's article, Tukey (1985) stated that a PEP study for a new census should probably consist of both dilute samples within each of a complete set of segments and more concentrated samples within each of a sample of local areas which might, but need not, involve all segments. According to Tukey, such a study would both lead to improved values for segments and provide a basis for attempting the use of synthetic, or more general regression, techniques in extending segment areas to local area results. Tukey further stated that another, quite distinct virtue of regression among segments is the reduction of sampling error

variation when the regression values, segment by segment, can be taken as better segment adjustments than the raw segment values.

However, estimates of the national net undercount and small area undercounts may be considerably biased. The various population lists used would have been constructed over differing time spans. According to Fay (1985), approximately 20% of the US population moves in a single year, and this high mobility will create problems in matching the census counts with alternative population lists. Any matching method used will also be subject to errors generated by mis-spelling of names, addresses recorded at different dates and so forth. When using a megalist, every person included in more than one list and not detected through matching would be duplicated, possibly even more than once, on the combined list, thus causing a significant positive bias to the resulting undercount estimate. Moreover, according to Fellegi (1985), there is no known method of generating undercount estimates even for a sample of local areas, since sample data at local area level are subject to nonrandom overcount or undercount biases that can only average out over a relatively large area.

Chapter 9

DISSEMINATION OF RESULTS

Introduction

The public response to each census will be determined by the questionnaire content of the current census, the way the previous census was conducted, the speed of publication, the confidentiality of data, the purchase price of specific data, the presentation of the data, the ease of access to the data, and the utility of the data.

Many people are unsure how to interpret statistics, and their lack of confidence is undoubtedly fed by the ability of opposing political groups to draw differing conclusions from the same basic set of data. In addition to tabulated data, the publication of nonpolitical, explanatory reports by statisticians, which outline trends and other points of interest and use pictorial displays such as bar charts, histograms and line graphs to illustrate the main features of the data will increase public interest in census data.

Availability of Census Data

For full usage to be made of census data, the data should be made available in several different media, including published tables, computer printouts, computer tapes or discs, and microfiche. In addition, the range and availability of the data must be well publicised.

It is also desirable to have sufficient flexibility to produce the required information in differing formats, or by differing classifications, or even to provide the capability for users to analyse the data themselves. With the advent of computer

technology, it has become possible to store the data files on computers, and for the user to specify the tabulations or classifications he or she is interested in. However, precautions must always be taken to ensure confidentiality of the data, whether it be by the usage of random rounding, the careful selection of sampled data, or refusal to supply small-area data (which could be used to discern data pertaining to specific individuals, thus violating the guarantee of confidentiality).

Pictorial Representation of Data

Persons unfamiliar with statistics will gain little from reading a table of data. Frequently, specific aspects of the data can be effectively displayed using diagrams and graphs. Piecharts are useful for comparing percentages in each of a few categories of some population characteristic for one or more censuses. Bar graphs can be used to compare percentages in each of several categories of some population characteristic for a single census, and line graphs are useful for displaying trends over a long period of time. Figures 9.1-9.3 on pages 210-212 show examples of diagrams and graphs used by the New Zealand Department of Statistics to illustrate selected aspects of census data.

Thematic maps are a relatively new development in the presentation of census data. They usually display the geographical distribution of one characteristic of the population. For instance, the drift from rural areas can be illustrated by using colours or symbols to represent the rate of population loss for each county. Figures 9.4-9.8 on pages 213-217 show examples of single-subject and multi-subject thematic maps.

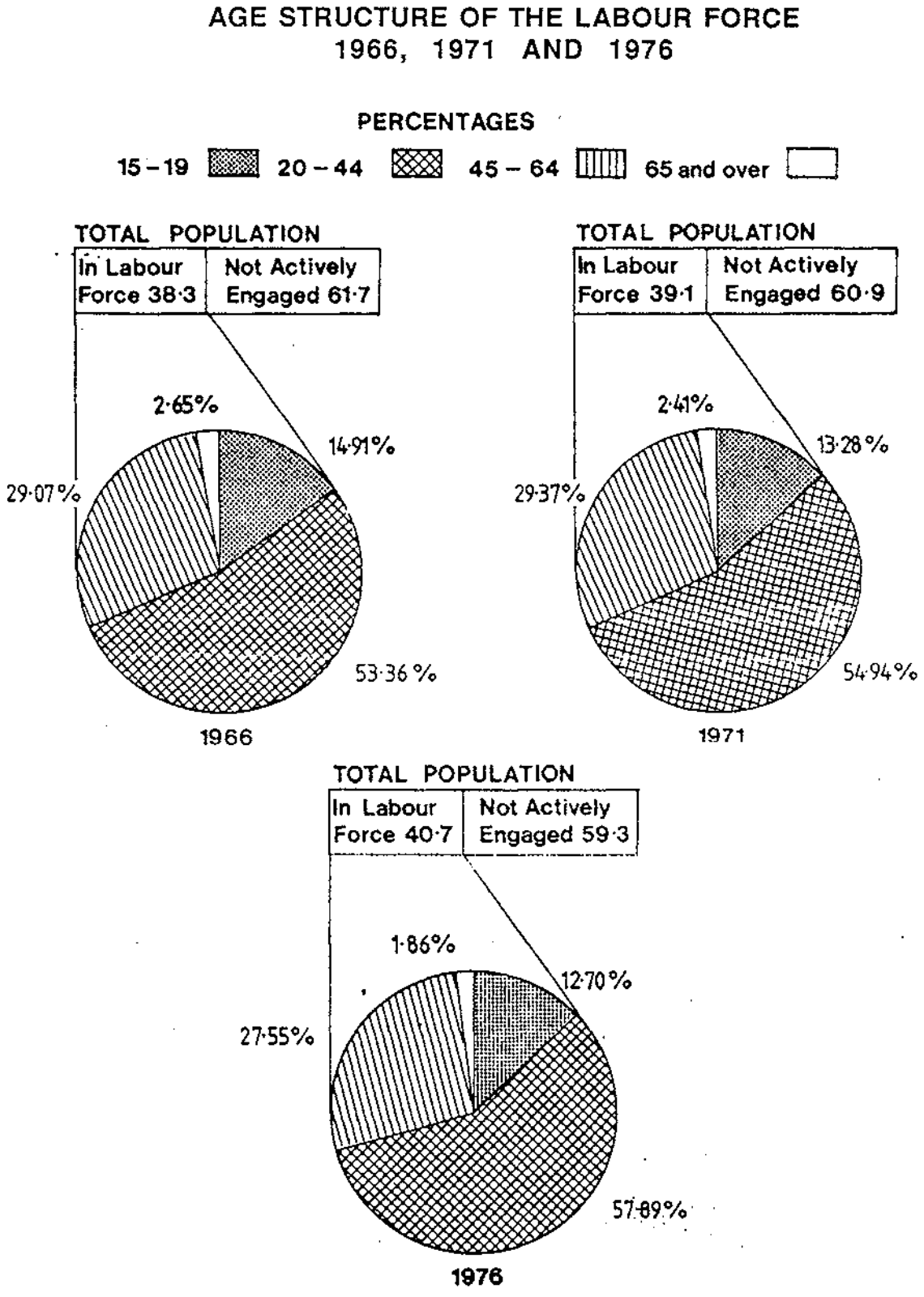


Figure 9.1 Example of Pie Chart

Source: *New Zealand Census of Population and Dwellings 1976 Labour Force*, 1980 (Department of Statistics, Wellington)

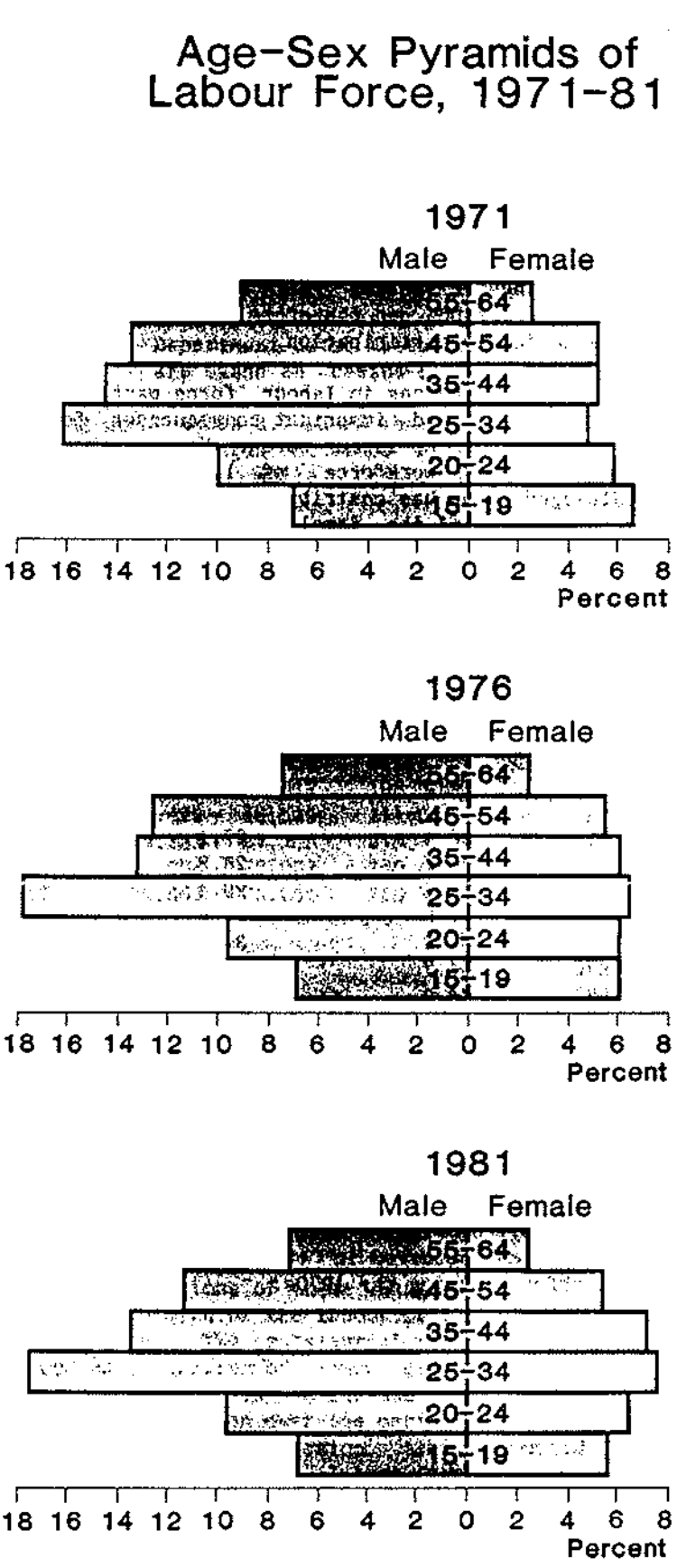


Figure 9.2 Example of Bar Chart

Source: *New Zealand Census of Population and Dwellings 1981 Population Perspective '81*, 1985, *Volume 12 (General Report)* (Department of Statistics, Wellington)

Separated or Divorced Females by Ethnic Group, 1981

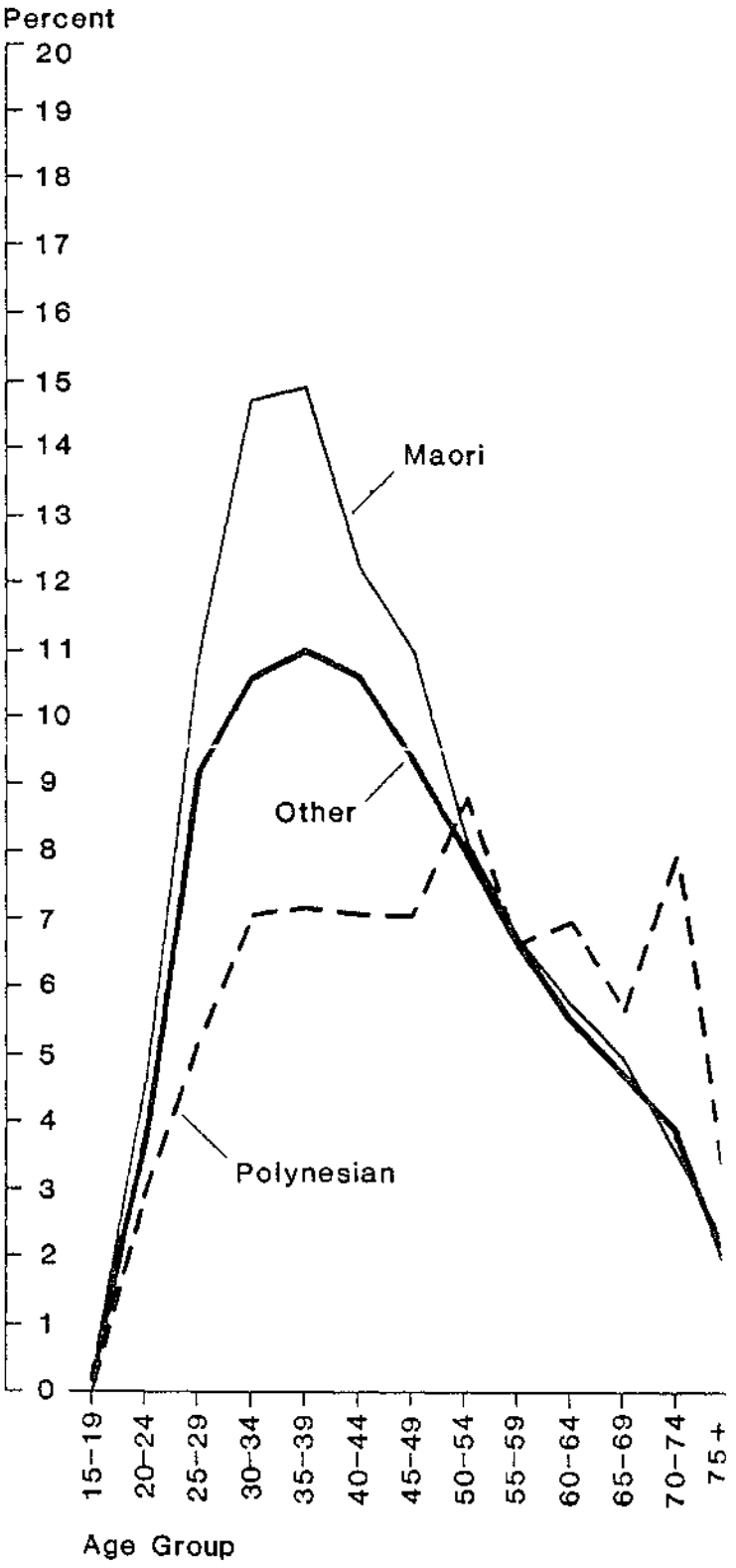


Figure 9.3 Example of Line Graph

Source: *New Zealand Census of Population and Dwellings 1981 Population Perspective '81*, 1985, *Volume 12 (General Report)* (Department of Statistics, Wellington)

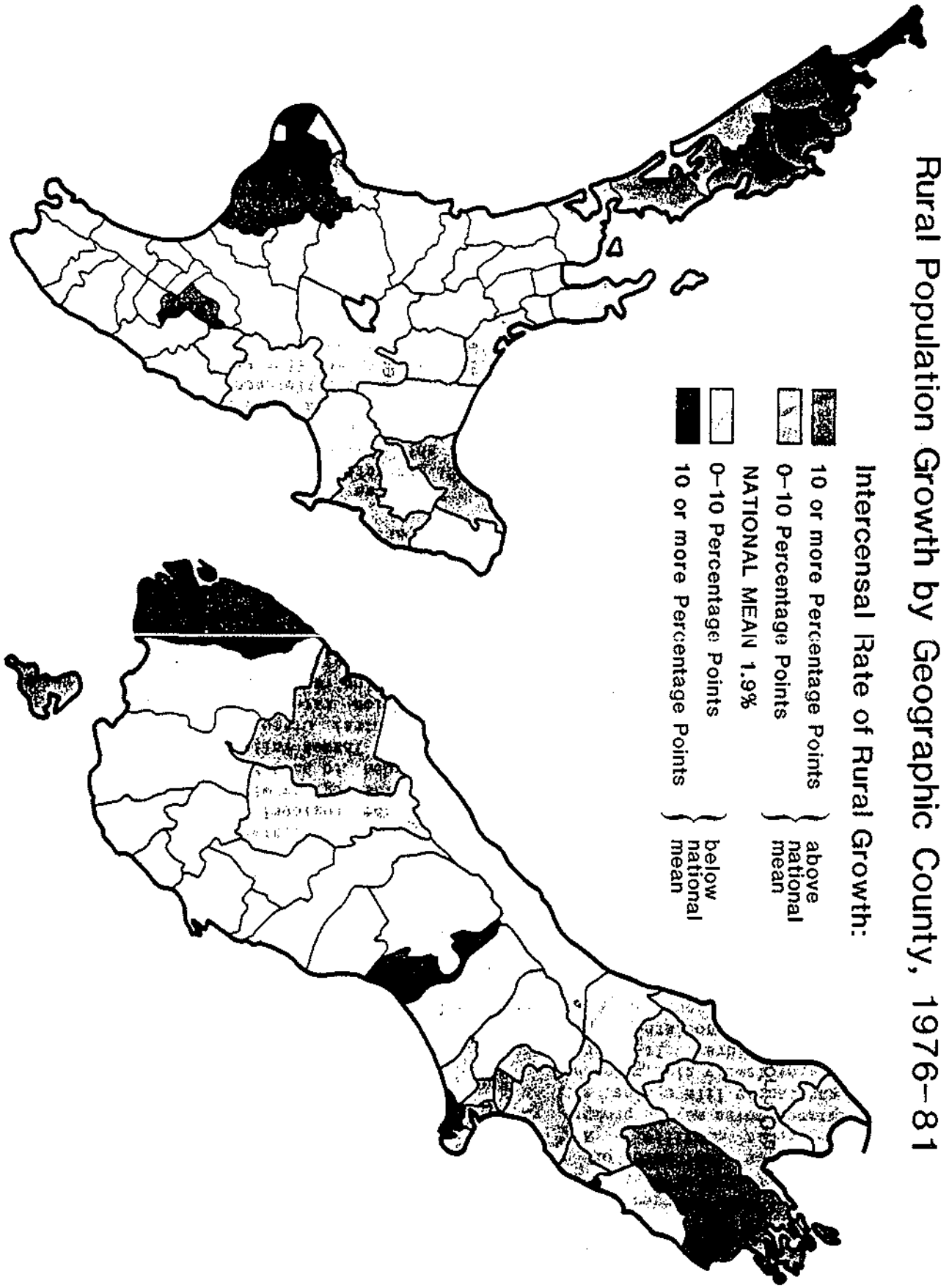


Figure 9.4 Thematic Map; Usage of Shading or Colouring to Represent Data Ranges

Source: *New Zealand Census of Population and Dwellings, 1981, 1985, Vol. 12* (Department of Statistics, Wellington)

The Growth of Adelaide's Built-up Area, 1880-1978

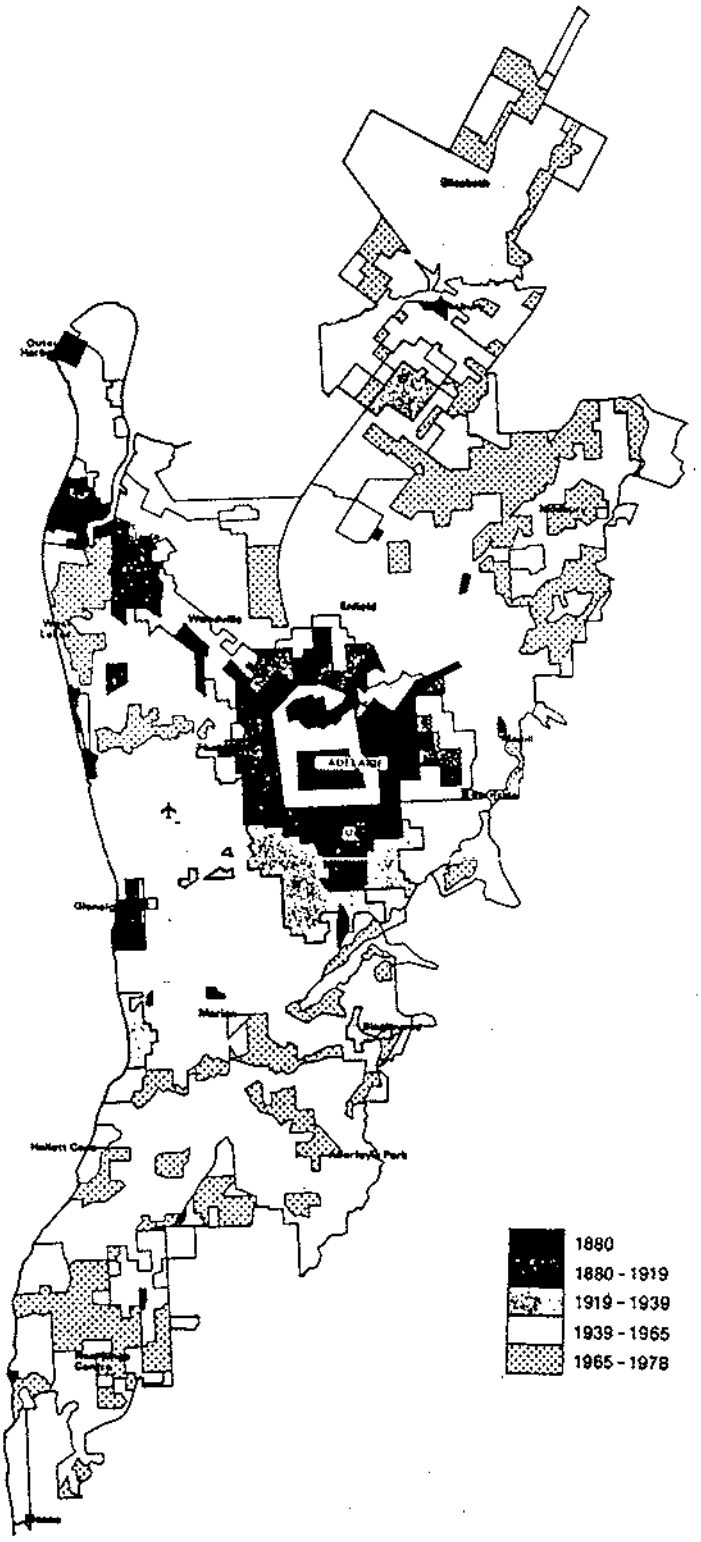


Figure 9.5 Thematic Map; Usage of Shading or Colouring to Represent Time Series Data

Source: Division of National Mapping and Australian Bureau of Statistics *"Adelaide, a Social Atlas"*

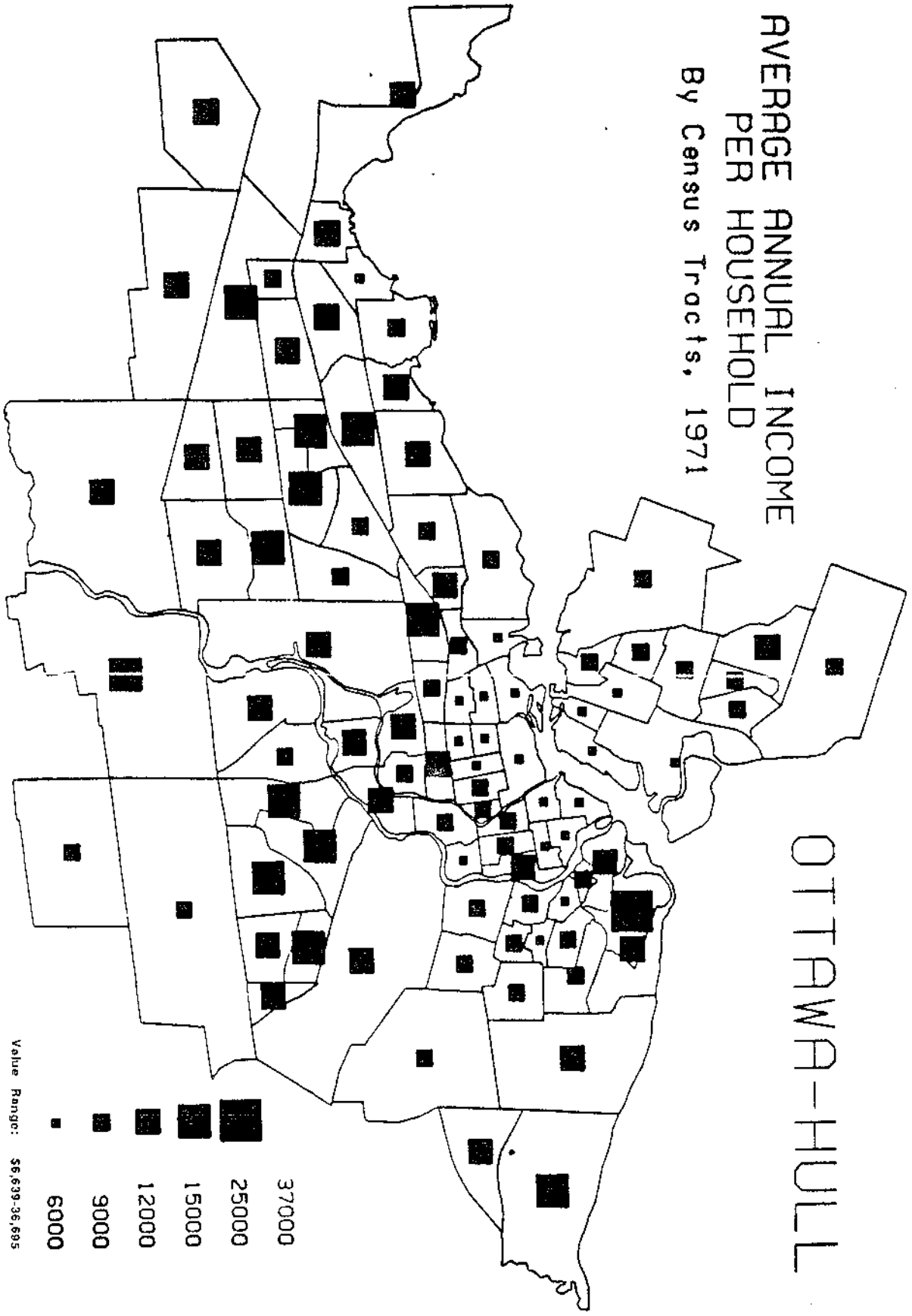


Figure 9.6 Thematic Map, Single Subject, Proportional Representation

Source: Weiss, Caroline C. "An Evaluation of the Gimms Computer Mapping Program" *Census Field Study Cartography Series*, 1977, *No1* (Statistics Canada)

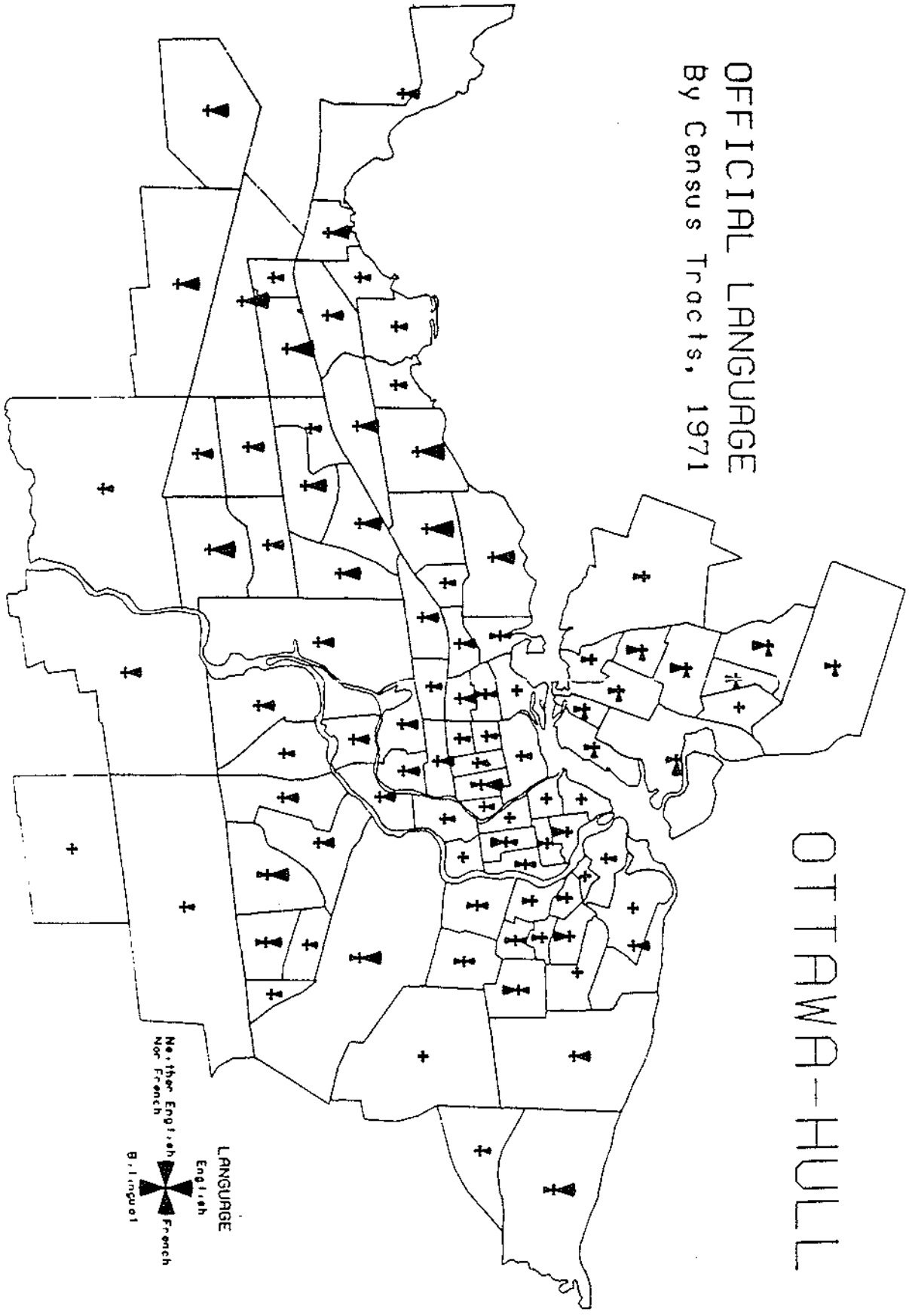


Figure 9.7 Thematic Map, Multi-subject

Source: Weiss, Caroline C. "An Evaluation of the Gimms Computer Mapping Program" *Census Field Study Cartography Series*, 1977, *No1* (Statistics Canada)

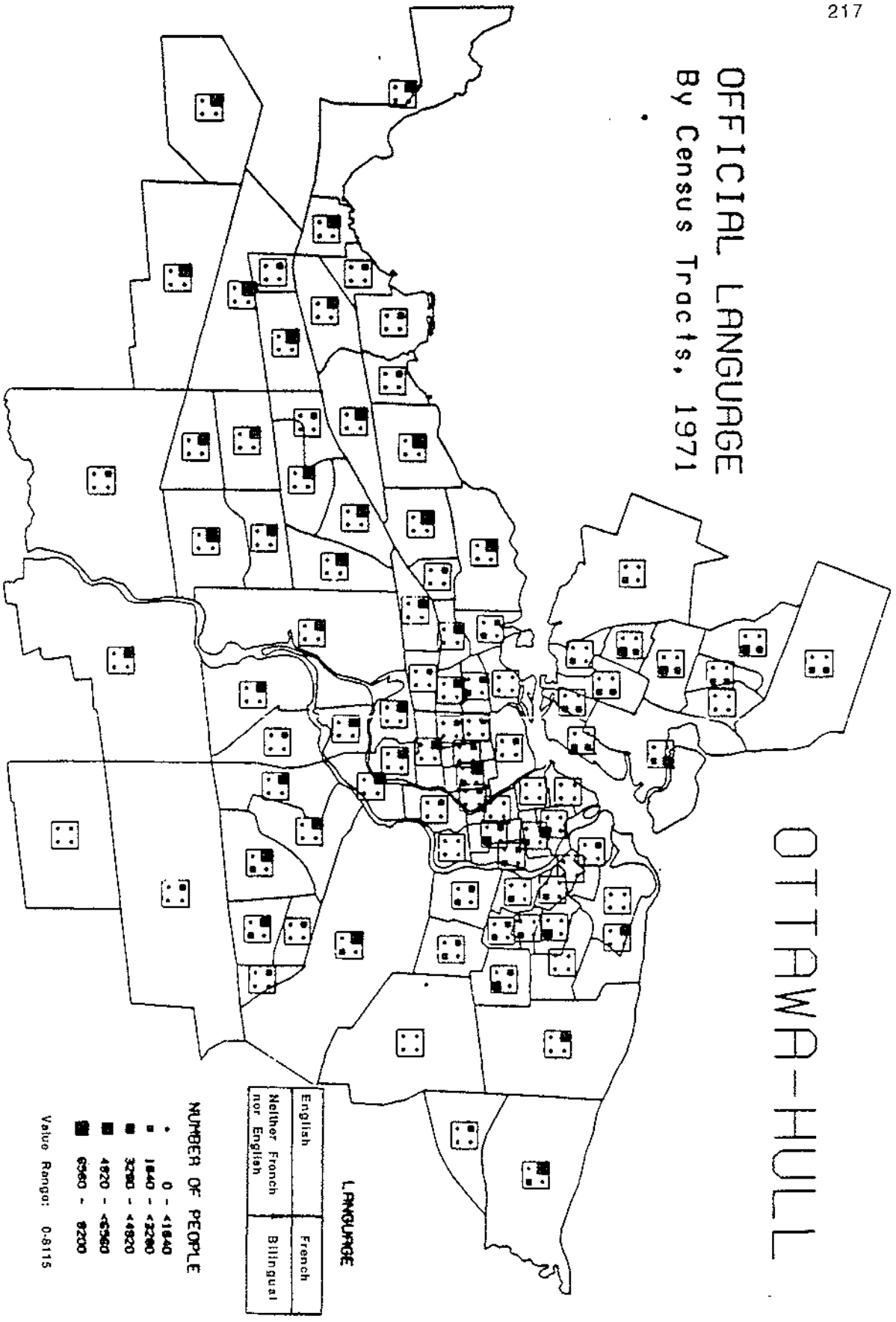


Figure 9.8 Thematic Map, Multi-subject; Alternative Representation

Source: Weiss, Caroline C. "An Evaluation of the Gimms Computer Mapping Program" *Census Field Study Cartography Series*, 1977, *No1* (Statistics Canada)

Confidentiality

It is of interest to note the change in attitudes towards confidentiality of census data. In the first six American censuses (1790 to 1830), Federal marshals were instructed to post copies of the completed enumeration sheets containing individual census information. To enable changes to be made for missing households or missing people, the sheets were to posted in two of the most public places within each jurisdiction, "*there to remain for the inspection of all concerned...*" As mentioned below, the confidentiality of American census data is now scrupulously observed.

Enumeration in India is still conducted on a communal basis, particularly in the rural areas. Enumerators interview respondents in the presence of other villagers, who are permitted to assist, contest or amend the responses of the individual being interviewed. In the light of such an open interview situation, there seems little cause for concern over confidentiality of published census data.

To prevent users of census data being able to trace published information back to individual respondents, it is standard practice in most countries to either withhold data from enumeration districts with only a few respondents, or to use some sort of method to mask the data. Because summary files contain only aggregated data, no masking procedure is necessary before releasing the data. To reduce the user cost of obtaining census data, sample files are produced from the summary files. Such files do not require masking, since the data in the files is randomly selected from the summary files.

The Australian Bureau of Statistics implemented the following measures to maintain confidentiality of the 1981 Census data:

(i) Any collection district with one to three males, one to three females or one to three occupied private dwellings was declared confidential. Each of these collection

districts was combined with another district prior to release of the data.

(ii) Cross-classified tables of population and dwelling characteristics were not released for collection districts which had less than 100 persons.

(iii) Detailed tabulations of population and dwelling characteristics were only released for local government areas and other standard statistical areas with a population of at least 5,000. Similar information for smaller areas was released in a condensed format.

(iv) Any small cells of three or less in the above detailed tabulations were randomly adjusted before publication. All cells in detailed cross-classification tables were randomly adjusted prior to release. The reader is referred to the information paper *Census 81 - Effects of Introduced Random Error,* Catalogue No. 2156.0 for further information about the random adjustment procedure.

The US Bureau of the Census produces summary tape files with varying degrees of small-area data for computer usage. The printed reports, produced in paperback volumes or on microfiche, contain summary statistics at varying geographic levels, down to places of 1,000 or more people. Data for individual block statistics are produced from samples, accompanied by only the barest information on the sample selection procedure.

To maintain confidentiality of New Zealand census data, as required by the Statistics Act 1975, users are not permitted to obtain individual census records. In order to extend the range of information available to census data users while complying with the Act, the Department of Statistics processed the Summary Files and all tabular information for the 1981 and 1986 Censuses using **random rounding** prior to release. All cell values, including cell and column totals, were rounded

using simple random rounding to base three. Using this method, zero counts and counts which are multiples of three are left unchanged, while any other count is rounded to one of the two nearest multiples of three. The probabilities of rounding up or down are set so that the long run expected value equals the original count. For example, suppose an original count of 46 was obtained. This could be rounded to either 45 or 48, since these are the two closest numbers which are multiples of three. If 48 is more likely to selected than 45, then the expected value is

$$45 \times \frac{1}{3} + 48 \times \frac{2}{3} = 47.$$

Conversely, If 45 is more likely to selected, then the expected value is

$$45 \times \frac{2}{3} + 48 \times \frac{1}{3} = 46.$$

Since the latter expected value tallies with the original count, the probability of 46 being rounded to 45 would be twice that of it being rounded to 48.

Similarly, the probability of the number 50 being rounded down to 48 would be $\frac{2}{3}$, whereas the probability of it being rounded up to 51 would be $\frac{1}{3}$.

The reason for applying random rounding is that users of such census data can never be sure what the original numbers were. If, as in the above example, a cell count in a published table was 48, there is no way of determining whether the original count was 46, 47, or 49.

This random rounding procedure enabled an increase in the range of small area census data made available in a short time frame to meet user demand, without violating the anonymity within small groups of figures or requiring additional staff resources. However, usage of this random rounding procedure means that a total will not necessarily be the exact sum of the component

parts, and tables generated at different times from the same data source will rarely contain exactly the same figures.

Timeliness

It is also important that the census data be made available as soon as possible. This is no mean feat, as census data can involve many detailed cross-classifications. It is also essential that the data has been thoroughly edited before being analysed and released for publication. However, several statistics, such as personal income, dwelling and occupancy rates, and family and household size, which are derived from census data, may become rapidly outdated, and it is therefore highly desirable that the publication of census data is achieved within as small a time frame as practicable. In order to achieve timely statistics, it has become common practice to employ questionnaires which allow data to be captured by Optical Character Recognition (OCR) or to use answer boxes which enable rapid data capture and Computer Assisted Coding (CAC).

Cost

The cost of producing the census data in tabular formats, reports, diagrams and maps must always be borne in mind. It is possible to use the computers to select a sample of the census data which may prove more than adequate for the user's needs, and will also be considerably cheaper than obtaining the entire census file.

Coverage of New Topics versus Continuity of Data

Another important requirement is that the criteria for selecting topics to be included in the census must be flexible. Obviously, persons researching trends in populations will want

the continuity of historical census data to be maintained, but it is also necessary for research into new fields to be made possible by the introduction of new classifications of census data. As the census questionnaires must be kept to a reasonable size to prevent antagonising the respondents, it is often necessary to introduce new questions at the expense of some of the existing questions. High quality data is of paramount importance, and this can only be achieved if the respondents can see the value of the questions asked, and if the questionnaires are not tediously long. Regrettably, questions which are of value to researchers who are investigating long-term trends may have to be sacrificed in order to include questions which cover new fields of enquiry.

Changes in Official Definitions and Categories

In addition to the deletion of questions, researchers also have to cope with the problem of changes in definitions between historical and current data as well as changing categories. For example, as noted in Appendix 4.1, the definition of New Zealand Maori was altered in 1974 and permitted persons of any degree of Maori descent to be classified as Maori. Prior to 1974, the only persons who had half or more Maori blood were classified as Maori. This change in the official definition of Maori descent rendered Maori population data for 1976 and ensuing censuses incompatible with earlier census data.

The allocation of a respondent to an ethnic group often proves to be difficult when a person is descended from more than one group. For the 1981 New Zealand Census, a person was coded according to the predominant group in terms of the stated proportion of blood mixture of races. In cases where no group predominated, the allocation was determined according to the following priority order: Maori, Pacific Island Polynesian, other ethnic groups (excluding European), and European. However, for the 1971 Census, persons of part-European part-Pacific Island Polynesian Ancestry were classified as Pacific Island

Polynesian, even if they were of less than half Pacific Island Polynesian descent.

For the 1986 Census, two optional definitions were adopted for each major ethnic origin category. The classifications were based on:

(i) Single Ethnic Origin. This is based on group affiliation which is in turn based on cultural and ancestral criteria; and

(ii) Ethnic Origin/Descent. This classification is closely comparable with previous census statistics on a descent basis. It provides the population with a common biological (or ancestral) background and, with the exception of New Zealand Maori, is considered by the Department to be of limited relevance to present user requirements.

Because data from the early New Zealand censuses used broader age categories than are currently used, data from different periods may need be manipulated before comparisons can be made. The earlier census data can broken down into narrower age groups using life expectancy tables or by modelling the data on a population which is similar to the population being studied, and has a sufficiently detailed age distribution. Alternatively, the recent census age categories can be broadened, to make them compatible with the earlier data.

History of Official Publications of New Zealand Census Data

New Zealand Census statistics have been traditionally published in a series of volumes containing each major subject area, while the final volume summarises the full range of information collected. Because of topic changes, the number and titles of the reports has varied and amalgamations of

reports have frequently occurred, but listed below in usual order of publication are the individual subject matter reports and General Report which are usually included in the historical series:

Location and Increase of Population
Ages and Marital Status
Religious Professions
Industries and Occupations
Incomes
Education
Birthplaces and Ethnic Origin
Maori Population and Dwellings
Dwellings
Households
Families and Fertility
Internal Migration
General Report

It has been the practice of the Department of Statistics to release census data as early as possible, and to this end, it has become standard practice to release provisional census data (that is, data which has not been officially verified) in news bulletins, followed later by publication of bulletins on national and regional data, as well as on specific topics, and finally the publication of the Census Subject-Matter Volumes.

1966 and 1971 New Zealand Census Publications

Prior to 1966, New Zealand census data was analysed and tabulated without the aid of computers. For the 1966 and 1971 Census data, a separate computer programme was written for each table. A statistical table generation package was used for the 1976 Census data, and the Department of Statistics estimated that this saved some 25-30 man-years of programming effort. The package was developed by the U.S. Bureau of the Census, and named COCENTS. Another milestone in the history of census data was reached in 1976, when

preliminary census statistics were published within 9 months of the Census data. With earlier censuses, detailed cross-classifications of census data were not available for up to 5 years after the censuses were taken. In 1976, a systematic 10% random sample was taken from the census data on a meshblock basis, and the 10% sample tables and the estimates derived from the tables were published as preliminary census statistics. This bulletin was preceded by a series of news releases on provisional counts for each territorial local authority area, and a bulletin containing provisional population and dwelling statistics for Statistical Areas, Statistical Divisions, Main Urban Areas and territorial local authority areas. In the following four years, regional bulletins based on locality within Statistical Areas, subject-matter bulletins such as the Ages and Marital Status Bulletin and Census Volumes were published.

1981 New Zealand Census Publications

Preliminary population counts from the 1981 Census were published in a series of news releases from mid-April to mid-July 1981. Provisional population counts for each of the 13 Statistical Areas, the 7 Statistical Divisions, the 23 Main Urban Areas, and the 14 Secondary Urban Areas, as well as a table of occupied dwellings in Statistical Areas were contained in a news release in August. Provisional population and dwelling counts for local authority areas including subdivisions were published in Bulletin No. 1 in September 1981 and a November 1981 news release contained the final New Zealand population total.

A 10% systematic sample of private households plus the associated dwelling and personal questionnaires from all non-private households were given priority in the coding operation, and enabled the November 1981 publication of Bulletin No.2, which contained 60 tables of selected cross-tabulations on national data and the February 1982 publication of Bulletin No.

3, which contained 22 tables of single-topic tabulations on regional data. Bulletin No. 3 contained statistical data for Statistical Divisions, Main and Secondary Urban Areas and a composite Minor Urban Area group made up of all boroughs, town districts, district communities, communities and townships outside the Main and Secondary Urban Areas with populations of 1,000 and over at 1981 Census.

In an effort to improve the timeliness of the 1981 Census Volumes, the Introductions that had been traditionally included were published as separate bulletins. The first publications in the standard volume series were released from May 1982. Volume 1A contained all population counts for all non-administrative regional areas, plus local authority areas including subdivisions. Volume 1B contained statistics of population density in New Zealand, together with final population counts for various small areas designated as townships, localities and vicinities. A series of regional Bulletins containing final census results were progressively released from April-November 1982. They included 23 single-topic tabulations for Local Authority Areas (including subdivisions) and were published by regions approximating Statistical Areas, a Regional Summary, and National Statistics.

A computer network of official statistical data was established in 1982, called INFOS (Information Network for Official Statistics). Census data is now stored on computer files, and the public is permitted access to these files through computer terminals, as well as the more traditional forms of output dissemination, such as published tables and reports, and computer printouts made available on request.

The INFOS system, being a new facility, initially had only a small range of census data loaded into the system, and did not permit access to historical data. However, as mentioned below in the section on 1986 New Zealand Census data, specifically selected historical New Zealand census data is now stored on a public data base, which can be accessed through INFOS.

In 1984, the Department of Statistics announced that data from the 1976 and 1981 Censuses of Population and Dwellings were available on **magnetic tape**. The tabular information could be obtained from user-written or use-specified programmes which could be run against the 1976 and 1981 Census data files, or from the Table Archive System which stores a range of printfiles produced from the 1981 Census.

The Census data was available in ICL and IBM readable format from ICL or IBM 4341/4361 computers. As mentioned above, the method of random rounding employed to maintain confidentiality of the 1981 Census data generated slightly different figures each time a table was produced, and additivity of the table was not always preserved.

The Table Archiving System was created in October 1983 and was based on the ICL 2980 computer. It was created to provide copies of the original tables that were generated for producing census bulletins and volumes. However, these tabular print files were not duplicates of the publications, as several tables were often required to make one published table, and at other times, only portions of the tables were used to produce one published table.

User-written or user-specified programmes using CENTSAID (ICL), SAS or TPL (IBM) were accepted by the Department and run against the 1976 or 1981 Census 10% sample, down to meshblock aggregation level, or full (100%) data files, as requested. All output from the full data files was random rounded to base three, and was provided on magnetic tape together with a hardcopy printout if required. If information was not required on a geographic basis (in other words, national tabulations were sufficient), a 1981 Census 1% data file which contained all variables other than geographic variables was available for user access. Output from the Census 1% and 10% sample data files was weighted accordingly to reflect national totals.

CENTSAID was developed for non-specialists to extract various sources of ad hoc information. It was easy to use and achieved large cost reductions in the table testing area because of the ease of programming, but had severe limitations in terms of the number of report format variables processed and did not produce visually acceptable census tables without manual intervention.

Other statistics from the 1981 Census which were available on microfiche, computer tape or computer printout are as follows:

(i) *Local Authority and Area Unit Tables, Meshblock Tables and Meshblock Listing* (not available on microfiche, but could be ordered as photocopy, printout or computer tape). A computer listing of Census Night populations, together with counts of inhabited and uninhabited dwellings for each meshblock.

(ii) *Listing of Rural Settlements.* A clerically compiled listing of Census Night populations in rural counties, together with counts of inhabited and uninhabited dwellings for each rural township, vicinity and locality (available as photocopied pages only).

(iii) *Meshblock Number Key.* A listing showing conversions from "standard" meshblock numbers to census administrative numbers and vice-versa.

(iv) *Area Unit Keys.* A listing showing area units within urban areas, statistical divisions and statistical areas. Another listing showing area units within each local authority and their included meshblocks was also available.

The Department also had supplies of maps in the NZMS.92 series showing statistical area, statistical division, urban area and county boundaries together with the boundaries of local authority sub-divisions contained within main and secondary urban areas and "User" maps showing local authority and meshblock boundaries as at the 1981 Census of Population and Dwellings.

1986 New Zealand Census Data

The 1986 Census questionnaires were designed to allow the rapid capture of data by using precoded answer boxes for the bulk of the questions, and Computer Assisted Coding was employed for the remaining questions, with the consequence that the Provisional Local Authorities Population and Dwelling Statistics bulletin was cancelled due to the finalised statistics being ready for publication several months ahead of schedule.

The following media were used for publishing the 1986 Census of Population and Dwellings data:

1. *Published tables and reports* (Census Bulletins and Volumes). These include Local Authority Statistics, Rural Population Statistics, Regional Statistics, Usually Resident Population, Hospital Board Districts and Health Districts, Electorate Profiles, Urban and Rural Profiles, Population Atlas (containing thematic maps), National Statistics, Ages and Marital Status, Labour Force, Birthplaces and Ethnic Origins, Internal Migration, Incomes and Social Welfare Benefits, New Zealand Maori Population and Dwellings, Pacific Island Polynesian Population and Dwellings, Dwellings, Households, Families, Religious Professions, Education and Training, Total Population Statistics, Overseas Visitor Statistics, Profiles of New Zealanders, Questionnaire Content and Submissions, Range and Availability of Statistics, Scope of Census, Census Data Files, Technical report on Census Media.

2. *Microfiche or photocopies of microfiche frames.* These include small area statistics at the meshblock level. Major topics are covered in a summary format.

3. *Printouts of unpublished tables.*

4. *Photocopies of publications or printouts.*

5. *Record layouts of census files* and associated data base dictionary.

6. *Copies of sample Cents-Aid/SAS/TPL computer programs.* These illustrate how to analyse 1981 and 1986 Census data by using a sample of questionnaires, from which the identifying names and addresses have been removed. Tables can be produced using Cents-Aid, and statistical analyses can be performed using SAS. Users must write their own programs, which are then processed by the Department of Statistics.

7. *Cents-Aid Manual* and a publication describing the contents of 1981 and 1986 Census Public Use Sample Files.

8. *Census data on magnetic tape*, produced in both ICL and IBM-readable format. The data is obtained from the Table Archive system, which stores a range of print files produce from the 1981 and 1986 Censuses. User-written or user-specified computer programs can be run against the full (100%) data files or sample (10%) files from the 1976-1986 Censuses. Users can request either magnetic tape or printout.

9. *Data available in IBM 4331 machine-readable form*, including:
   (i) *Data stored on INFOS*. This essentially includes indexed cross-tabulations which have been specifically selected because of their utility in association with data from earlier Censuses of Population and Dwellings or from other surveys;
   (ii) *Geographic reference files* containing 1971, 1976, 1981 and 1986 Census meshblock identifiers and meshblock aggregates such as area units, local authorities, and electoral districts;
   (iii) a *Rural Settlements key* containing descriptions of rural settlements, meshblock identifiers, area unit

codes and counts of population and dwellings in each meshblock and/or part meshblock for the settlement;

(iv) a *Meshblock Control File* containing population and dwelling counts for each meshblock, the requirements of the Representation Commission, control counts and history of the meshblock;

(v) *Indexed cells or summary files* which are computer readable files of indexed cells of cross-tabulations which permit record and cell selection and/or aggregation as specified by the user and a series of predetermined tabulations summarised at meshblock level;

(vi) *Unit record data files (full census and sample)* as record files or SAS data sets for ad hoc statistical requirements that cannot be satisfied from generated tables, INFOS or summary files. These data files are indexed by region, and are available as separate regional files, containing an incomplete set of fields and any additional fields required to be created from fields or values on the masterfile;

(vii) *Digitised boundaries of area units, and Territorial Local Authorities* which can be linked to the geographic reference file and the census data files. The file of digitised boundaries at the unit area level will enable mapping of those areas which comprise aggregates of whole area units;

(viii) *Meshblock Centroids*, which are required to calculate the production of statistical aggregates for ad hoc areas defined by users.

10. *Documentation relating to the census series on INFOS.* A Census Information Aid (CIA) for the 1986 Census is stored within INFOS, and contains an indexing and documentation system. It was designed as an aid to census users who wish to know more on the range and availability of statistical outputs from the Census, and allows users to specify the variables they require.

Features of the CIA include details of expected date of availability of data, the costs of obtaining the requirements in the desired media, and the opportunity to view the formats of tabulations.

Australian Census Data

In addition to the traditional publications of Australian census data, users can now obtain 1981 and 1986 Census data on an IBM-compatible compact disk. Supermap, the accompanying computer programs on the disc, will display the data pictorially in table, graph or map format. While Supermap cannot be used to alter the census data, sections of the information can be reproduced and stored on floppy disks. Personal programs can then be adapted to manipulate information between Supermap and standard spreadsheets, word processing or data processing packages. Figure 9.9 on page 233 shows a sample of compact disc-generated information displayed using a thematic map.

Future Developments in Data Dissemination

The availability of census data in public data base files, and transferral of selected data to compatible computer systems has resulted in the data being available in a far shorter time frame, and a drastic reduction in the rate of transcription errors. The increased availability of programmes which can present the data in the format and by the classifications specified by the user means that census data users are no longer limited to tables produced in standard formats, or having to wait for the release of ad hoc tables which had to be officially requested by the user.

New programming techniques should make it possible for users who are not computer experts to obtain full descriptions of the methods used in existing systems, and to explore the consequences of changing those methods. For example,

Distribution of Household Incomes above $40,000
in Central City Areas of Sydney and Melbourne

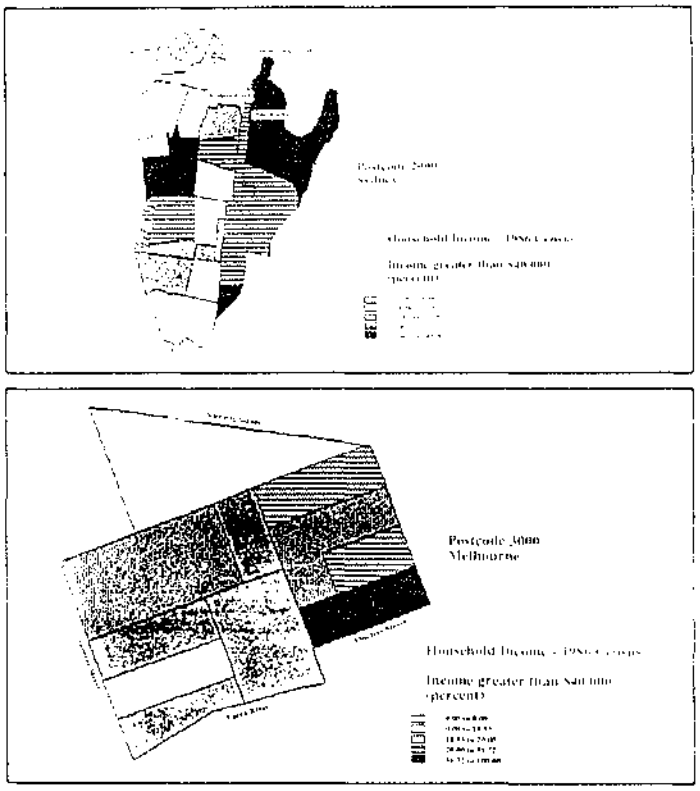


Figure 9.9 Example of CD-Generated Thematic Map

Source: CSIRO "Census Data on CD" *ICON IICT,* December 1988, *No.3* (Institute of Information and Commercial Technology, Australia)

provided that confidentiality of individual responses is retained, users should be able to specify which variables are to be included in each table produced, and the level of classification required. Users will be able to produce tables using differing classifications, obtaining immediate feedback on the effect of varying the level of classification. For example, the age-grouping can be changed according to the level of detail required, or one particular age group could be selected, and the effect of classifying the data by varying combinations of sex, geographic region, occupation and race observed. Because such systems will relieve statisticians of the burden of standard table production, they will have more time to devote to interpreting the census data, and producing more extensive general reports, or acting as consultants to clients interested in specific areas.

Users will be able to make their own analyses, employing specific assumptions. The effect of these assumptions on the results may result in further anaiyses being initiated, and comparison of the results. For example, predictions of future population counts could be obtained by specifying various birth, death and migration rates. The composition of the workforce for a specified future period could be predicted, as could the effect of a decreasing family size on the workforce, and the effect of a "baby boom".

Users will be able to produce their own multi-colour graphs of census data, displaying one or two variables, depending on the amount of information involved. Thematic maps of the entire country or of specified local areas will be produced on demand, according to the statistical aggregates defined by the users. Allocation of centroids to meshblocks will permit census users to define their own geographical areas for the maps, rather than having to use the standard statistical regions.

Chapter 10

THE FUTURE OF THE CENSUS

Introduction

The demand for subnational statistics has grown markedly over recent years. This upsurge in demand is partly due to a desire to obtain economic statistics that will enable the identification of sections of the population which are in need of assistance. Administrative data which is available for small-area units is an obvious potential alternative to census data. For instance, in New Zealand: the Department of Inland Revenue collects information on incomes, the Department of Labour collects data on the workforce, and the Department of Social Welfare has data on welfare benefits. While there is duplication of collected data, existing law do not permit the provision of information between government departments or to any other agency or individuals. A current proposal that Department of Inland Revenue and the Department of Social Welfare cross-check their records to detect benefit fraud has met with a markedly adverse reaction from various sections of the community; indicating that many members of the public are not yet ready to accept a pooling of data by several agencies.

The collection of similar data by several agencies is a constant source of irritation to many respondents, and it would appear that much of the redundancy of data collected could be eliminated by improved coordination of the data sources and common usage of the data. This coordination would not only reduce the frequency of the collection of data, but would also markedly reduce the cost of producing the required statistics. The idea of supplementing survey and census data with data already available in statistical records is not new, but it does have several inherent problems, many of which are discussed in this chapter.

Minimising Costs

A major problem for those responsible for the production of census data is how to achieve the best possible results for the least cost. The high cost of censuses are incurred not only because of the large number of staff involved with the actual enumeration, but also because of the massive public relations exercise which must be launched prior to the enumeration in an effort to elicit a good response from the public.

Traditional censuses are becoming more and more expensive, and cannot be undertaken with sufficient frequency to satisfy the rapidly expanding data needs. The effects of budget cuts on sample surveys and censuses are readily identified by decreased sample sizes in surveys and pilot tests, longer delays in publication of the census or survey results, decreased frequency of surveys or censuses, or the more extreme cases of cancelled pilot tests, surveys or censuses.

Confidentiality

At this point, it should be mentioned that hand-in-hand with the idea of the increased usage of data from administrative records and from census and survey data comes the problem of confidentiality. On collection of the data, respondents are assured of complete confidentiality, and this trust must not be violated. The advantages of inter-usage of data between statistical records and censuses and surveys must be tempered by a built-in system which prevents the identification of characteristics with respondents, thus ensuring the protection of privacy, and also preventing improper usage of the data, such as for financial gain by commercial firms. Australian and New Zealand census data are rounded to multiples of three. Another procedure being investigated by the U.S. Inland Revenue Service which may prevent violation of privacy is known as "blurring" (Jabine, 1984), and can be used when a small number of continuous variables are involved in a particular data field.

The procedure involves sorting the data records into groups of a specified size, and then substituting group averages for individual values. Since records are not simultaneously grouped on all variables, less information is lost than when traditional grouping procedures are used, while still ensuring confidentiality.

Differing Objectives of Data Collection

Administrative records are collected and maintained for specific administrative purposes. These purposes usually differ from statistical purposes. The main aim of administrative systems is to record the relevant details about individuals which are required for particular administrative purposes, and while statistical summaries are often required, the production of information is not a major objective. In contrast, the primary objective of censuses and surveys is to provide detailed information about various subgroups of the population, while stringently ensuring the privacy and confidentiality of individuals.

In order to produce statistics for all areas of a nation from administration records, the records must be detailed, comprehensive and consistent. A researcher interested in obtaining data on social security benefits may well find that the administrative data available is not sufficiently specific. For instance, the data may not be classified into sufficiently small age groups, or the researcher may be interested in relating the social security benefits to both marital status and age groups, or even to the birthplaces of the recipients.

The practice of compiling facts from multiple sources can prove to be a difficult, frustrating and expensive exercise. Because data is collected for specific and often peculiar reasons, the classifications used in published data will vary markedly, resulting in varying table formats and with dissimilar definitions. The definitions used for each publication must be

carefully noted, as these may vary for publications obtained from different sources, and even from everyday usage.

Different administrative records used to generate statistics have frequently produced differing results. Moreover, statistics based on census or sample survey results have conflicted with those obtained from administrative records. This is of particular concern to those involved with census and survey design, as administrative records are often used as sampling frames for statistical surveys and for post-census evaluation. This lack of consistency between data obtained from different sources can be attributed to differences in the procedures employed to define reporting units, to identify and code reporting unit characteristics, and to develop local standards for data tabulations.

Quality Control

Data from administrative records should not be taken at face value, but must be thoroughly checked for accuracy. Like all other administrative procedures, the maintenance of administrative records is subject to financial constraints, and agencies often rely on the public to advise them of changes in address or circumstances, rather than instigating surveys or other procedures to update their records. Administrative data is usually collected on a continuous time frame, whereas census data, as far as practicable, is collected on one particular night so that a "snapshot picture" is obtained of the population. Hence if administrative data is to be used as an alternative source or as a substitute for census data, the different time frames must be taken into account.

Naturally, administrative organisations would be reluctant to cooperate with statistical agencies unless there were obvious benefits for both parties. One possibility would be for statistical agencies to process administrative records and link them with census or survey data to produce aggregate data

which would yield information on the coverage and accuracy of the administrative records. The statistical agencies could provide files of the linked administrative and statistical data to the administrators, provided that steps were taken to ensure that these files could not be compared with the original administrative files, to yield identification of particular individuals.

Conflicting Definitions

One of the major obstacles to increased usage of administrative records is the lack of a uniform set of establishment codes and definitions. For instance, there are at least four possible definitions of "family":

(i) An 'economic family' is defined to be a group of individuals related by blood, marriage or adoption who share a common dwelling unit.

(ii) For the purposes of the New Zealand Census, a family is defined as *"a husband and wife with or without never married children of any age or a lone parent with one or more never married children, living in a private household"* . (For purposes of the New Zealand Census, a **private household** is defined to include all persons who live on the same premises and who usually eat one or two meals together daily, or at least share the same cooking facilities. This definition includes temporary guests, but excludes persons living in group living quarters.)

A 'census family' is not necessarily all related people in a household, but only those related by blood, marriage or adoption, who normally live together as a single family unit, as defined above, and who are present on census night. Thus one or more census families may be contained within one economic family.

(iii) The term 'nuclear family' has recently come into vogue, and this is a restricted form of the definition of a census family, with age limits placed on the children.

(iv) A 'tax family' is defined to be an individual with or without a spouse, but with dependent children. The New Zealand Inland Revenue Department, for the purposes of a family rebate in relation to a taxpayer and to any income year, defines a family to be:

" (a) The taxpayer and another person not being the housekeeper who has the care of a qualifying child.

(b) In any case where only the taxpayer has care of a qualifying child. "

Changing Procedures

A further potential problem is that administrative records are subject to change, and notification of any such changes would need to be well publicised, and given in sufficient time to allow a comfortable adaptation to the changes. Changes in laws, such as placing new data requirements on a system, or eliminating old requirements, will effect major changes in the content of administrative records. A changeover to a new computing system or data-base management system, or new policies covering the disclosure of data will affect the accessibility of administrative records. Changes in report forms, new compliance procedures and new data edits will affect the quality of administrative information. The credibility of census and survey data depends on the timely production of consistently high quality data, and undue delays in publication or erroneous documented data will adversely affect public response to the entire census or survey procedure.

To facilitate the increased usage of statistical records, high quality control would need to be assured, and globally standard definitions and codes would need to be adopted by all data-collecting agencies. Data would need to be made available in machine-readable form so that it could be automatically edited

(and, where necessary, imputed) and then immediately analysed, using standard statistical computer packages or computer programs which had been written specifically for particular analyses.

Current Usage of Administrative Records for Census Purposes

Currently, administrative data are used to complement census and survey data. As mentioned in Chapters 5-8, which discussed Pilot Testing and Quality Control, under certain conditions administrative records can be used to assess proposed questionnaires in terms of:

(i) the accuracy of the responses;
(ii) providing a sampling frame which is independent of any census or survey;
(iii) providing more accurate estimates at subnational levels of certain highly correlated variables; and
(iv) to employ multiframe designs, thus improving the coverage of groups not represented adequately in area sampling frames, enabling oversampling of potentially rare groups of special interest in the population, and reducing the sample variance of estimates yielded from the survey.

The New Zealand Department of Statistics has not yet used administrative data to supplement or replace census data. While the Department would like to improve the quality of its data and increase its statistical services, current budget constraints and the issue of confidentiality will prevent the usage of administrative data in this context in the near future.

In 1980, the U.S. Bureau of the Census used administrative records as a complementary source to statistical data. A sample of persons on individual tax records was matched to census records to evaluate census coverage (by identifying errors and imputing data when necessary), and administrative

records were also used to identify special target populations for improved coverage in the census. The Census Bureau plans to release its annual preliminary intercensal estimates more than a year earlier than prior estimates for some states by implementing a methodology which is based heavily on state and local administrative records. Another proposal is to explore the possibility of substituting administrative data for existing questions on the census form. However, the issue of confidentiality and accuracy of the data will need to be resolved before any attempts are made to access administrative files.

The US Current Population Survey (CPS) uses local building permits as a basis for updating the CPS address sample for new construction between censuses, and the Bureau's intercensal estimates program incorporates local administrative records of births, deaths and migrations to produce estimates of the population of all states, counties and local governments.

In contrast to the New Zealand situation, Statistics Canada enjoys access to the following administrative records:

(i) tax records, family allowance records and unemployment insurance records are used to obtain information on the labour force, demographic data (such as age, sex, marital status and birthdates of children) and data on incomes and migration. Government programmes have been evaluated using individual tax records;

(ii) interprovincial migration of the Canadian population has been estimated using family allowance data;

(iii) data on the labour force, including employment income, gross labour force and unemployment, have been obtained from personal income tax files and unemployment insurance files;

(iv) the potential of conducting marketing studies and developing marketing strategies using individual tax records has been investigated, bearing in mind the requirement of confidentiality (it has been proposed that administrative social data be produced for small areas and that these data be then used to target direct mail advertising agencies);

(v) also under investigation is the feasibility of employing small area data generated from tax files to provide aggregated data for a federal agency which is engaged in management advice and loans to small businesses; a sample frame for a survey designed to assess the incomes and pension plans of Canadians was defined using tax files;

(vi) the return-migration phenomenon, limited time-series data on segments of the population and assessment of the income patterns of migrants and non-migrants have been investigated using data from tax files;

(vii) data on the ages of children obtained from family allowance data have been used by individual school boards to assist in the administration of their school districts;

(viii) an assessment of experimental small area unemployment indicator based on counts of unemployment insurance beneficiaries.

Data Banks

With the advent of public data base systems, it is been mooted that censuses will soon become redundant. Much of the information collected by censuses is already available in administrative files, and this information could be loaded into data bases for access by various users. However, as mentioned

earlier in this chapter, the data in administrative files is collected in different time frames, using varying definitions, and it is often not of sufficient detail to satisfy census requirements. Moreover, most data is collected under the guarantee that the information will only be used by the agency collecting the data, and will not be released to other agencies or individuals.

Increasing usage of plastic credit cards, EFT POS, smart cards and deposit cards provides a pool of information on credit transactions and, more importantly, the movements of individuals which could be tapped for various analyses, but again this threatens violation of personal privacy. Without information on internal migration, sub-national and small-area estimates would not be obtainable from data bases. Moreover, in order to link the various data files for a particular respondent, a unique personal identifier would have to be employed. Such identifiers would have to be issued to every member of the population and some system would have to be introduced to ensure that the identifiers were adopted and correctly used; a very costly operation in terms of time, money and effort. There is much resistance to such a scheme in New Zealand, as it is interpreted by many as a gross infringement of personal liberty.

Conclusion

In summary, provided that the provisions are made to ensure that the confidentiality of responses is not violated, there is undeniably tremendous potential for expanded usage of administrative records at several stages of the development and execution of censuses and surveys, including the definition and stratification of sampling units, development of sample frames, sample selection, evaluation of questionnaires, imputation of missing or erroneous data, estimation and evaluation. However, if administrative records are to be increasingly used to complement statistical data, or even be

substituted for statistical data, then the content, collection, processing, quality assurance and dissemination of administrative data must be tightly controlled.

A further note of caution is that substitution of administrative data for statistical data reduces the redundancy necessary to identify content error and undercoverage, and even more importantly, it removes the "safety net" of a secondary source of data, should part of the data collection system fail unexpectedly.

Censuses must continue because they are the only effective method of obtaining the information necessary to provide for current and future needs of the population; censuses enjoy a higher response rate than surveys, and administrative data are seldom collected at a common reference point in time, use non-standard definitions, and are of varying quality and level of detail. Strict confidentiality laws currently prevent access to many alternative data sources, such as administrative files and data bases.

The following challenges must be addressed for censuses to remain credible and acceptable:

(i) Proliferation of surveys. With the growth of commercial market research firms, the public is being exposed to an increasing number of surveys. Over-exposure to such surveys will result in an increasing level of nonresponse. Collectors of census data face an ongoing battle to keep the nonresponse rate as low as possible. Methods currently employed include publicity campaigns, liaison with leaders of minority groups, translation of questionnaires and leaflets into other languages, establishing information centres and educational campaigns.

(ii) Confidentiality of responses must be assured.

(iii) Expenses must be kept as low as possible, while still retaining effectiveness. Thorough pilot testing must be conducted to ensure that the questions asked obtain the desired data; effective follow-up procedures must be maintained to minimise nonresponse; exhaustive quality control procedures must be maintained; post-censal evaluations must be continued to ensure that the experience of previous censuses is fully exploited to refine and improve census procedures.

(iv) Data must be released for publication in a time frame which is acceptable for users, while still ensuring the the data is of the highest possible quality. To reduce the time required for data processing and analysis, the latest computer technology must be utilised.

(v) Data must be made available in several different media, and accompanied by explanatory reports, diagrams and graphs which highlight the main features of the data.

(vi) Adoption of universal definitions. Standard definitions within a country would ensure that, within the restrictions of confidentiality, the fullest possible usage could be made of data from other sources, reducing the amount of duplicated data collection. More extensive information could then be collected, as the redundant questions could be replaced with questions on new topics. Standard definitions between countries would facilitate comparison of social conditions in various countries.

(vii) Adoption of definitions which are acceptable to respondents. Definitions such as Ethnic Origin classifications must conform to current social attitudes. Issues such as who is Head of Household must be confronted, particularly in joint-income families. Other methods of relating all persons in a household unit need to be examined. Again, these demands conflict with the

requirement of data which is consistent with previous census data for analysis of trends in society.

(viii) Continuous appraisal of census questions to ensure topics required by users are covered, while still satisfying the demand for data which is compatible with historical data.

(ix) The cost to the consumer (data user) must be kept to a realistic level, particularly as respondents are not paid for furnishing the data.

The demand for statistical data on the population and its characteristics is ever increasing, while the desire for personal privacy on the part of respondents is steadily growing. Collectors of census data face an interesting future, in which these conflicting demands must be addressed and reconciled.

Appendix 1.1

SELECTED EXTRACTS FROM SELECT COMMITTEES REPORTS ON NEW ZEALAND; ESTIMATES OF THE SIZE OF THE NON-MAORI AND MAORI POPULATIONS

...

J.L. Nicholas: We arrived in New Zealand the latter End of December 1814 and left it the latter End of February in the following Year; I was there about Ten Weeks.

...

Earl of Devon: You went from New South Wales?

J.L. Nicholas: We did, to the Bay of Islands, where the first Missionary Establishment was settled.

...

Earl of Devon: Had you an opportunity of judging whether the Island was thinly or thickly peopled at that Time?

J.L. Nicholas: I should say it was very thinly Peopled, considering the Extent of the Island. The Villages we came to were small and contained but a scanty Population. It is impossible to give any correct Account of what the Population might be. I think Foster, who accompanied Captain Cook, supposed the Population of the Northern Island to be 100,000; in the Book I wrote when I came back I put it down at 150,000; it is of course all Guess-work, but the Population is well ascertained to be very inadequate to the immense Extent of the Country.

...

Earl of Devon: Did the Missionaries settle there?

J.L. Nicholas: They did; three Missionaries with their Families, which constituted the first Mission of the Island, were settled in the Bay of Islands.

Earl of Devon: You state that when you were there, in the End of 1814 and the Beginning of 1815, you thought the Country

was very healthy; did you observe many old People among the Natives?
J.L. Nicholas: No; some, but not many.
...
Earl of Devon: You have stated that they are very healthy People. Is it a fact the a great Depopulation has been going on of late Years?
J.L. Nicholas: I understand very much so.
Earl of Devon: Has that been to a great Extent?
J.L. Nicholas: I only know from reading Publications.
...
Earl of Devon: Was there any Village of considerable Size in the Bay of Islands at that Time?
J.L. Nicholas: The largest Village in the Bay of Islands was close to where the Missionaries purchased this Land, and which I think contained a Population perhaps of about 200 People.
Earl of Devon: You have spoken of the Areekee, the principle Chief; do you know what the Amount of the Tribe under that Areekee was, in the Neighbourhood of the Bay of Islands?
J.L. Nicholas: I think it was said the Shunghi, the Areekee with whom we came much into contact, could muster a Thousand Warriors. We went to visit his Fortress in the Interior, a large fortified Place; but that did not contain, I should think, more than Three to Four hundred People; but a number of Villages and a very large Extent of Country belonged to him.

...
Earl of Devon: Have you at any Time been in New Zealand?
J. Watkins: I was there in the Years 1833 and 1834.
...
Earl of Devon: How long did you remain there?
J. Watkins: About Three Months altogether in New Zealand collectively. I availed myself of the Opportunity of traversing the Country and searching for Flowers and natural Curiosities, - Botanizing; these were my Objects.

Earl of Devon: Did that lead you to walk about the Island a good deal?
J. Watkins: It led me to walking a good deal; Forty or Fifty miles into the Interior, in various Directions, about the Bay of islands. I went over to Hukianga.
....
Earl of Devon: Can you form any judgement as to the Amount of European Population when you were there?
J.Watkins: That must be of course a very rough Estimate. The runaway Sailors and Convicts, and all that low Class, may amount to 400 or 500 Individuals; I should fancy that from One or Two being in every Tribe, and in many Five or Six, exclusive of the Missionaries, and who call themselves respectable English People; those have Shops there, and Stores for Ships, and such like Things.
....
Earl of Devon: All the Missionary Establishments are in the North Island, are they not?
J. Watkins: They are.
Earl of Devon: Have you been on the South Island?
J. Watkins: I have not. The Missionaries themselves, I believe, have not been much upon the South Island; it is very little known, I believe.
....
Earl of Devon: When you were there, in 1834, you state that the European Population consisted of Five or Six Traders, the Missionaries, and 400 or 500 European Sailors and Convicts?
J. Watkins: Yes. There were other English Residents in Hokianga and Wangaroo, and various other Parts, and about the Thames; there were a few there again of Traders, I cannot exactly enumerate the Traders, but there were very few.
Earl of Devon: Have you seen many old People there?
J. Watkins: I have seen some.
Earl of Devon: Did they appear a long-lived Race?
J. Watkins: Their Hair was white, and they appeared to be aged.
....

Earl of Devon: Your Evidence applies principally or entirely to the Northern Part of the North Island?
J. Watkins: Yes; the Southern Island, I understand, from what I have heard, is very little known; I have not visited any Part of that.
....

Earl of Devon: When did you go first go out to New Zealand?
J. Flatt: I went out to New Zealand in December 1834.
Earl of Devon: How long did you remain?
J. Flatt: Till May 1837.
....
J. Flatt: I left shortly afterwards, on a visit to England, for the purpose of being married, and I cannot say what has been done since. I came to the Bay of Islands and remained there from February to May.
...
Earl of Devon: During the time you were there, what do you consider the Extent of the European Population; did it increase or diminish while you were there?
J. Flatt: The Families of the Missionaries increased rapidly; there were upwards of 100 children before I left.
Earl of Devon: How was the other Population?
J. Flatt: There were very few of the other Europeans married...
Earl of Devon: What do you consider to be the Number of Europeans in the Island when you were there?
J. Flatt: I consider there were 500 Convicts and runaway Sailors on the Seacoast; not in the Interior.
...
Earl of Devon: With regard to the Traders who are settled in the Northern Part of New Zealand, what may be their Number?
J. Flatt: I believe the Number is greater than has been stated; they are settled in every Bay; not only in the Northern Bays but in the Southern Island.
Earl of Devon: Will you confine yourself, at first, to the Northern Island?

J. Flatt: I cannot give a distinct Statement of the Number of them.

Earl of Devon: A Witness has stated that he did not consider the Traders in the Bay of Islands as more than Five or Six?

J. Flatt: Not the respectable Traders in the Bay of Islands; only the runaway Convicts have begun carrying on Trade to a large Extent, and some of the Sailors as well.

Earl of Devon: Do you think there are as many as Fifty or Sixty respectable Traders in the Bay of Islands?

J. Flatt: No; but there may be that Number including all the Stations. There some in the Neighbourhood of the Wesleyan Station.

Earl of Devon: As to the Proprietors of Land; are there an Europeans who are Proprietors of Land except the Missionaries?

J. Flatt: Yes, some.

Earl of Devon: Are there many?

J. Flatt: No, not that I am acquainted with; but the major Part of my Time has been spent in the Interior.

Earl of Devon: Are there any other Descriptions of Europeans except Bay of Islands, Traders, and Runaways, and their families?

J. Flatt: No.

....

....

Earl of Devon: Had you any Opportunities of forming any Opinion as to the Extent of Population as compared with the Extent of the Country?

J.B. Montefiore, Esq: That has varied very much. I have heard many say it was 1,000,000; I have heard others say 500,00; but I think it is impossible to state the Fact.

Earl of Devon: Did it appear thickly or thinly populated, according to the Extent of Surface?

J.B Montefiore, Esq: Very thickly; I have seen as many as 2,000 or 3,000 Natives together in particular Parts.

Earl of Devon: Do you mean that generally, with reference to the whole Surface of the Country throughout, it is thickly peopled?
J.B. Montefiore, Esq: No; I have seen Numbers collected for particular Purposes.
Earl of Devon: The Length of the Island is 800 or 900 Miles; taking the whole Surface, is the Population large with reference to the whole Extent of Country?
J.B. Montefiore, Esq: The Population is very large, for I have heard them say, "Give us two or three Days, and we will get - such a Number, say some Thousands, - together." The Population of the Northern Island is very great; the Southern is much more thinly populated; quite a different Race of men...
....

....
J.S. Polack: ... there are about five Natives to every three square Miles of Land. The Northern Island is the most populous; at the same time it is the smallest. I have been many Miles without seeing a Native; I have been many Nights in the Bush without the chance of seeing a Native.
Earl of Devon: Has the Influx of Europeans been greater or less of late Years?
J.S. Polack: The Influx of Europeans have been wonderfully increasing... Queen Charlotte Sound, in the Southern Island, Cloudy Bay, Otargo, and all down to the South West is inhabited by Europeans there for the last Five and Thirty Years past, what the Natives call Kou Matuas; that is, old Men living there for the last Forty Years on the Coast. There are innumerable Europeans; they were principally Sealers; lately Whalers.
....

Earl of Devon: At what Period of the year were you there?
Rev. F. Wilkinson: From February to the 17th of May.
....

Earl of Devon: Were you there much towards the Bay of Plenty?
Rev. F. Wilkinson: I was there about Six Weeks.
....
Earl of Devon: Is there much European Population on the West Coast, where the Wesleyans are?
Rev. F. Wilkinson: Yes; a good deal, but not so much as the the Bay of Plenty. They were generally in separate small Groups...
....
Earl of Devon: Where is it that the Europeans principally live?
Rev. F. Wilkinson: At Kororaika and Otoiku...
....
Earl of Devon: You do not know the Amount of the Population of that Class?
Rev. F. Wilkinson: No; if I were to guess I should say about Five Hundred.
Earl of Devon: Are they permanent Settlers, or do they belong to Ships which come in there to trade?
Rev. F. Wilkinson: A good many of them are permanent Settlers
...

....
Earl of Devon: Do you happen to know what the Number of Europeans Landholders in the Islands is?
Hon F Baring, M.P.: I cannot tell what the Number of Europeans is; the Number of Europeans, as it is represented to us, increases; there are a Number of Europeans settled, but not for permanent Purposes, such as the Whalers and the Sealers, and other who are established for some temporary Object. The Number of those of whom any Census has been take is about 1,800 by the last Accounts.
Earl of Devon: They are not exclusively Land Owners?
Hon F Baring, M.P.: No; a great number are Grog-shop Keepers, and employed by the Shipping.
....
Earl of Devon: Are the Europeans Settlers on the Islands increasing in Number?

Hon F Baring, M.P.: They are increasing very rapidly, and the Natives are decreasing; a Letter lately received from a Mr Stack still more confirms that.

....

Hon F Baring, M.P.: ...the best Estimate we can get of the Population of the Two Islands does not set it above 150,000....the White Population is on the increase....

....

....

Earl of Devon: (To Coates) You are Secretary to the Church Missionary Society?

D. Coates, Esq: I am Lay Secretary to the Church Missionary Society.

Earl of Devon: (to Rev. Beecham) You are one of the Secretaries to the Wesleyan Missionary Society?

Rev. J. Beecham: I am.

Earl of Devon: Can you state to the Committee what are the Number and the State of the Native Population of New Zealand?

D. Coates, Esq: I have some Information on that Subject; it is contained in a Letter from the Reverend William Williams, one of the Missionaries of the Church Missionary Society in New Zealand, which bears Date the 10th February 1834. He states, "I believe the Population of this Island does not exceed 106,000, of which about 4,000 are in connection with our Station at Kaitaia to the Northward, 6,000 with the Wesleyan Station at Hokianga, and 12,00 connected with our Four Stations in the Bay of Plenty. The Number in the Thames is about 4,800; while those at Waikato, a District in the same Parallel coast of the Bay of Plenty, as far as Hick's Bay, are about 15,600. From Hick's Bay to Hawkes' Bay the Number is about 27,000, concentrating in Two principal Places. There are now on other Inhabitants in the Southern part of the Island, except in the Neighbourhood of Entrey Island, where the Number is 18,000. " This is the most distinct Statement of the Population of the Northern Island of New Zealand which I have ever seen, and appears to have been formed, by the tenor of Mr

Williams's Letter, with considerable Care, and therefore probably approximates to the Truth. In a Letter, written 4th September 1835, Mr Williams says, "The Population of the two Islands is small, not exceeding 200,000." We have no Missionary stationed on the Southern Island, therefore our Information respecting it is not of a detailed Character.

Earl of Devon: (to Rev Beecham) Are you able to speak to these Facts?

Rev. J. Beecham: We have not any specific Information from our Missionaries on that point. We have understood that the Population at Hokianga, in connection with our Mission there, may be about 5,000. I should incline to the Opinion, from the Information which I have had from various Quarters, that the Statement made by Mr Coates respecting the total Amount of the Population is perhaps as near to the Truth as can be obtained.

Earl of Devon: (to Coates) Have the Accounts you have received led you to believe that the Population has increased or decreased since 1835, or that it has a tendency to increase or decrease?

D. Coates, Esq: I am quite aware there are Causes in operation in New Zealand which necessarily tend to the Decrease of the Population ... But though those causes are in operation, there are some Considerations which lead me to infer that the Depopulation has not been to that Extent which has been represented... I find that the Tribes are migratory; a certain District will be observed to be thickly populated at one Period, and being visited at a subsequent Period, will be found to have scarcely any Population at all... I understand that Doctor Foster, who was with Captain Cook, and who estimated the then Population of the different Islands in the South Seas, stated that of the Northern Island of New Zealand at 100,000. Now if that Estimate approximated at all to the Truth, and I think the Presumption is that it rested on less satisfactory Data than that of Mr Williams, which I have stated to the Committee, it will be found that the Population in 1834, the date of Mr Williams's Letter, was to the full extent as considerable as it

was supposed to be by Doctor Foster when Captain Cook visited the Island...

Earl of Devon: Have you any Information from the Missionaries upon that point?

D Coates: Yes; the increasing Depopulation of the Island is stated in our Information.

Earl of Devon: The Opinion of the Missionaries who are stationed there is that the Population is wasting away?

D Coates: Yes.

Earl of Devon: It is the Opinion of the resident Missionaries that the Population is diminishing there, unquestionably.

....

Earl of Devon: *(to Coates)* You have no doubt that the irregular Population of the Whites is increasing?

D Coates, Esq: That it is increasing I think most probable, but that its Amount has been exaggerated I am very much impressed with the Persuasion. I have seen a Statement which spoke of 2,000. I certainly have the strongest Impression upon my own Mind that it does not amount to more than One Fourth. I have no very accurate Information, but I observe in the Petition appended to Dr. Hind's Pamphlet that the Population Northward of the River Thames is stated to amount to 500. Now unquestionably that Part of the Island is that in which this Population abounds, and if according to the latest and most authentic Statements, as far as I am aware, they are spoken of as about 500, I think that the Presumption is that the total Number is not very much beyond that.

....

....

Earl of Devon: Are you aware of the Extent of the White Population in New Zealand?

W.A.Garret: I have no distinct Knowledge upon the Subject, but I am aware that there are Differences of Opinion on the Subject.

Earl of Devon: You are aware that it is an increasing Population?

W.A. Garret: It is so considered.

....

....

Earl of Devon: Mr Busby, the Resident, in his Report to the Governor of New South Wales, states, "In this way has the Depopulation of the Country been going on, till District after District has become void of its Inhabitants; and the Population is even now but a Remnant of what it was in the Memory of some European Residents. "Do you consider that to be an exaggerated or true Statement of the Condition of the Country? *Captain R Fitzroy, R.N.:* I am not able to say, decidedly; but the general Opinion when I was there was, that the Population was decreasing very fast. The Natives themselves said so; their common Expression was (in their own Language), "the Land is not for Us; the Land is for the White Men", alluding to the Depopulation.

....

Source: *British Parliamentary Papers/Colonies/NZ/1/Sessions 1837-38,40*
Reports from Select Committees on New Zealand
The State of the Islands of New Zealand, 3-340.

Appendix 1.2

SELECTED EXTRACTS FROM SELECT COMMITTEES REPORTS ON NEW ZEALAND; STATEMENTS ABOUT CHARACTER OF NON-MAORIS IN NEW ZEALAND

....

Earl of Devon: Can you form any Judgement as to the amount of European Population when you were there?

J Watkins: That must of course a very rough Estimate. The runaway Sailors and Convicts, and all that low Class, may amount to 400 or 500 Individuals; I should fancy that from one or two belong in every Tribe, and in many Five or Six, respectable English People; those have Shops there, and Stores for ships, and such like Things...

J Watkins: ....The Missionaries have immense Influence among the Natives; they are respected there as much as any Gentleman of character are respected here, and a good deal more; indeed I may say they have unlimited Influence.... The Missionaries have much Influence among the Natives, as teaching them the various Arts - Carpentry and Blacksmithery, and so on. I think the Natives esteem the Missionaries as much for their moral Character as their mechanical Knowledge... Nothing can be more lawless than the Europeans who are there...

Earl of Devon: You have spoken of the general Immorality of the Settlers; besides that, is not open Violence frequent, such as Theft, Robbery, and so on?

J Watkins: Yes.

Earl of Devon: Has that been directed against the Settlers themselves and against the Zealanders?

J Watkins: ...Yes. Mr Mayor was in the habit of having his Stores broken open by the runaway Sailors; and he had a Band of Natives to protect his Stores, who kept watch for him.

Earl of Devon: What Proportion of European Women are there among the lower Class of Sailors?
J Watkins: ...None at all. The European Women, if they go there, will not stay, I never know any Instance but One of that kind. She was in the Country with an English Person; she came from Hobart Town, and went away with him.
Earl of Devon: There are some European Women, the Missionaries Wives?
J Watkins: ...Yes, and a few Others.
Earl of Devon: Those are respectable Persons?
J Watkins: ...O Yes, their Character is unblemished.
....

Earl of Devon: What do you consider to be the Number of Europeans in the Island when you were there?
J Flatt: I consider there were 500 Convicts and runaway Sailors on the Sea Coast; not in the Interior.
Earl of Devon: In what sort of Way do they live?
J Flatt: They lead a most reckless Life, by keeping Grog Shops, selling Spirituous Liquors, both to Europeans and Natives...
Earl of Devon: Is there much of Violence and Theft among the European Population?
J Flatt: They frequently fell out; while I was at the Bay there were two Murders.
Earl of Devon: And no Punishment for them?
J Flatt: No; Mr Busby had not the Power.
....

....
J.B Montefiore, Esq: ....We have in very great Measure lost our Character with the New Zealanders, in consequence of the very bad Character of some runaway Convicts, refractory Seamen, and others who are now residing there. I think, with the exception of Myself and some Others, very few Gentleman have ever visited the Country. There are a great many bad Characters on the Island...

Earl of Devon: Nothing can be worse than the State of Society in the Bay of Islands?
J.B Montefiore, Esq: I understand it is as bad as it can be; it consists of Persons who have run away from Sydney in Debt, Convicts, and very bad Characters who have left Whaling Stations, and the worst of Characters that can be.
....

....
Earl of Devon: What is the state of Manners or Morals among the European Population generally?
J.S. Polack: Decidedly bad.
Earl of Devon: There exist no Law to control or correct them?
J.S. Polack: None, but that of Force.
....

....
Earl of Devon: Were you much towards the Bay of Islands?
Rev. F Wilkinson: I was there about six Weeks.
Earl of Devon: What was the State of the European Population there?
Rev. F Wilkinson: I think it is as bad as can be. I was at Koroika. I do not know that I ever saw such a bad Community; there was Drunkenness and Profligacy of all kinds.
Earl of Devon: Is there much European Population on the West Coast, where the Wesleyans are?
Rev. F Wilkinson: Yes; a good deal, but not so much as at the Bay of Islands. They were generally in separate small Groups. They go there, not be make Money, for that they cannot do, but to live in a loose profligate Manner. They have easily found a Living there. They are generally Persons who have been obliged to get away from New South Wales on account of Debt.
....
Earl of Devon: The Population in the Bay of Islands, as far as the Europeans are concerned, are outcasts of Society, the Majority of them, are they not?
Rev. F Wilkinson: Yes; the Mixture of Americans and English. I do not think that they are Convicts, many of them. It is

supposed there are a good many Convicts that have escaped from the Colony; but I do not think that.
Earl of Devon: But Men quite as immoral as the Convicts?
Rev. F Wilkinson: Quite as immoral.
...
Earl of Devon: When you were there was Mr Busby there?
Rev. F Wilkinson: He was.
Earl of Devon: Where did he reside?
Rev. F Wilkinson: He resided near Waitangi; I forget the name of the Place.
Earl of Devon: Is that in the Bay of Islands?
Rev. F Wilkinson: Yes.
Earl of Devon: Does he exercise any Authority over these casual Settlers?
Rev. F Wilkinson: Not any at all.
Earl of Devon: Do the Chiefs retain any Authority, and exercise any Authority over them, in the Country round the Bay of Islands?
Rev. F Wilkinson: I do not think they do at all.
Earl of Devon: At present they are under no Law whatever?
Rev. F Wilkinson: They are under no Law whatever.
Earl of Devon: You do not know the Amount of the Population of that Class?
Rev. F Wilkinson: No; If I were to guess I should say about Five Hundred.
Earl of Devon: Are they permanent Settlers, or do they belong to Ships which come in there to trade?
Rev. F Wilkinson: A good many of them are permanent Settlers, but not respectable Settlers; several are Persons who are respectable, and who have permanent Establishments for the supply of Ships; but there are a Parcel of Fellows who keep Grog Shops and Beer Shops, who are Runaways from Ships.
...

....
W.A. Garratt, Esq: ....those who settle there now are principally of two Classes; first, the Missionaries, and those connected with them, whom for that Purpose I may throw very much out

of the Question; and then there are those who go there unconnected with the Missions, and for Purposes of their own, to pursue agricultural or commercial Objects, or rather to run away from Ships or from other Settlements. I apprehend there would be very few respectable Persons who would settle in New Zealand in the present unsettled State of Things, while there is no British Government to protect them.

Source: *British Parliamentary Papers/Colonies/NZ/1/Sessions 1837-38,40*
Reports from Select Committees on New Zealand
The State of the Islands of New Zealand, 3-340.

Appendix 1.3

EXCERPTS FROM THE NEW ZEALAND STATISTICS ACT 1975

The Statistics Act (1975) , Part III, provides the legal basis for the Census of Population and Dwellings of New Zealand. Section 23(1) states:

The Census of Population and Dwellings of New Zealand shall be taken by the Department in the year 1976 and in every fifth year thereafter.

The questions which are mandatory (i.e. must be included) in every census are listed in Section 24, subsection (1), which states:

At every census of population and dwellings particulars relating to all of the following matters shall be obtained from every occupier or person in charge of a dwelling:

(a) The name and address, sex, age, and ethnic origin of every occupant of the dwelling:

(b) Particulars of the dwelling as to location, number of rooms, ownership, and number of occupants on census night.

Section 24 (2) lists standard questions (i.e. are considered to be of sufficient importance to be included in every census), and states:

At any census of population and dwellings the Statistician may, if he considers it in the public interest so to do, obtain from every occupier or person in charge of a dwelling particulars relating to all or any of the following additional matters:

(a) The profession or occupation and industry in which employed, nationality and citizenship, health, marital condition, religion, birthplace, duration of residence in New Zealand, address where living at previous census or previous year, number of children, number of hours worked per week for wages or salary or financial reward, status in employment, name and address of employer, mode of transport to and from work, time taken to travel to work, income, address of usual residence, and service in the armed forces of every occupant of the dwelling:

(b) Particulars of the dwelling as to type and tenure of dwelling and nature of materials of structure, household amenities, rent paid, and details of any livestock:

(c) Any information relating to the kinds of statistics for which information may be required pursuant to section 4 of this Act or as may be prescribed by regulations under this Act.

Other relevant sections of this Act include the appointment of Enumerators, Sub-enumerators, etc. (Section 19), Evidence of appointment (Section 20), Declaration of Secrecy by every employee (Section 21), Duty of persons to obtain census schedules (Section 25), Duty of the occupier and other persons in the dwelling (Section 26), Security of information (Section 37), Failure of Enumerator, Sub-enumerator to carry out duties (Section 39), Neglect or refusal to supply particulars (Section 43), and General Penalty (Section 47).

Source: *The Statutes of New Zealand 1975*, *1*, 16.

Appendix 1.4

DATES OF NEW ZEALAND CENSUSES 1851-1986

Date of Census	Period since last Census (Years)
November-December 1851	-
24 December 1858	1.75(1)
16 December 1861	2.98
1 December 1864	2.96
19 December 1867	3.05
27 February 1871	3.19
1 March 1874	3.01
3 March 1878	4.01
3 April 1881	3.08
28 March 1886	4.98
5 April 1891	5.02
12 April 1896	5.02
31 March 1901	4.97
29 April 1906	5.08
2 April 1911	4.93
15 October 1916	5.54
17 April 1921	4.50
20 April 1926	5.01
24 March 1936	9.93
25 September 1945	9.51
17 April 1951	5.56
17 April 1956	5.00
18 April 1961	5.00
22 March 1966	4.93
23 March 1971	5.00
23 March 1976	5.00
23 March 1981	5.00
4 March 1986	4.95

(1) Period since incomplete enumeration of March 1857.

Sources: *Population Perspectives '81, New Zealand Census of Population and Dwellings 1981,* 1985, *12* , 146 (Department of Statistics, Wellington).

Questionnaire Content and Submissions, *New Zealand Census of Population and Dwellings Census '86* , 1985, 118 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings Census '86 General Information , 1988, *Series D Report 3,* 9 (Department of Statistics, Wellington).

Appendix 2.1

SOME APPLICATIONS OF DATA FROM THE NEW ZEALAND CENSUS OF POPULATION AND DWELLINGS 1981

PERSONAL QUESTIONNAIRE

Question Number	Content	Purpose
1.	Full name	Identifier for linkage with other questionnaires and other surveys.
2.	Sex	Population profile - sex composition of population, fertility trends.
3.	Date of birthday	Confirmation of accuracy of next question - if conflicting data, date of birthday taken as correct.
4.	Year born or age last birthday	Cohort fertility rates, population profile - age distribution, assessment of current and future needs for kindergartens and day-care facilities, prediction of future primary and secondary school rolls, prediction of future rolls at tertiary institutions (universities, teachers colleges, polytechnics or technical institutes, business colleges, nursing schools, pharmacy schools, community colleges), working-age population, the elderly population, estimates of "at risk" populations with special needs for local authority services - for instance, current and future health needs for the

		elderly, estimates of number of current and future pension payments.
5.	Relationship to occupier	Family structure, trends in family formation and dissolution - elderly living alone, single parent families, proportions of family, non-family and one-person households by age groups and sex, growth in non-family households, persons usually living alone.
6.	Full address on Census night	Identifier, regional distribution of population, rural and urban population growth, regional differentials in population growth, composition of small-area populations - used for development of community health care facilities, current and future planning for educational facilities, future housing, shopping, commerce needs.
7.	Usual residential address	Population statistics on de jure basis.
8.	Usual residential address one year ago	Migration trends - for example from rural to urban areas, from South Island to North Island, from areas outside Auckland to Auckland.
9.	Usual residential address at previous census	Migration trends, linkage to questionnaires from previous census.
10.	Country of birth	Long-term migration statistics.
11.	Religious denomination	Trend in religious affiliations.
12.	Ethnic origin	Ethnic distribution, ethnic growth, Maori and non-Maori fertility declines, age distribution of ethnic groups, rural-urban

		distribution of ethnic groups, distribution of ethnic population in urban areas, secondary school attendance by ethnic and age groups, tertiary attendance by ethnic groups and sex labour force participation by ethnic groups and sex, age-specific fertility by ethnic and age groups, unemployment rates by ethnic groups and sex, marital status by age, sex and ethnicity, family households by type and ethnic group.
13.	Cigarette smoking	Profile of cigarette smokers.
14.	Present marital status	Trends in marital status - comparison of numbers of persons married, separated and divorced with those enumerated in earlier censuses, both for total populations and by age groups and ethnic origin, trends in family formation and dissolution, trend towards de facto marriages, marital status of de facto couples, comparison of one- and two-parent families.
15.	Number of children born	Fertility trends, family size trends - including assessment of trend towards childless families.
16.	Hours worked per week	Trend in number of hours worked per week, part-time labour force by marital status, sex, age groups and occupation, female participation rates by age of youngest child, number of full-time workers with part-time jobs by sex.

17.	Employment status	Current and future composition of labour force, labour force participation of mothers in one- and two-parent families, median age of the labour force by sex, trend in composition of labour force by age groups and sex, trend in unemployment rates by sex, unemployment rates by sex and age groups, regional unemployment, marital status of employed and unemployed, living arrangements of unemployed, educational qualifications of unemployed by sex, highest educational level of unemployed by sex, percentage unemployed in each sex, age group and highest school attendance group, percentage of unemployed in each sex and marital status group receiving social welfare benefits, estimates of "at risk" populations from unemployment indicators, small area unemployment and employment indicators.
18.	Occupation	Profile of workforce - growth of workforce, occupational distribution of workforce, participation rates by age groups, employment in community services, male-female occupational differences, occupational effect on median incomes.
19.	Name of employing organisation	Identifier for linkage with other surveys.
20.	Address of workplace	Identifier for linkage with other surveys.
21.	Type of work carried out by employing organisation	Industrial distribution on small-area and regional basis.

22.	Main means of travel to work	Trend in transportation - future demands for road networks, parking, public transport.
23.	Social security benefits	Profile of welfare payments - numbers and types, proportion of population receiving welfare payments, percentage of unemployed in each sex and marital status group receiving social welfare benefits.
24.	Income from social sec. benefits	Profile of incomes, low-income households, comparison with Social Welfare expenditure records, estimates of "at risk" populations from poverty indicators.
25.	Income from other sources	Profile of incomes - personal and family, non-benefit income distribution for both unemployed and employed, household income by age and ethnic groups.
26.	Highest level attended at school	Profile of education level achieved at secondary level, percentage unemployed in each sex, age group and highest school attendance group.
27.	Highest school qualification gained	Profile of educational attainment - measure of success of secondary education, educational qualifications of unemployed by sex, highest educational level of unemployed by sex.

28.	Other places of education attended	Profile of past and current attendance at university, teachers college, polytechnic or technical institute, business college, nursing school, pharmacy school, community college.
29.	Academic, vocational, or professional qualifications gained since leaving school	Profile of education level of population achieved at tertiary level.

DWELLING QUESTIONNAIRE

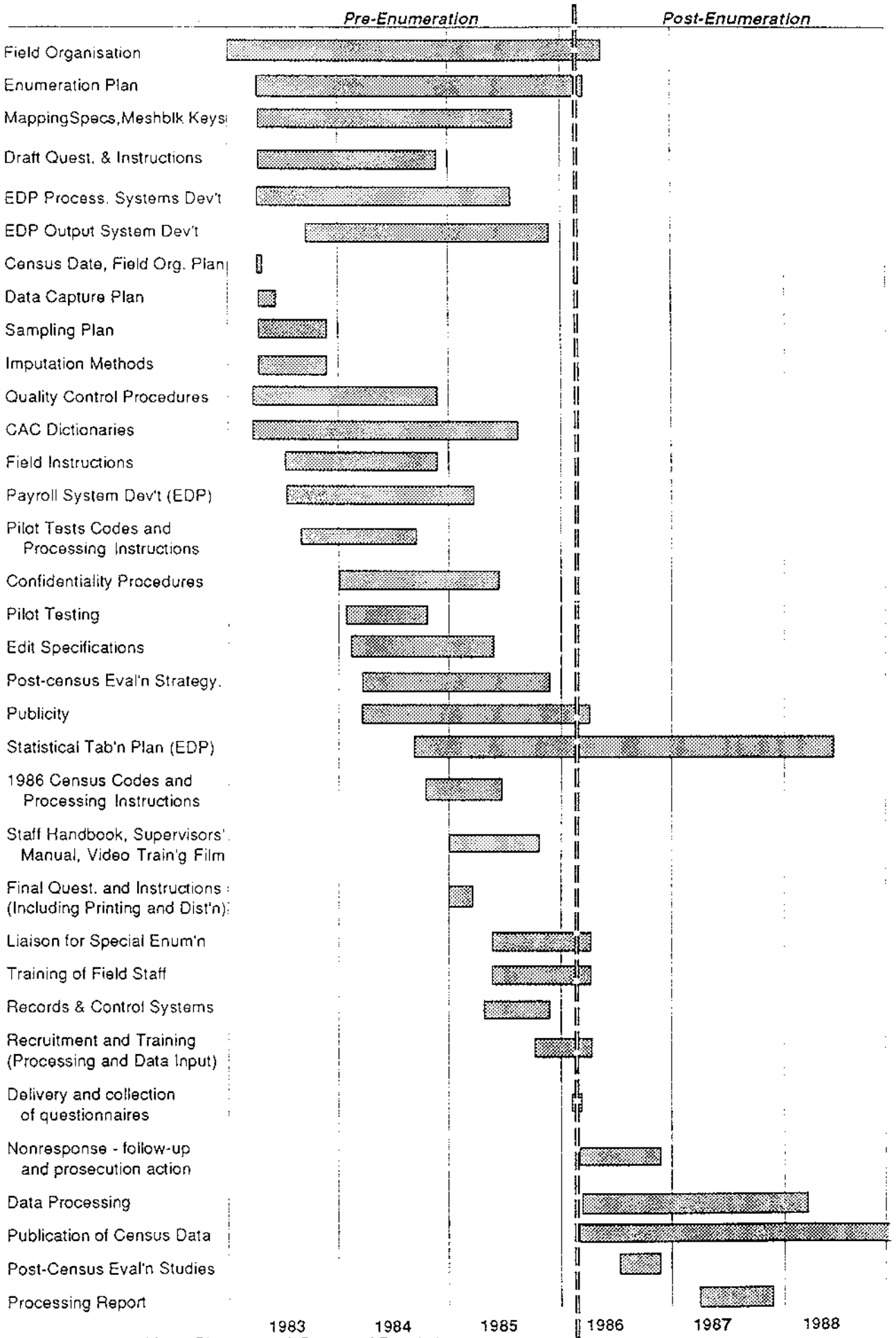
Question Number	Content	Purpose
1.	Name of occupier	Identifier.
2.	Full address of dwelling	Identifier, small-area and regional statistics on housing density.
3.	Number of occupants on Census night	Occupancy and 'crowding', crowding and ethnicity, average household size.
4.	Type of private dwelling	Number of households in private dwellings, trends in dwelling types, numbers of rooms in private dwellings, dwelling insulation by regions.
5.	Type of dwelling if other than a private dwelling	Distribution of dwelling types on small-area and regional basis.
6.	Principal means of cooking	Statistics on heat sources used for cooking.
7.	Type of hot water supply	Statistics on sources used for heating water.

8.	Heating of dwelling	Statistics on sources used for heating dwelling.
9.	Tenure of dwelling	Trend in tenure of private dwellings, tenure by marital status and age groups, home ownership by income, age and ethnic group, home ownership of women by age groups, major tenure and landlord groups housing corporation tenants by ethnic and age groups.
10.	Rent	Median rents by landlord types, median rents by urban areas and furnished-unfurnished properties, characteristics of renters by age, ethnic groups and sex, tenure and landlord characteristics by age groups and sex, relative rents for male renters by age groups and ethnic origin, relative rents by age groups of male renters and landlords, relative rents for female renters by age groups (sample too small to permit analysis of rentals for females by ethnic groups).
11.	Roof material	Statistics on various types of materials used for roofing.
12.	Material of outer walls	Statistics on various types of materials used for roofing.
13.	Number of rooms	Number of rooms in private dwellings, socio-economic indicators.
14.	Heat insulation	Proportion of private dwellings with wall and ceiling insulation.
15.	Amenities present in dwelling	Socio-economic indicators.

16.	Holiday residence	Socio-economic indicators, linkage to personal questionnaires for coverage assessment.
17.	Number of vehicles, caravans, and boats	Socio-economic indicators, trends in car ownership and strategic implications for the road network.
18.	Persons absent on Census night	Linkage to personal questionnaires for assessment of coverage.

Source: *Population Perspectives '81* New Zealand Census of Population and Dwellings, 1981, *12* , 1-144, 158-169 (Department of Statistics, Wellington).

RELATIVE TIMING OF VARIOUS STEPS IN 1986 NEW ZEALAND CENSUS OF POPULATION AND DWELLINGS



Source: Adapted from *Plan for 1986 Census of Population and Dwellings*, 1983, 71-79 (Department of Statistics)

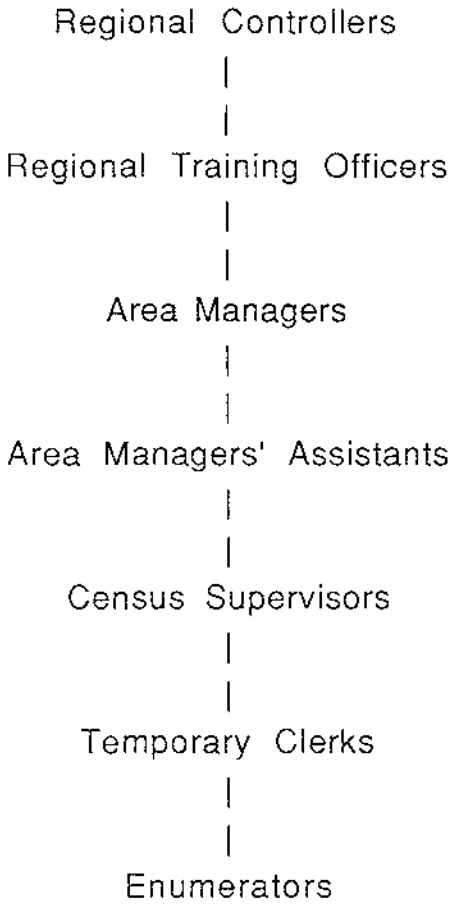
Appendix 3.2

ORGANISATION OF THE ENUMERATION PHASE OF THE NEW ZEALAND CENSUS OF POPULATION AND DWELLINGS 1986

For all New Zealand censuses prior to 1986, the staff and facilities for organising and controlling the census field work have been provided by the postal sector of the Post Office (now New Zealand Post). The postal sector has an established network of facilities and staff who are thoroughly familiar with their local areas, and who generally have been involved with previous census field work. Senior officers at selected Post Offices were appointed by Post Office Headquarters, upon approval by the Department of Statistics, to act as Census Enumerators. Each Enumerator selected Sub-enumerators for his or her District and, upon approval by the Department of Statistics, trained the Sub-enumerators and supervised their fieldwork. However, the postal staff have been required to perform the census duties in addition to their normal workload, and in 1981 the Post Office notified the Department of Statistics that all census work would be done outside normal office work hours and that, in some districts, postal staff would not be available for enumeration duties. The increased overhead costs incurred by paying postal staff overtime rates and training non-postal personnel as enumerators in several districts led to the break with tradition, and the Department of Statistics decided to undertake the enumeration phase of the 1986 Census itself. All census field staff were selected and trained by the Department, and the Post Office provided the normal postal facilities, local knowledge, and issued questionnaires on request to members of the public who for some reason had not been supplied with questionnaires by their local enumerator.

Prior to the 1986 Census, Enumerators were assigned Enumeration Districts, and Sub-enumerators were assigned Sub-districts. This procedure was revised for the 1986 Census, when Districts were controlled by Census Supervisors, the Sub-districts became the responsibility of the Enumerators, who supervised the house-to-house enumeration by the Sub-enumerators.

The management structure for the 1986 New Zealand Census of Population and Dwellings was as follows:



Sources: *New Zealand Census of Population and Dwellings 1981 Enumerator's Handbook* , 1980, 38 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings Census '86 Enumerator's Handbook , 1985, 11 (Department of Statistics, Wellington).

Appendix 3.3

DUTIES OF SUPERVISORS AND ENUMERATORS FOR THE NEW ZEALAND CENSUS OF POPULATION AND DWELLINGS 1986

The New Zealand Department of Statistics employs Census Supervisors and Enumerators under contract for the period necessary for the execution of the census. All employees of the Department of Statistics are legally bound by the statutory Declaration of Secrecy, both during the period of employment, and at all times thereafter. To aid identification, each Census Supervisor and Enumerator wears an identity badge when performing official duties, and is also supplied with a Certificate of Authority. All information collected or entered on the questionnaires must be kept confidential, and the Census Supervisors are not permitted to obtain unauthorised information. Census staff are not permitted to enter a dwelling if there is no occupant present.

Census Supervisors are responsible for the selection, and, upon approval, the training and performance of the Enumerators. They also visit all non-private dwellings, such as hospitals, prisons and temporary work camps, in their Census District and acquaint the person in charge with their responsibilities. Persons in charge of **non-private dwellings**, or **group-living quarters**, are responsible for ensuring that all persons in their establishment on census night complete and return a Personal Questionnaire, as well as those arriving the following morning who have not completed a Personal Questionnaire elsewhere, provided that they were in New Zealand at midnight of Census night. A Dwelling Questionnaire covering the establishment must also be completed.

Under the Statistics Act 1975, a **dwelling** is defined as *"any building or structure, whether permanent or temporary, which*

is wholly or partly used for living purposes". Thus the definition includes any shelter in which people are located on census night, whether it be a house, hospital, flat, boarding house, hotel, motel complex, hospital, prison, hut, tent, car, caravan, ship, which means that even people on an overnight tramp must be located and enumerated.

For census purposes, dwellings are divided into 2 categories: **group-living quarters** (non-private dwellings) and **private dwellings**. Group-living quarters cover institutions and other dwellings which have been designed to cater for large groups of individuals or a large number of families, and a private dwelling is any dwelling lived in by one private household and which has its own separate sleeping, cooking and dining facilities.

A **private household** comprises all persons who live on the same premises and who usually eat one or more meals together daily or, at least, *share the same cooking facilities.* Temporary guests will also be part of the household in which they are staying with. Persons in group living quarters are, for New Zealand census purposes, *not* regarded as living in 'private households'.

Maps, produced by the Department of Lands and Survey, are supplied to Census Supervisors and Enumerators. The maps show the boundaries of each Sub-district within the District, and are used to ensure that no area of land is either omitted or duplicated during enumeration. The Enumerator's maps show the exact boundaries of the Sub-district, together with the precise boundaries of each meshblock into which the Sub-district is divided. On average, each Sub-district contains 8 mesh-blocks.

Census Supervisors must make special arrangements for the enumeration of people on board ships, campers, travellers and homeless persons. It is their responsibility to instruct the Enumerators when to start delivery of the Questionnaires and

to supervise, monitor and control their progress. Census Supervisors open sealed envelopes received from their Districts, checking them for completeness. Each enumerator must endeavour to ensure that all the persons in the district at census time are counted, following up on contingencies such as persons shifting since census day who haven't completed a questionnaire, omissions in replies to the questionnaires, non-cooperative respondents and enquiries from the public about the census.

Census Supervisors despatch provisional population and dwelling totals for each Sub-district as they become available, recruit and instruct clerical staff to check all submitted questionnaires for correctness and completeness, and arrange for the return of Questionnaires to respondents when a major proportion of the questionnaire is incomplete. They also carry out follow-up work on the Summary of Outstanding Questionnaires List, supply a summary of population and dwellings in Sub-districts, send the census questionnaires to the Census Processing Section and follow-up and advise on queries from the Census Processing Section.

In an effort to ensure complete coverage, prior to the census, Enumerators list all private dwellings which will be unoccupied on census night, dwellings under construction, vacant sections, non-residential properties, business premises in rural areas, and diplomatic residences. It should be noted that Enumerators make no contact with diplomatic residences, as delivery and collection of questionnaires to these buildings is the responsibility of the staff of the Ministry of Foreign Affairs.

Every Enumerator has the responsibility of ensuring that a Personal Questionnaire is completed for every person in their Sub-district on census night, as well as for every person who arrives there before noon the next day without having filled in a questionnaire elsewhere (provided that those persons were in New Zealand at midnight on census night), and that a Dwelling

Questionnaire is completed for each separate dwelling which is used as living accommodation on census night.

Enumerators are required to insert the necessary reference numbers on questionnaires, which consist of the meshblock number on every Dwelling and Personal Questionnaire and an identical questionnaire number on every Personal Questionnaire issued at the same dwelling. They are also responsible for the safe custody of completed questionnaires.

If respondents do not wish their completed questionnaires to be viewed by another member of the household or by the Enumerator, they are permitted to seal them in envelopes which are endorsed with their names and the reference numbers (as printed at the top of the Personal Questionnaire).

To ensure maximum coverage, Enumerators are required to re-visit every dwelling within their Sub-district, including those previously classified as unoccupied on Census night. Dwellings which were inhabited between the delivery of the questionnaires and Census night and any other dwellings which were missed are then issued with questionnaires, and listed. The questionnaires are collected, and the Identity Numbers on each are checked, with particular attention given to any extra questionnaires which were obtained from a Post Office or brought to the dwelling by visitors.

Enumerators also check that a Dwelling Questionnaire has been completed by the person nominated as the 'occupier' of each private dwelling, that every person who was present in a private dwelling on Census night has completed a census questionnaire, that envelopes containing sealed questionnaires have been appropriately endorsed, that unsealed questionnaires are complete, that the number of questionnaires collected corresponds with the number delivered and that no flats, baches, or makeshift accommodation have been overlooked at an address. As mentioned earlier, Census Supervisors and Enumerators are not permitted to enter a private dwelling if

there are no occupants present. If all occupants of the dwelling are absent at a second or later visit, a 'Return Envelope' is left.

Sources: *New Zealand Census of Population and Dwellings 1981 Enumerator's Handbook* , 1980, 6-9, 97 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings Census '86 Enumerator's Handbook, 1985, 5-63 (Department of Statistics, Wellington).

Appendix 3.4

STRATIFICATION OF GEOGRAPHICAL AREAS OF NEW ZEALAND INTO STATISTICAL UNITS FOR 1986 CENSUS

Level 1

Census Regions

Administered by Area Managers.
Determined by density of population, number of field staff, topography, accessibility, etc.

Level 2

Census Districts

Areas administered by Census Supervisors
Generally follow boundaries of territorial local areas

Level 3

Census Sub-districts

Administered by Enumerators
Consist of a number of adjoining meshblocks
Average of eight meshblocks covered by each Enumerator

Level 4

Meshblocks

Areas of land whose boundaries are readily identifiable on the ground. The meshblock is the smallest areal unit for which the census data is compiled. The meshblocks may vary considerably in physical size, depending on the nature of the area. In urban areas, a meshblock may consist of a single street block, but in rural areas, it may cover up to 20 or 30 square kilometres, with the boundaries conforming to road patterns or to other topographical features.

Sources: *New Zealand Census of Population and Dwellings 1981 Enumerator's Handbook* , 1980, 8-10, 38-39 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings Census '86 Enumerator's Handbook , 1985, 11 (Department of Statistics, Wellington).

Appendix 3.5

NUMBERING AND BOUNDARIES OF SUB-DISTRICTS CREATED FOR NEW ZEALAND CENSUSES OF POPULATION AND DWELLINGS

The following guidelines were used by the Department during the creation of Sub-districts:

(i) Every Sub-district is one continuous area (that is, it cannot consist of two or more detached parts) and, except where a Sub-district comprises the whole of a very small **community** plus a portion of the surrounding **county**, a sub-district must be contained within a single local authority area.

(ii) Sub-district boundaries always follow meshblock boundaries, and in **urban areas** follow urban sub-division boundaries, although one sub-division may be divided into two or more Sub-districts when necessary.

(iii) The Sub-districts must together form the whole of the District.

(iv) Roads or streets are used for Sub-district boundaries in cities and boroughs, thus avoiding the problems of locating boundaries which run between sections. In **rural areas**, well defined and marked natural boundaries which can easily be identified on the map and on the ground by Sub-enumerators are used as Sub-district boundaries. Road boundaries are not desirable in thinly populated areas, as usage of them may result in two Sub-enumerators travelling long distances to enumerate a few houses on each side of the road.

(v) All Sub-districts within a District are numbered to a set pattern, according to local authority areas. Where a District contains parts of more than one local authority area, then all Sub-districts within one local authority are numbered consecutively from the last number in the previous local authority area, and so on. (For instance, if District 245 consisted of three Sub-districts in a borough, one in a community and 6 in a county, then the borough Sub-districts would be numbered 245/01, 245/02 and 245/03, the community Sub-district would be numbered 245/04 and the county Sub-districts would be numbered 245/05-10. If a community were not sufficiently large to be a separate Sub-district, it would be treated as part of the parent county for numbering purposes.

Sources: *New Zealand Census of Population and Dwellings 1981 Enumerator's Handbook* , 1980, 38-40 (Department of Statistics, Wellington).

Appendix 4.1

HISTORY OF QUESTIONS ASKED AT EACH NEW ZEALAND CENSUS OF POPULATION AND DWELLINGS, 1851-1986

Note: Unless otherwise stated, the questions listed below were included in both European and Maori Schedules. From 1951 onwards, a common schedule was issued to all respondents.

PERSONAL QUESTIONNAIRE

Question	*Period asked*	*Comments*
Full Name	1851-1986	
Sex	1851-1986	
Age	1851-1986	Asked in 2 parts in 1981 in order to improve the quality of response (asked to specify either year born or age at last birthday). In 1986 Census, only date of birth requested.
Marital Status	1851-1986	Prior to 1926 Census, only included in General European Census. For 1851-1916 and 1971-86 censuses, asked only of persons aged 15 years and over. For 1921-66 censuses, asked only of persons aged 16 years and over.
De Facto Status	1981,1986	

Duration of Current Marriage	1911-21	Included in General European Census, and asked only of married women. Used to ascertain fertility of marriage.
Whether married to a European	1921	Asked only of Maoris living in North Island and in Chatham Islands.
No. of Children		
- Born Alive within marriage	1911-21,1945, 1971-76	Asked only of married women. For 1911-21 censuses, only included in General European Census. For 1945 Census, also asked of Maoris living in the North Island and in Chatham Islands. In 1911-21, only children born alive to existing marriage were sought. In 1945 and 1971-76, children born alive within all marriages were sought. In 1971, question worded 'state the number of children born alive to you during your lifetime'. The resultant criticism caused the Department of Statistics to publish advice that only children born alive within marriage were required to be disclosed.
- Born Alive	1981	Asked of all resident females aged 15 years and over.
- Still Living	1911-21	Asked only of married women, and only included in General European Census.

No. of Children		
- Deceased	1911	Asked only of married women, and only included in General European Census.
- Dependant (i.e.under 16yrs)	1921-66	Prior to 1951, only included in General European Census. Each married man, widower or widow asked to state number of his or her living children under 16 years of age, including stepchildren and adopted children.
Orphanhood	1921-36	Only included in General European Census, and required only of persons under 16 years of age. War of 1914-1918 and influenza pandemic of 1918-19 resulted in large increase in number of orphans in New Zealand.
Infirmities	1911-16	Only included in General European Census.
Relationship to Occupier or to person in charge of dwelling on Census night	1851-1986	Prior to 1945, only included in General European Census. Used to determine the type of household that occupants of each dwelling constitute. Also used to establish family groupings for subsequent production of family statistics. Enables exclusion of guests in hotels, patients in hospitals, etc. from population results for revision of electoral boundaries.

Religion	1851-1986	Prior to 1926, only included in General European Census. The only question on the questionnaire which gives respondents the statutory right to object to giving an answer, providing the word 'OBJECT' is written on the questionnaire by the respondent.
Social Security Benefits	1976-86	Asked only of persons aged 15 years and over.
Social Security Income	1981	
Income Group	1926-86	Prior to 1951, only included in General European Census.
Life Insurance Sum Assured	1921	Only included in General European Census.
Address on Census Night	1926-86	Up to 1921, information available from Householder's Schedule. From 1926 onwards, a specific question on address was included in the Personal Schedule.
Usual Residential Address		
- on Census night	1921-86	For 1936 and 1945 censuses, only included in General European Census.
- 1 year ago	1971, 1981	Virtually the only source of information on internal migration of residents within N.Z.
- 5 years ago	1971-86	

Usual Residential Address		
- No. of Years lived at	1976,1986	
Number of Years lived in N.Z.	1851-58, 1901-86	Prior to 1951, only included in General European Census. In 1926, N.Z.-born persons identified from response to question on duration of residence in N.Z.
Country of Birth	1851-1921, 1936-86	Prior to 1951,only included in General European Census.
Father's Country of Birth	1921	Only included in General European Census.
Nationality		
- whether British or foreign	1971-1921	Only included in General European Census. In 1921, aliens required to state actual nationality.
- how acquired	1901-1921	Only included in General European Census. Prior to 1921, only applied to British subjects. In 1921, applied to all Europeans.
Race or Ethnic Origin	1874-1901, 1916-86	Prior to 1921, only included in General European Census. Prior to 1916, only persons of Chinese race required to identify themselves as such. From 1916 onwards, all races were enumerated, producing detailed race statistics.

Descendant of a N.Z. Maori	1976, 1981	Prior to 1974, both Maori Affairs Acts and Electoral Acts defined Maoris as persons who had half or more Maori blood. The 1974 Maori Affairs Act and 1975 Electoral Act amended the definition to enable persons of any degree of Maori descent to be defined as Maori.
Hours Worked	1945-81	In 1945, only included in General European Census. For 1971-81 censuses, asked only of persons aged 15 years and over. For 1981 Census, distinction made between main job, second job and other job.
No. Hours worked last week	1986	Asked only of persons aged 15 years and over. Distinction made between main job and other jobs.
Occupational Status	1891-1986	Prior to 1951, only included in General European Census. For 1971-86, asked only of persons aged 15 years and over.
Occupation	1874-1986	Prior to 1921, only included in General European Census. For 1971-86, asked only of persons aged 15 years and over.
Industry	1851-1986	Prior to 1951, only included in General European Census. For 1971-86, asked only of persons aged 15 years and over.

Name of Employing Organisation	1945-86	Prior to 1951, only included in General European Census. For 1971-86, asked only of persons aged 15 years and over.
Address of Workplace	1971-86	Asked only of persons aged 15 years and over.
Means of Travel to Work	1971-86	Asked only of persons aged 15 years and over. Initially asked in 1971 on 'ad hoc' basis, but repeated since 1971 because of high level of interest shown in results.
Travel time to work	1945,1961,1971	In 1945, only included in General European Census. In 1971, asked only of persons aged 15 years and over.
Unemployment		
- Persons unemployed	1896	Only included in General European Census.
- Sickness	1921-36	Only included in General European Census.
- Duration of	1916,1921,1936	Only included in General European Census.
- Whether registered	1936	Only included in General European Census.
Working time lost through		
- sickness/injury	1926,1936,1945	Only included in General European Census.
- unemployment	1926,1936,1945	Only included in General European Census.
- other causes	1936	Only included in General European Census.
Seeking work in last 4 weeks	1986	

Main activity at workplace	1986	
Service		
- wars in which served	1936-71	In 1936, only included in General European Census. For 1971 Census, asked only of persons aged 15 years and over.
- Forces in which served	1936-66	
- whether receiving war pension	1936	
Intention to postpone retirement	1945	Only included in General European Census. Directed at persons who had voluntarily postponed retirement or who only took up paid work because of war. This and the following 2 questions related only to immediate postwar conditions.
Intended Industry in peacetime	1945	Only included in General European Census.
Intended occupation in peacetime	1945	Only included in General European Census. Asked of persons in Armed Forces or man-powered into work other than intended work in peacetime.
Literacy	1851-1921	Only included in General European Census.

Education		
- Sunday Schooling	1858-1911	Only included in General European Census.
- Establishments being attended	1851-1921, 1966-81	Only included in General European Census.
- Establishments attended in past	1966-81	Asked only of persons aged 15 years and over.
- Duration at each level	1966-71	Asked only of persons aged 15 years and over.
Qualifications		
- educational	1966,1971,1981, 1986	For 1966 Census, asked of all respondents. For other 3 censuses, asked only of persons aged 15 years and over.
- vocational	1971,1981,1986	Asked only of persons aged 15 years and over.
Courses for personal interest	1976	Asked only of persons aged 15 years and over.
Affliction causing inability to work	1851, 1861-1916	Only included in General European Census.
Sugar Diabetes	1971	Whether insulin injection or other methods. Sought at request of Medical Research Council.

Cigarette smoking	1976-81	Asked only of persons aged 15 years and over. Ad hoc question requested for 1976 Census by Medical Research Council of N.Z. and National Health Statistics Centre of the Health Department. Repeated in 1981 at request of National Health Statistics Centre and N.Z. Heart Foundation.

DWELLING QUESTIONNAIRE

Question	*Period Asked*	*Comments*
Address of Dwelling	1926-86	
Name of Nearest Post Office	1916-86	Prior to 1951, only included in General European Census. Asked only of respondents in rural areas (i.e. not for dwellings in a city, borough or town).
Distance to Nearest Post Office	1916-86	Prior to 1951, only included in General European Census. Asked only of residents in rural areas.
Type of Dwelling		Prior to 1926, only included in General European Census. For early censuses, a single large questionnaire was used for each dwelling, but in 1921 a Personal Schedule was introduced for use in nonprivate dwellings, with continued usage of a Householder's schedule for private dwellings. In 1926 and 1936, 3 types of schedule were used, namely the Householder's Dwelling Schedule, Householder's Family Schedule and Personal schedule. Dwelling and personal Schedules were used for the 1945 Census and thereafter, but renamed 'Questionnaires' in 1976 Census.
- Private Dwelling	1926-86	
- Non Private Dwelling	1911-86	

Name of Non-Private Dwelling	1926-36, 1981	For 1926-36, only included in General European Census.
Total No. of Rooms in:		
- Private Dwellings	1861-1981	Only included in General European Census.
- Non-Private Dwellings	1901-76	Only included in General European Census.
No. of Bedrooms	1976,1986	
Material of Outer Walls	1851-1981	Prior to 1951, only included in General European Census. From 1976 onwards, required only for private dwellings.
Material of Roof	1961-81	From 1976 onwards, required only for private dwellings.
Heat Insulation (ceiling & outer walls)	1976-81	Required only for private dwellings.
Year or Period built	1956	Only included in General European Census.
Permanent or Temporary Purpose Built	1926	Only included in General European Census.
No. of Storeys	1981	
Flat, built as such	1926,1951	In 1926, only included in General European Census.
Name of Occupier or Person in Charge	1945-86	In 1945, only included in General European Census.

Occupants, no. of	1916-86	Prior to 1951, only included in General European Census. Prior to 1926, information available from Householder's Schedule without specific enquiry required.
Domestic Servants, no. of	1901	Only included in General European Census.
Absent Persons		
- no. of each sex	1926	Only included in General European Census.
- Details of each absentee	1971-86	Required only for private dwellings.
Tenure	1916-86	Prior to 1926, only included in General European Census. From 1971 onwards, required only for private dwellings.
Rent		Required only for rented or leased dwellings.
- Paid per week	1911-86	Prior to 1951, only included in General European Census. From 1971 onwards, required only for private dwellings.
- From whom rented	1981,1986	
Degree of Furnishing	1936, 1951-81	In 1956, only included in General European Census. From 1971 onwards, required only for private dwellings.

Distance to Nearest - shopping facilities - transport services - public primary school	1945	Only included in General European Census. Asked in an effort to ascertain whether adequate services were being provided to the N.Z. population, and to project future needs.
Cooking, Means of	1945,1956-61	In 1945, only included in General European Census.
Water Heating		In 1945, only included in General European Census.
- type of main supply	1966-86	
- type of secondary supply (if any)	1976-86	
Heating of Dwelling		
- principal means of	1961-71	
- source of energy for	1971,1976,1986	
- appliances used for	1976,1981	
Source of Electrical Supply	1945	Only included in General European Census.
Electric Light	1945	Asked only of Maoris living in the North Island and Chatham Islands.
Piped Water Supply Laid On	1945,1956	In 1945, only included in General European Census.
Type of Power Supply	1945,1961-71	At 1945 Census, asked only of Maoris living in the North Island and Chatham Islands. Asked of all respondents at 1961,1966 and 1971 censuses.

Amenities	1945-81	In 1951, required for flats only. From 1971 onwards, required only for private dwellings. The small range of items included provides some indication of prosperity for social research purposes. Items that could be termed as "luxury items" are inappropriate for inclusion in a census. When census shows that an amenity has reached saturation level, that amenity is discontinued and a different amenity substituted for it.
Vehicles - no. owned or in the care of household members		
- motor cars	1971,1986	Required only for private dwellings.
- motor cycles or scooter	1981	Required only for private dwellings.
- power cycles or bicycles	1981	Required only for private dwellings.
Pleasure boats	1971,1981	Required only for private dwellings.
Caravans	1976-81	Required only for private dwellings.
Holiday residence, Address of	1971,1981	Required only for private dwellings. Ad hoc question in 1971 Census. Repeated in 1981 Census as a result of 4 submissions.
Home vegetable production	1956,1971	In 1971 , required only for private dwellings.

Poultry	1861-91,	Prior to 1945, only included in General European Census.
	1906-71	In 1971 , required only for private dwellings. Dropped after 1971 Census, when it became apparent that the backyard flock was dying out and the trend was towards large commercial flocks.
Bees	1906-16	Only included in General European Census.

Source: *New Zealand Census of Population and Dwellings 1976 Internal Migration,* 1981, *Vol 11,* 161-163 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings 1951 General Report, 1956, *Vol VIII,* 9-10 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings, Census '86 General Information, Series D Report 3, 1988, 153-154.

Appendix 4.2

SPECIMEN COPIES OF THE PERSONAL AND DWELLING QUESTIONNAIRES USED IN THE NEW ZEALAND CENSUS OF POPULATION AND DWELLINGS 1986

epartment of tatistics

CENSUS '86
We all count for New Zealand

New Zealand Census of Population and Dwellings

Tuesday, 4 March 1986

SPECIMEN ONLY

CS/00/03

District No.

Sub District No.

Meshblock No.

Questionnaire No.

For Office Use Only

ersonal Questionnaire

e information you provide will remain confidential to the Department of Statistics

A Personal Questionnaire is legally required to be filled in by or for every man, woman and child (including baby) living in New Zealand at midnight on 4 March 1986.

This completed questionnaire will be seen only by employees of the Department of Statistics who have taken a statutory declaration of secrecy. The information you provide will be used for statistical purposes only and identifiable details about you or your household will not be disclosed to any other government department, organisation or person.

- This Census is taken under the authority of Section 23 (1) of the Statistics Act 1975.

S. KUZMICICH
Government Statistician

ase refer to the guide notes before you fill in this questionnaire

r each question, please either tick box ✓ *or enter number eg:* 4 *or print answer eg:* Surname or family name SMITH

Name

Surname or family name

First or ristian names

Address of where you are on Census night

- *DO NOT give P.O. Box or Rural Delivery numbers*

Street number and name

Suburb or rural locality

City or town or county

1 Where do you usually live?

1 Usually live at the above address

2 Usually live elsewhere in New Zealand at the address below

Street number and name

Suburb or rural locality

City or town or county

3 New Zealand resident with no fixed address

4 Usually live overseas

Please state country

2 How long have you lived at your usual address?

years *(If under 1 year write '0')*

3 Where did you usually live five years ago?

(at the last Census on 24 March 1981)

1 Same as usual address now

2 Lived elsewhere in New Zealand at the address below

Street number and name

Suburb or rural locality

City or town or county

3 Not alive 5 years ago

4 Lived overseas in March 1981

Please state country

4 Sex

6 Male

7 Female

5 Date of birth

day *month* 1 *year*

6 What is your relationship to the occupier?

(the person answering the Dwelling Questionnaire)

01 I am the occupier

02 Husband or wife of occupier

03 Daughter or son of occupier

Other relative *(such as grandchild, mother-in-law)*

Please state

Not a relative *(such as flatmate, boarder, hotel guest)*

Please state

7 What country were you born in?

1 New Zealand
2 Australia
3 England
4 Scotland
5 The Netherlands
6 Western Samoa
7 Cook Islands
8 Other country *(such as Eire, India, Fiji)*

Please state

8 If you were born overseas, how many years have you lived in New Zealand?

years *(If under 1 year write '0')*

9 What is your ethnic origin?

Tick the box or boxes which apply to you

(1) European
(2) New Zealand Maori
(3) Samoan
(4) Cook Island Maori
(5) Niuean
(6) Tongan
(7) Chinese
(8) Indian
(9) Other *(such as Fijian, Tokelauan)*

Please state

10 What is your religion?

01 Anglican
02 Presbyterian
03 Catholic
04 Methodist
05 Baptist
06 No religion
07 Other religion *(such as Ratana, Hindu)*

Please state

08 Object to answering this question

- **Answer the remaining questions if you are aged 15 years or over**
  (that is, if you were born on or before 4 March 1971)
- **If you are under 15 years, answer no more questions, but please sign at the bottom of the back page.**

11 What are your living arrangements?

1 Living with legal husband or wife
2 Living with a partner as a couple *(de facto marriage)*
3 Living alone
4 Living with other persons *(such as parents, flatmates)*

Please state

12 What is your present marital status?

1 Never married
2 Married, first time
3 Remarried
4 Separated
5 Divorced
6 Widowed

13 What Social Welfare payments have you received during the last 12 months?

Tick the box or boxes which apply to you

(1) None
(2) Family Benefit
(3) Family Care
(4) National Superannuation
(5) Domestic Purposes Benefit
(6) Unemployment Benefit
(7) Sickness or Invalid's Benefit
(8) Widow's Benefit
(9) War Pension
(0) Other

Please state

Guide to filling in your Personal Questionnaire

Department of Statistics

CS/00/04

New Zealand Census of Population and Dwellings

Tuesday, 4 March 1986

otes:

- Do not fill in your questionnaire before 4 March 1986.
- Your completed questionnaire will be collected by the Census enumerator along with the other Census questionnaires completed by persons in this dwelling.
- If additional privacy is desired, you may hand your completed questionnaire in a sealed envelope to the occupier or the Census enumerator. Write your name and the questionnaire numbers on the outside of the envelope. The numbers are in the four boxes on the top right hand side of the front page.

ame and address on Census night

- Name and address are asked to make sure that all persons in each dwelling are counted once. Names are also used to allocate household members to a family group.
- Names and addresses are not recorded in the computer processing of Census statistics.
- If a baby has not yet been given a name, write 'baby' instead of first names.

ıe Questions

Where do you usually live?

- Indicate the address you have lived at (or plan to live at) for at least three months.
- If you have more than one New Zealand address, write the one where you spend the most time.
- Primary and secondary school pupils who board away from home and usually return home at the end of each term, should tick box 2 and write the usual address of their parent or guardian.
- If you are a university or tertiary student, trainee, or other person living away from home, give your present address if you have been (or will be) there for three months or more.
- If you usually live overseas, write the name of the country lived in.

How long have you lived at your usual address?

- Exclude temporary absences when estimating how long you have lived at your usual address.
- If your usual address is overseas, write '0' for this question.

Where did you usually live five years ago?

- If you are a New Zealand resident who was overseas at the last Census (on 24 March 1981) then:
  - — if you were absent for less than three months, write your usual address before you left New Zealand.
  - —if you were absent for more than three months, write the name of the overseas country in which you were staying at that date.
- If you are not a resident of New Zealand and you did not live in New Zealand on 24 March 1981, tick box 4 and write the name of the country you lived in at that time.

5 Date of birth

- Write the day, month and year you were born, in the boxes.
  For example, if you were born on 5 December 1942, you would write in the boxes as follows:

- If you cannot recall your exact date of birth, please write the year in which you were born. If this is not known, write your age at your last birthday alongside the question.

6 What is your relationship to the occupier?

- The occupier in a private dwelling is the person chosen to fill in the Dwelling Questionnaire.
- The occupier in a non-private dwelling is usually the proprietor, manager or other person in charge on Census night. He or she need not be living in the dwelling on Census night.

- If you are the de facto husband or wife of the occupier, tick box 02.
- If you are an adopted child or stepchild of the occupier, or a child of the de facto husband or wife of the occupier, tick box 03.
- If you are a flatmate, boarder, foster-child, visitor (not related to the occupier), paying guest in a motel or hotel, patient, inmate or a resident staff member, tick the 'Not a relative' box and write your relationship to the occupier in the space provided.

7 What country were you born in?

- Give the present name of the country in which you were born.

8 Years lived in New Zealand

- Count the total number of years you have lived in New Zealand. Leave out temporary absences.
- If you were born in New Zealand, do **not** answer this question.

9 Ethnic Origin

- If you have more than one origin, tick as many boxes as are necessary to describe your ethnic origin.

10 Religion

- You have a legal right not to answer this question, provided you tick the 'Object to answering' box.
- For children, tick the box or write the religion (if any) according to the way in which they are being brought up.

11 Living arrangements

- If you or your partner are temporarily away from your usual place of residence, answer this question as if you were both living at your usual address. Ignore any temporary absence which is due to a business trip, holiday, hospital stay, or similar reason.

12 Present marital status

- If you are single and have never been married, tick box 1.
- If you are legally married, that is you have signed a marriage certificate, tick box 2 or 3, as appropriate.
- If you are married, but permanently living apart from your husband or wife, tick box 4 whether or not you have a legal separation agreement or order.
- If your marriage has been dissolved, tick box 5.

13 Social Welfare payments received

- Payments received by one family member or behalf of other family members should be shown on the questionnaire of the person th payment is made to. For example, the Famil Benefit is generally paid to the mother. This should be recorded on her questionnaire ON
- If you receive an emergency benefit, tick the box for the benefit you have applied for. For example, if you receive an Emergency Maintenance Allowance, tick the box 'Domes Purposes Benefit'.
- If you are unsure of the type of benefit you receive, tick the 'Other' box and describe th benefit.
- If you receive a pension from another count tick the 'Other' box and write 'Overseas Pension'.

14 Income from all sources

- Income consists of money received from the sources listed in the question, as well as th assessed value of fringe benefits and payments in kind.
- DO NOT include housekeeping allowances, pocket money, board and other payments which you receive from members of your household.
- Include the income you have received since 1 April 1985 and estimate how much you wil receive to 31 March 1986.
- Add the money you receive from all sources and tick the appropriate box.
- If you receive fringe benefits or income in k such as housing, board and goods or servic supplied free by your employer, include the assessed value in your income estimate.
- DO NOT include lump sums you have receiv from Accident Compensation, superannuatic funds, insurances, lottery wins or other sucl sources.
- If you receive income from overseas, conver to New Zealand dollars.
- Include the total amount of National Superannuation payments you have receivec The estimated total amount for the year end 31 March 1986 would be

BEFORE TAX

	Per Year	Per Week
	$	$
Married (each person)	6,118	121
Single	7,478	148

- Unemployment Benefit is taxed before you receive it, unless you have dependent childr Include the before-tax amount in your incom estimate

5 Hours of Voluntary Work

- Voluntary work is unpaid work you do for a community organisation or privately, which will benefit persons outside your household or family.
- Estimate the average number of hours per week and tick the appropriate box.
- DO NOT include time spent doing housework in your own home.
- DO NOT include any time you spend working in a family business without pay or profit.

6 Main work or activity

- Tick the category which applies to your main work or activity.
- If more than one category is applicable to you, choose the one at which you spend the most time.
- If you regularly work without pay or profit in a family business, tick box 07.
- If you are a hospital patient, sickness beneficiary or invalid, tick box 05 and describe your situation.
- If you are on a temporary work scheme, such as PEP or VOTP, tick box 06.

7 Highest school qualification

- DO NOT include partially completed courses.
- If you have obtained a school qualification since leaving school, tick the appropriate box.
- If you have school qualifications from another country, tick the New Zealand equivalents. The following table may assist:

New Zealand School Qualification	Usual Age (Years) Qualification Obtained	Usual Years at Secondary School
School Certificate	15–16	3
6th Form Certificate, Endorsed School Certificate	16–17	4
University Entrance, Matriculation	16–17	4
Higher School Certificate, Higher Leaving Certificate	17–18	5
University Bursary, Scholarship	17–18	5

18 Qualifications obtained since leaving school

- Include all certificates, degrees or diplomas obtained since leaving school.
- If you obtained school qualifications since leaving school, tick the appropriate box in question 17.
- DO NOT include partially completed qualifications.
- If you have qualifications from another country, tick the New Zealand equivalent. If there is no suitable New Zealand equivalent, tick the 'Other qualification' box and write the qualification in full.

19 Did you look for paid work in the last 4 weeks?

- Looking for work means doing at least one of the following over the last four weeks:
  - — applying for a job (such as contacting an employer or answering a newspaper advertisement)
  - — contacting an employment agency
  - — contacting the Department of Labour's Employment Service
  - — placing an advertisement for work in a newspaper
  - — asking friends or relatives for a job.
- Full-time work is 30 hours or more per week.
- Part-time work is less than 30 hours per week.

20 Do you work in a job, business, farm or profession?

- If you work on a temporary work scheme such as PEP, answer 'yes' and complete questions 21 to 27.
- If you are an unpaid worker in a family business, answer 'yes' and complete questions 21 to 27.
- If you are not working now, but are waiting to start a new job, answer 'yes' and complete questions 21 to 27 for your previous job.
- If you are a school leaver who has been hired and is waiting to start a first job, answer 'yes' and complete questions 21 to 27 for your new job.
- If you do not work in a job, business, farm or profession, answer 'no', and then sign the questionnaire.

21 Employment status

- If you have two or more jobs answer this question for your main job (the one at which you spend the most time).
- If you are working on a commission only basis, tick box 1.
- If you are a partner in a business, tick box 2 or 3, as appropriate.
- If you regularly work without pay or profit in a family business, tick box 4.

22 Hours worked last week

- If you have two or more jobs, state the hours for your main job first. State the hours for the other job or jobs in the second box.
- If you have only one job, write '0' in the second box.
- Include overtime hours worked last week.
- If on holiday, sick or temporarily absent from work last week, give usual hours worked.
- Do not include unpaid household or voluntary work.

23 Present occupation

- If you have more than one job, business or professional practice, write down what you do in your main job.
- If you are a trainee or apprentice, include that in the description.

Examples of occupations, tasks and duties are:

Occupation	Task or Duties
bookkeeper	preparing company accounts
civil engineer	designing and supervising bridge construction
clerical supervisor	checking the work of insurance clerks
council labourer	weeding gardens, keeping parks tidy
factory supervisor	controlling a frozen vegetable process line
interior design consultant	advising customers on colour schemes
key punch operator	keying data into a computer
knitwear machinist	making cardigans
laboratory technician	carrying out diagnostic tests on blood
motor vehicle mechanic	repairing cars
process worker	assembling television sets
shop assistant	making and selling takeaway foods
truck driver	driving petrol tanker
working proprietor	selling books and stationery

24 Who do you work for?

- If your employer does not have a trading nam give his or her name.
- If you work in your own business, give your trading or business name. If you do not have trading or business name, write 'self'.

25 Where do you work?

- If you do not know the street name and number, give the name of the building.
- If you have more than one job, write the address of your place of work for your main jol
- If you have no fixed workplace (for example, you are a milk vendor, driver, sales representative), give the address of the depot headquarters or reporting point you operate from. If you have no fixed reporting or assembly point, but travel from your home to various work locations, write 'no fixed address'.
- If you work at home or in the same building you live (for example, if you are a farmer, dressmaker or dairy proprietor) tick the 'Work at home' box.
- If you are working on a ship, write the name the port at which the ship is berthed on Census night. If the ship is not berthed on Census night, write the name of the next por of call.

26 The main activity at your place of work

- Main activity refers to your employer's predominant business, industry or service.

Examples of employers' activities are:

beef farming
brewing
cartage contracting
commercial printing
health food retailing
leather bag manufacturing
market gardening
meat freezing works
metal furniture manufacturir
motor vehicle upholstering
Post Office mail service
private school
property valuation
road construction
town milk supply
trading bank

27 Main means of travel to work

- If you have more than one means of travel to work, tick the box for the means of transport you usually use to travel the greatest distanc
- If you travel to work in different ways during the week, or if you have no fixed place of work, tick the box that describes how you travel most often.
- If you are a farmer living and working on a farm, tick box 0.
- If your place of work is in the same building you live, tick box 0.

14 What will be your income before tax for the year ending 31 March 1986?

Include income from all sources

- Wages, salary
- Social Welfare payments *(including National Superannuation)*
- Family Care, Family Benefit
- Interest, dividends, rent, commission
- Fringe benefits or income in kind
- Business or farming income *(less expenses)*
- Accident Compensation weekly payments
- Bursary, Scholarship
- Superannuation

01 Nil or Loss

02 $1000 or less per year (Less than $19 per week)

03 $1,001–$2,500 per year ($19 and less than $48 per week)

04 $2,501–$5,000 per year ($48 and less than $96 per week)

05 $5,001–$7,500 per year ($96 and less than $144 per week)

06 $7,501–$10,000 per year ($144 and less than $192 per week)

07 $10,001–$12,500 per year ($192 and less than $240 per week)

08 $12,501–$15,000 per year ($240 and less than $288 per week)

09 $15,001–$17,500 per year ($288 and less than $337 per week)

10 $17,501–$20,000 per year ($337 and less than $385 per week)

11 $20,001–$25,000 per year ($385 and less than $481 per week)

12 $25,001–$30,000 per year ($481 and less than $577 per week)

13 $30,001–$35,000 per year ($577 and less than $673 per week)

14 $35,001–$40,000 per year ($673 and less than $769 per week)

15 $40,001–$50,000 per year ($769 and less than $962 per week)

16 $50,001 and over per year ($962 and over per week)

15 How many hours of voluntary work do you do on a regular weekly basis?

- *For example, Meals on Wheels, sports administration, marriage counselling, Te Kohanga Reo*

1 Nil hours

2 1–4 hours per week

3 5–9 hours per week

4 10–14 hours per week

5 15 or more hours per week

16 What is your main work or activity?

01 Home duties—looking after children

02 Home duties—not looking after children

03 Full-time student

04 Retired

05 Unemployed

06 Paid job, business, farming or profession

07 Unpaid work in a family business

08 Other *(such as hospital patient)*

Please state

17 What is your highest school qualification?

1 No school qualification

2 School Certificate, 1 or 2 passes

3 School Certificate, 3 or more passes

4 6th Form Certificate, Endorsed School Certificate

5 University Entrance, Matriculation

6 Higher School Certificate, Higher Leaving Certificate

7 University Bursary, Scholarship

8 Other school qualification

Please state

18 What qualifications have you obtained since leaving school?

Tick the box or boxes which apply to you

01 Still at school

02 No qualification since leaving school

03 Trade Certificate, Advanced Trade Certificate

04 Nursing Certificate or Diploma

05 Teachers Certificate or Diploma

06 Technicians Certificate

07 New Zealand Certificate or Diploma *(awarded by the TCA or AAVA)*

08 University Certificate or Diploma below Bachelor level

09 Bachelors Degree

10 Postgraduate Degree, Certificate or Diploma

11 Other qualification

Please state

19 Did you look for paid work in the last 4 weeks?

1 ☐ Yes — looked for full-time work

2 ☐ Yes — looked for part-time work

3 ☐ No — did not look for work

20 Do you work in a job, business, farm or profession?

6 ☐ Yes — working ► **Answer all questions 21 to 27**

7 ☐ No ► **Answer no more questions. Please sign box at bottom of page**

21 In your work, are you . . .

1 ☐ Working for wages or salary

2 ☐ Self-employed and not employing others

3 ☐ Employer of others in own business

4 ☐ Unpaid worker in a family business

22 How many hours did you work last week?

- *If on holiday, sick or absent for other reasons, state usual hours*

☐ Hours worked last week in main job

and

☐ Hours worked last week in other jobs
(If nil hours write '0')

23 What is your present occupation?

- *For example, builder's labourer, maintenance fitter, sheep farmer, primary teacher, general office clerk.*

In your work what are your main tasks or duties?

24 Who do you work for?

- *Please state name of business, firm, government department or other organisation*

25 Where do you work?

- *If street address is not known, give building name or shopping centre*

Street number and name

Suburb or rural locality

City or town or county

OR ☐ Work at home

26 What is the main activity at your place of work?

- **State fully**
  For example, public health nursing, video hire, shirt manufacturing, sheep farming.

27 What is your main means of travel to work?

Tick one box only

1 ☐ Public bus

2 ☐ Train

3 ☐ Drive a private car, truck or van

4 ☐ Drive a company car, truck or van

5 ☐ Passenger in car, truck, van or company bus

6 ☐ Bicycle

7 ☐ Motor cycle, power cycle

8 ☐ Walk

9 ☐ Other means

0 ☐ Work at home

Signature: I declare that the information I have given is true and complete as far as I know:

Sign here

partment of
tatistics

CENSUS '86
We all count for New Zealand

New Zealand Census of Population and Dwellings

Tuesday, 4 March 1986

CS/00/01

SPECIMEN ONLY

District No.

Sub District No.

Meshblock No.

Questionnaire No.

welling Questionnaire

structions

A Dwelling Questionnaire is legally required to be filled in for every dwelling which is occupied on Census night.

If this is a private dwelling, the person who fills in the Dwelling Questionnaire should be either:
— a person who owns this dwelling,
— or, if a rental dwelling, a person in whose name the dwelling is rented,
— or any other responsible person.
This person, who must be living in the dwelling on Census night, is called the **occupier.**

If this is not a private dwelling, the person who fills in the Dwelling Questionnaire should be either the proprietor, superintendent or other person in charge on Census night. This person is called the **occupier.**

The **occupier** is also required to ensure that a Personal Questionnaire is filled in for every person (including baby) present in the dwelling on Census night.

Persons arriving at or returning to this dwelling between midnight on 4 March and noon on 5 March must fill in a Personal Questionnaire at this dwelling, unless one has been filled in at another dwelling.

The information you provide will remain confidential to the Department of Statistics.

- This completed questionnaire will be seen only by employees of the Department of Statistics who have taken a statutory declaration of secrecy. The information you provide will be used for statistical purposes only and identifiable details about you or your household or dwelling will not be disclosed to any other government department, organisation or person.
- This Census is taken under the authority of Section 23 (1) of the Statistics Act 1975.

S. KUZMICICH
Government Statistician

se refer to the guide notes before you fill in this questionnaire

r each question, please either tick box ✓ *or enter number eg:* 4 *or print answer eg:* Street number and name 4 HIGH ST.

Name of occupier

Surname or family name

First or Christian names

Address of this dwelling

DO NOT give P.O. Box or Rural Delivery numbers

Street number and name

Suburb or rural locality

City or town or county

If an institution, hotel, motel, hospital, school hostel, camp, boarding house or ship

lease give name

If this dwelling is in a rural locality, rural township, or county . . . *please state*

Name of nearest Post Office

Distance to nearest st Office by usual route | km or | miles

How many persons are present in this dwelling on the night of 4 March 1986? *(include babies)*

persons

2 Is this dwelling

either

(a) a private dwelling?

01 Separate house

02 Two flats or houses joined together

03 Three or more flats or houses joined together

04 Flat or house attached to a business or shop

05 Bach, crib or hut *(not in a work camp)*

06 Caravan, cabin or tent in a motor camp

07 Other

Please state

or

(b) not a private dwelling?

11 Hotel, motel or guest house

12 Boarding house or rooming house

Other *(such as hospital, construction camp)*

Please state

- The following questions are for ***private dwellings*** only.
- If not a private dwelling — please sign at bottom of this page

3 How many bedrooms are there in this dwelling?

bedrooms

4 Is this dwelling

01 Owned with a mortgage

02 Owned without a mortgage

03 Provided rent-free

04 Rented or leased

5 If you rent or lease this dwelling:

(a) How much is the WEEKLY rent?

$ dollars . cents per week

(b) Who do you rent or lease from?

1 Private organisation, person or real estate agency

2 Housing Corporation

3 Other government department *(including hospital or education board)*

4 Local authority *(including council, electric power board or harbour board)*

(c) Is this dwelling rented or leased on a furnished basis?

5 No—unfurnished

6 Yes—furnished

(d) Do you rent or lease from your employer?

7 No—not rented from employer

8 Yes—rented from employer

6 What do you use to heat this dwelling?

Tick one or more boxes

1 Electricity

2 Gas

3 Wood

4 Coke or coal

5 Oil *(including kerosene)*

6 Other

Please state

7 No means of heating

7 What type of hot water supply do you have i this dwelling?

Tick one or more boxes

1 Electric

2 Gas

3 Other *(such as wood, solar)*

4 No hot water supply

8 How many motor vehicles available for privat use do persons in this dwelling have in their care on Census night?

Do not include: motor bikes, scooters, tractors

0 None

1 One

2 Two

3 Three

4 Four

5 Five or more

9 Persons away on Census night (4 March 1986)

List below:

- Those persons who are temporarily away (for less than 3 months) on Census night, but who usually live at this dwelling, such as persons on business, on holiday, or in hospital.
- Children at boarding school.
- Mother and baby at a maternity hospital.

Do NOT list below:

- Those persons who are away for longer than 3 months, but who usually live in this dwelling, such as long-term hospital patients, armed forces personnel overseas.
- University or tertiary students who live away from this dwellir for most of the year.

	ABSENT PERSON 1	ABSENT PERSON 2	ABSENT PERSON 3	ABSENT PERSON 4
Surname or family name				
First or Christian names				
Sex				
Age (*in years*)				
Marital status (*such as married*)				
Relationship to you (*such as son*)				
Address or location on 4 March 1986				

FOR OFFICE USE ONLY

Signature: I declare that the information I have given is true and complete as far as I know:

Sign here

;uide to filling
ı your
)welling Questionnaire

Department of Statistics

New Zealand Census of Population and Dwellings

Tuesday, 4 March 1986

ɛlecting an Occupier

- Where a private dwelling is owned or rented jointly by two or more persons, choose **any** one who is living in the dwelling on Census night to be the **occupier**.
- The occupier in a **non-private** dwelling is usually the proprietor, manager or other person in charge on Census night. He or she need not be living in the dwelling on Census night.

hat is a dwelling?

- The term dwelling means any building or structure, whether permanent or temporary, which is being wholly or partly used as living quarters for a household.
- A **private dwelling** is one which is lived in by a private household.
- A **private household** consists of one or more persons living together, who usually have one or more meals together daily. Boarders and temporary visitors are also household members.
- Caravans, cabins or huts in motor camps which are occupied on Census night are private dwellings if the occupants have been, or will be, living in them for three months or more.
- Each self-contained flat or house within a retirement village or complex is a private dwelling.
- Managers, proprietors, persons-in-charge or staff of institutions, camps and other non-private households are considered to be living in a separate private dwelling, if they have self-contained living quarters within the institution or camp.
- A **non-private dwelling** is an accommodation unit catering for a number of generally unrelated persons, comprising a non-private household. Examples are hotels, motels, boarding houses, camps, ships or institutions such as a hospital or prison.

ıe Questions

How many persons are present?

- Count all persons, including children and babies, who spent the night of Tuesday 4 March in this dwelling and enter the number in the space provided.
  For example, if four persons were present on Census night, complete the box as follows:

  [4] persons

- Include persons who are away on Census night, but who return before noon on Wednesday 5 March and have not completed a questionnaire somewhere else, e.g. shiftworkers.
- Include visitors who arrive at this dwelling before noon on Wednesday 5 March and have not completed a questionnaire elsewhere.

Type of dwelling

- Answer either question (a) or (b). Tick only one box.

(a) *Private dwelling*

- A 'separate house' is a house which stands apart from any other house, flat or business premises.
- 'Joined together' means connected or joined by a common wall, garage, carport or shed to another dwelling.
- If you live in a town house, apartment, ownership unit or villa, which is joined to another dwelling, tick box 02 or 03, as appropriate.

(b) *Non-private dwelling*

- If you live in a hostel, convent, prison, youth camp, recreational camp, commune, ship or military barracks, tick the 'Other' box and state the type of dwelling.

3 Bedrooms

- Count all bedrooms, including spare bedrooms.
- If this dwelling is a bed-sitter, cabin or caravan which does not have a separate bedroom, count this as one bedroom.
- Include caravans and sleepouts which are used as extra bedrooms.

4 Is this dwelling owned or rented?

- This question only applies to the dwelling you occupy and not to the land on which it is situated.
- If this dwelling is being bought under hire-purchase (such as a caravan), tick box 01.
- Owner-occupants of ownership flats purchased under unit title, stratum title, composite leasehold title, licence to occupy or similar, should tick box 01 or 02, as appropriate.

5 If you rent or lease this dwelling

- Do not include mortgage repayments as weekly rent.
- If you do not rent on a weekly basis, give the equivalent weekly rate.
- If this dwelling is a hired caravan in a motor camp, include site fees in the total rent paid.
- If the rent for this dwelling includes rent or lease payments for a farm, business or shop, estimate the weekly rent for your living quarters only.
- If this dwelling is rented on a partly furnished basis, tick box 6.

6 Heating this dwelling

- If you use more than one type of heating, tic the boxes which apply.

7 Type of hot water supply

- If you have more than one method of heating the water supply, tick the boxes which apply

8 Motor vehicles

- *Include:*
  - cars, station-wagons, vans, trucks, utility vehicles, four-wheel drive vehicles and oth vehicles used on public roads;
  - privately-owned and business vehicles, if they are available for the private use of persons in the dwelling;
  - vehicles which are hired, borrowed or leased or supplied by an employer;
  - vehicles which are temporarily under repa
- *Do not include:*
  - motor bikes, scooters, or farm machinery;
  - vehicles used exclusively for business.

Appendix 5.1

TESTING PROGRAMME TIMETABLE AND QUESTIONNAIRE TOPICS FOR THE NEW ZEALAND CENSUS OF POPULATION AND DWELLINGS 1986

Date	Kind of Test	Topics Tested
September 1982	OMR	Pre-coded answer options.
August 1983	Pretest	Ethnic Origin, Dwelling Type, Income, Marital Status, Age, Unpaid Household Work.
February 1984	Pilot Test	Full range of questions.
May 1984	Pretest	Employment Status and Main Activity.
June 1984	Pretest	Marital Status, Employment Status, Relationship to Occupier, Ethnic Origin, Voluntary Work.
June 1984	Skirmish	Name and Distance from Nearest Post Office.
July 1984	Pretest	Income, Employment Status, Tenure and Rent.
August 1984	Skirmish	Occupation and Industry.
November 1984	Pilot Test	Full Range of questions.

Source: New Zealand Census of Population and Dwellings Census '86 Questionnaire Content and Submissions, 1985,11.

Appendix 6.1

PUBLISHED 1970 U.S. CENSUS COVERAGE EVALUATION STUDIES

Population Coverage	Purpose
A. Demographic Analysis 1970	To estimate the net coverage for the total U.S. population. This is the major undercount study for the Census.
B. Medicare Record Check	To measure the omission and overenumeration of persons 65 years of age and over. Medicare enrolment files were used.
C. Birth Registration Test	To measure the completeness of birth registration for children under 5 years of age.
D. CPS-Census Match (1)	To identify, primarily, "within-household" omissions in order to discover why such omissions occurred. The March 1970 CPS sample units were matched with census records for this experiment.
E. Special Procedures to Improve 1970 Census Coverage	To measure the effect of various census procedures on coverage improvement.

Housing Coverage	Purpose
A. CPS-Census Match (1)	To provide estimates of housing undercount for the total U.S. Again, 1970 CPS sample units were used.
B. Mail Area Coverage	To provide estimates of omission, overenumeration, and the net coverage error for housing units in mail census areas.
C. Definitional Housing Unit Errors family	To measure the extent of definitional error in the census housing count (e.g., two households at an address which the census lists as a one-home).
D. National Vacancy Check	A large-scale post-census survey conducted to determine the misclassification rate of housing. A sample of the units classified as vacant were revisited to determine whether any of them had been misclassified.

Other Related Studies	Purpose
A. Geographic Coding Study	To measure the amount of residential address miscoding to geography (e.g., wrong block or tract).

B. Mail Extension Test	To determine the feasibility of conducting a mail census in rural areas.

(1) CPS stands for Current Population Survey, which involves a sample of 66,000 households each month, and has been conducted since 1940.

Source: US Department of Commerce *Census '80: Continuing the Factfinder Tradition* , 1980, 265-266 (Bureau of the Census).

Appendix 6.2

SPECIAL PROCEDURES EMPLOYED TO IMPROVE THE 1980 US CENSUS COVERAGE AND POST-CENSAL EVALUATIONS

Special Procedures

(a) *Field procedures*

The US Census Bureau has posted out the bulk of the questionnaires for the 1970 and 1980 Censuses. If respondents received their questionnaires through the mail, they were required to post back the completed questionnaires to specified collection centres. The remaining areas, which were less intensely populated, or were traditionally difficult to enumerate, were covered by door-to-door visits by enumerators.

Commercial mailing lists are available for the more urban areas of the United States. These were purchased by the Census Bureau, updated with postal checks and the precanvass operation, and assigned geographic codes by computer. The areas covered by these address lists are referred to as Tape Address Register (TAR). In the remaining areas, the Census Bureau compiled the address list by first having the areas canvassed by census enumerators, and then updating the list by further checks conducted by the Post Office.

Prelist Recanvass Operation

The Prelist Recanvass Operation was conducted to add and enumerate housing units that had been missed in previous census operations, to delete listed units that were located outside the boundaries of the recanvassed enumeration district, to reinstate deleted listings where appropriate,

and to eliminate duplicate listings. A quality check on every precanvass enumerator's work was conducted by deliberately suppressing a sample of housing units from the Precanvass Address Register. If an insufficient number of suppressed units were not reinstated by an enumerator, then the enumerator's area was recanvassed.

Coverage Improvement Evaluation Sample (CIE)

The CIE was designed as a 2-stage stratified cluster sample. Prior to sampling, the 409 district offices (DOs) that were set up for the census were separated into 6 strata according to the following criteria:

Stratum I :	Centralised DOs in cities with 1,000,000 or more population.
Stratum II :	The balance of centralized DOs.
Stratum III :	Decentralised DOs without Prelist Recanvass.
Stratum IV :	Decentralised DOs with Prelist Recanvass-Urban.
Stratum V :	Decentralised DOs with Prelist Recanvass-Rural.
Stratum VI :	Conventional[1] plus two-procedure DOs (i.e. DOs where both conventional and decentralised operations were used for operations).

[1] DOs in which the questionnaires were delivered and collected by enumerators.

The first stage of sampling consisted of a randomly selecting samples of district offices in each of the 6 strata, and supplementing these samples several times. At the second stage of sampling, a 15% sample of enumeration districts was systematically selected with equal probability from within the sample district offices.

Misclassified Occupied Operation

All Pretests for the 1980 US Census included attempts to determine efficient methods of detecting misclassification of occupied housing units based on actual enumeration. Extensive follow-ups of vacant or non-existent units were conducted in several of the Pretests. This procedure was evaluated using a Coverage Improvement Evaluation Sample, which used the 6 strata listed above to produce an estimate of the number of units whose occupancy status had been misclassified.

H4-Edit Procedures

The editing of question H4 on the census questionnaires, which asked respondents the number of units in the structure in which they resided. The respondents' answer was then compared to the number of units in the census records, in an effort to detect small multi-unit structures with missing housing units. The effectiveness of the H4 Edit was evaluated by using the Census Allocation Program Evaluation Sample to estimate the number of housing units added by the H4 Edit.

(b) ***Special Communication Efforts***

These placed special emphasis on racial and ethnic populations. In addition to standard publicity efforts directed through the media, they included:

(1) Hirage of minority Public Information personnel to implement the programme.

(2) Development of T.V. and radio spots specifically designed to reach minority and ethnic audiences.

(3) Development of printed literature of all types for distribution in minority and ethnic communities.

(4) Development of programmes for working with minority disk jockeys to help promote the census.

(5) Obtaining testimonials from prominent leaders in minority or ethnic communities.

(6) Arranging for minority advertising agencies to play a vital role in the national advertising campaign for the census.

(7) Development of special public service announcements on the hiring of minority or ethnic persons as temporary census employees.

(c) ***Local Review***

In an attempt to improve the quality of data received from field operations, pre-enumeration counts of housing units by small areas were made available for intensive review to local authorities. Preliminary population and housing unit counts for the smallest areas for which summarisations could feasibly be made were made available in local field offices. Local authorities were thus provided with the opportunity to review counts both before and after the actual field work on the census and to notify the Census Bureau of discrepancies in sufficient time to affect the completeness of the count. The availability of small area data facilitated pinpointing the areas of possible discrepancies. However, detailed supportive evidence was required for any application for amendment of a count.

(d) ***Non-household Sources (NHHS) Programme***

The NHHS programme was instigated in an effort to reduce the differential undercoverage of minority populations. It was conducted only in certain areas with large minority populations and was designed to enumerate persons who had been missed in households for which a census questionnaire had been received. In other words, the aim

Lists of names and addresses were acquired from the Department of Motor Vehicles for 43 states and the District of Columbia, from the U.S. Immigration and Naturalisation Service and from the 1979 New York City Public Assistance file. The address lists were matched to the Master Address Register and amended as necessary. Each viable case was then matched where possible to a census questionnaire. Follow-up interviews were conducted to ascertain whether the NNHS person should have been listed on the questionnaire, and whether any other persons in the household were missed. Finally, all missed persons enumerated during the follow-up were added to the census questionnaires and the address register was updated.

Post Censal Evaluations

Post Enumeration Survey (PES)

The PES was based on a case-by-case matching. Because the cases were individually matched, the PES was adaptable to any demographic group and to any level of geography. The procedure is to conduct a household survey soon after the census, and to then search the census records for each person counted in the PES. The proportion of the PES cases not found is taken as the estimate of the proportion of the population missed by the census; in other words, the *gross undercoverage rate*. For the 1980 Census, two months of the Current Population Survey were matched to the census. In addition, a sample was selected from the census to estimate the rate of erroneous enumerations (i.e. nonexistent persons, persons counted in the wrong place, and multiple enumerations). The difference in the gross undercoverage rate and the rate of erroneous enumerations is the *net undercoverage rate*.

A second part of the Post Enumeration Programme was designed to estimate gross overcoverage in the census, caused by duplicate and erroneous enumerations and geocoding errors. A second sample was drawn from the census and the sampled households were interviewed to determine whether residents were actually living at that address at the time of the census, the number of persons listed in census who were born after Census Day or died before Census Day and to uncover fabricated persons or households. In addition, at the time of the interview, the address of the unit was re-geocoded, and later compared to the one assigned to the housing unit by the census.

Source: Sledge G., Harahush T. and O'Brien R. (1984) "Misclassified/Occupied and H-4 Edit Coverage Operations" *1984 Section on Survey Research Methods* , 525-536 (American Statistical Association).

Appendix 6.3

ESTIMATED ERRORS OF CLOSURE FOR NEW ZEALAND CENSUSES OF POPULATION AND DWELLINGS 1956-1986

Census Date	Total New Zealand Population		Difference between Enumerated and Expected Population[1]	
	Enumerated	*Expected*	*Number*	*Percent*[2]
17 April 1956	2,174,062	2,177,564	- 3,502	-0.16
18 April 1961	2,414,984	2,415,435	- 451	-0.02
22 March 1966	2,676,919	2,690,904	- 13,985	-0.52
23 March 1971	2,862,631	2,869,885 [3]	- 7,254	-0.25
23 March 1976	3,129,383	3,148,774	- 19,391	-0.61
24 March 1981	3,175,737	3,163,894	12,166	+0.38
4 March 1986	3,307,086	3,315,600	8,516	+0.26
23 March 1971	2,862,631	-		
24 March 1971	3,175,737	3,182,962 [4]	- 7,225	-0.23

(1) A minus sign (-) indicates an apparent under-enumeration of population.

(2) Taken as a percentage of the expected population.

(3) Derived using the previous census as a base.

(4) Derived using the 1971 Census as a base for 1981.

Sources: *New Zealand Census of Population and Dwellings, 1981*, 1985, *Volume 12 (General Report)*, 156 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings Census'86 Usually Resident Population, 1987, *Series B Report 25*, 13 (Department of Statistics, Wellington).

Population Division, Department of Statistics, Christchurch.

Appendix 6.4

UNDERENUMERATION OF BABIES IN THE NEW ZEALAND 1976 CENSUS OF POPULATION AND DWELLINGS

A brief report on the study, conducted by the Department of Statistics, of the maximum possible underenumeration of babies aged 0-52 days.

Only nuptial births were included in the study (a nuptial birth being defined as a birth occuring after the date of marriage), because of the high probability that a large proportion of ex-nuptial births would not, at Census date, have been at the residential address of their (natural) mothers as stated on the Birth Registration Form. Each address supplied on the Birth Registration Form was used to check whether a Personal Questionnaire had been completed at that address for the baby concerned. If no such questionnaire was present, but the available questionnaires indicated that the baby's family was still residing at the address, then it was noted whether the baby's personal details were listed in the Absent Persons question of the Dwelling Questionnaire. A search was also made of the Census questionnaires from the address where the baby had been born. (This address was usually that of a maternity hospital, maternity home or maternity annex.) If the search was unsuccessful, the Birth Registration Form was used to search Census questionnaires from major hospitals in New Zealand and Death Registration Forms for the period 1 February-23 March 1976. Because of practical considerations, no attempt was made to locate the Personal Questionnaires of any babies which were permanently or temporarily staying at a private residential address on Census night that was not their mother's Birth Registration Form address. The study yielded a maximum possible undercount rate of 3.38% for the North Island, 2.04% for the South Island and 3.02% for the whole of

New Zealand. Omission rates for babies in the Absent Persons question of the Dwelling Questionnaire for these regions were 11.75%, 11.41% and 11.66%, respectively.

Source: *New Zealand Census of Population and Dwellings 1976 Internal Migration*, 1981, *Vol 11*, 158-159 (Department Of Statistics, Wellington).

Appendix 7.1

NEW ZEALAND CENSUS DATA PREPARATION AND CODING PROCEDURES

For the New Zealand Census, the Sub-enumerators in the Sub-districts check the questionnaires for any obvious errors and omissions when the questionnaires are collected from each household. The questionnaires are then submitted to the Enumerator, who gives the questionnaires a further check before forwarding them to the Census Division of the Department of Statistics.

On receipt of the questionnaires by the Census Division, a series of checks are performed on the questionnaires and any queries referred back to the relevant Enumerator. Prior to 1986, once the questionnaire checks were completed, a 10% systematic sample of the questionnaires was processed, and the results published as bulletins of sample estimates. In 1976, the sample was selected on the basis of meshblocks (the smallest statistical units identified for the processing of geographic data), and in 1981 the sample was selected on the basis of private households.

However, in 1986, by capitalizing on the benefits of automated data processing, the aim was to conduct a 100% sample (i.e. a full sample) of all those questions which were answered by ticking or entering a particular number in the answer boxes. Such questions constituted 90% of the census questionnaire, and it was hoped that these responses would be processed within six months of the Census date, and published without editing, to provide a range of small area data earlier than previously. The responses to the remaining 10% of the questions would be entered using CAC (Computer Assisted Coding).

In 1981, the coding of the questionnaires was undertaken in several phases, and was designed to enable related questions, such as questions on the labour force, to be coded in the same operation. Priority was given to the questionnaires selected in the 10% systematic sample of private households, and the associated Dwelling and Personal Questionnaires of non-private dwellings.

Coders entered the coded 1981 data onto specially designed coding sheets which were then read by an OCR (Optical Character Recognition) machine. The OCR process electronically reads written data onto a disc. To make the coded data easier to read, the coding forms were made from dense white paper, and because the characters had to conform to specified shapes and proportions and have a minimum variation in the density of the pencil impressions, the coded data was entered carefully onto the forms, and all characters carefully positioned on the document in coding boxes which had been printed using OCR non-readable ink. When characters were rejected as being unreadable by the OCR reader, the cases were entered on the disc file and flagged, and were subsequently corrected by an operator. As coding errors increase the expense of the Census, the performance of the coders was constantly monitored, and further training was given when required. After the disc file was cleared of all reading errors, the data was transferred to tape to be used as input for the computer edit programme.

The editing and imputation phase is where the errors on records are detected, identified and corrected. For the 1981 Census, after the tape containing the sample data was returned from the OCR Centre, it was sorted and reformatted into separate Dwelling and Personal records for each household, then given an initial edit and stored on tape.

In the initial edit, each record was checked to ensure it belonged in the batch and checked for errors, and then a check on households was performed. The system used for editing,

TIDE (The Identification of Errors), was developed by the Department of Statistics, and was based on principles of editing developed by Fellegi and Holt of Statistics Canada. TIDE incorporated over 750 edit checks!

In the Correction Edit phase, only those households containing records that required correction were accessed. The data from the Initial Edit phase was loaded out to an indexed sequential disc file, and when a correction was made to a record the whole household was re-edited to ensure that no further errors had been introduced. The corrections were input to the system through the OCR capture process, an Inforex key punching facility, or through three terminals situated in the coding section. The correction runs were scheduled, and ran in batch mode until all detected errors were corrected.

After each batch was cleared and the control total of the number of records expected matched the number on file, the final edit phase was run. In the final edit phase, all records were checked to ensure that no definite errors had been overlooked during the correction edit phase. If errors were found, the batch was retained for correction. Otherwise, it was transferred to tape for subsequent table production runs.

A late decision to obtain a dedicated machine for edit processing and the resultant tight conversion time frame meant that computer imputation procedures for the 1981 Census were limited to the simple derivation of several fields, such as the age of a respondent derived from her date of birth. For the 1986 Census, computer imputation procedures were used more extensively. For instance, if a respondent's sex was not supplied, and it was not apparent from the supplied christian names, then it was automatically imputed using a probability-based allocation of Male and Female classifications. If neither age nor date of birth were supplied, then the distribution of the family relationship was examined, and a value imputed.

For censuses prior to 1986, additional (temporary) staff have been employed by the Department of Statistics for the coding and transcription of questionnaires. For the 1986 Census, the Department elected to eliminate the transcription stage by entering the data directly off the questionnaires. The questionnaire design was altered to ensure that the maximum amount of information could be entered directly off the questionnaires, the data being captured by key-to-disc, and CAC used for the remaining questions. The editing stage was combined with the CAC, and it was hoped that CAC would eliminate a lot of errors, improve the consistency of the data and substantially reduce the costs and time for the operation. For recent censuses, when coding occupations, a list of occupations was manually referenced and the selected code manually entered on the respondent's questionnaire coding sheet. Research conducted by the Department indicated that approximately 40% of coding time was spent transcribing information form one form to another. By using CAC for the 1986 Census data, the Department of Statistics required only half the number of staff employed to process the 1981 Census data.

Source: In-house report (1982), Department of Statistics.

GLOSSARY OF SAMPLING AND CENSUS TERMS

ad hoc questions
See **questions, ad hoc**

address reference files
See **files, address reference**

Administrative County (New Zealand)
See **County, Administrative**

age-heaping
The phenomena of respondents, particularly older respondents, reporting their ages as ending in particular digits, usually 0 or 5.

analysis, demographic.
A method of updating the previous census count of the total population by taking into account the births, deaths and external migration registered during the intercensal period, i.e.

$$P^* = P + B - D + I - E,$$

where P^* = Estimated population count
P = Previous census count
B = Number of births recorded in inter-censal period
D = Number of deaths recorded in inter-censal period
I = Number of immigrants leaving the country in inter-censal period
E = Number of emigrants entering the country in inter-censal period

Area, Statistical (New Zealand)
A statistically defined area and has no administrative or legal basis. Statistical Areas were introduced in New Zealand in 1961, and the whole of New Zealand is covered by the 13 Statistical Areas, many of which conform closely to the old provincial districts.

Areas, Rural (New Zealand)
All areas not specifically designated as main, secondary or minor urban areas. Extra-county islands are included, but shipboard population is excluded.

Areas, Urban (New Zealand)
As from the 1981 Census, the classification of Urban Areas consists of **Main Urban Areas, Secondary Urban Areas and Minor Urban Areas**. Wellington and Auckland are each divided into four Main Urban Areas, but all other **Urban Areas** are broadly designed to encompass completely built-up urbanised centres or towns and are usually centred on territorial local authorities. Urban Areas boundaries are drawn to include a centralised population and associated hinterland, with an element of geographical containment and to allow room for growth of the urban population within the boundaries to facilitate comparability over a 15-20 year period.

Main Urban Areas. The present boundaries for Main Urban Areas were fixed at the 1971 Census and remain unchanged except for minor adjustments. Prior to 1981, one of the criteria for classifying Main Urban Areas was a population of 20,000 or more, centred on a major city or borough. This was increased to 30,000 at the 1981 Census, reducing the number of Urban Areas from 24 to 23.

Secondary Urban Areas were introduced at the 1981 Census and are based on similar criteria to Main Urban Areas, except that their population are between 10,000 and 29,999. Fourteen Secondary Urban Areas were created, including Masterton, which was formerly a Main Urban Area.

Minor Urban Areas
Defined to be all other population centres with populations of 1,000 or over which are not already classified as urban (that is, which do not fall within a Main or Secondary Urban Area).

Area, Urban
Boundaries drawn to include a centralised population and associated hinterland, with an element of geographical containment, and to allow room for growth of the urban population within the boundaries. Broadly designed to encompass a complete built-up urbanised centre or town, and usually centred on a territorial local authority; exceptions are Wellington and Auckland, which are each divided into four Main Urban Areas. Classification revised for 1981 Census, and components are:

(i) **Main Urban Areas.** Population of at least 30,000, centred on a major borough. Prior to 1981 Census, criteria was population of at least 20,000; number of Main Urban Areas reduced from 24 to 23; present boundaries remain as fixed for 1971 Census except for minor adjustments.

(ii) **Secondary Urban Areas.** Population of between 10,000 and 29,999, centred on a major borough. Introduced at 1981 Census; fourteen Secondary Urban Areas created including Masterton which was formerly a Main Urban Area.

(iii) **Minor Urban Areas.** All other population centres with populations of at least 1,000 and which do not fall within a Main or Secondary Urban Area.

bias, interviewer
The influence of of the interviewer on the answer supplied by the respondent.

Borough (New Zealand)
A legally defined area which may be created when the population of a specified area attains 1,500, provided that the average density of the population is not less than one person per acre (approximately 2.5 people per hectare). A borough may be called a **City** on reaching a population of 20,000.

census
A periodical enumeration restricted, in modern times, to population, and occasionally to industries and industrial resources, but formerly extending to property of all kinds, for the purpose of assessment.

census coverage
See **coverage, census**

closure, error of
The difference between the census count and its estimate at the census date, obtained by using records of inter-censal births, deaths and external migration to adjust previous census data.

cluster sampling
See **sampling, cluster**

coding errors
See **errors, coding**

cold-deck imputation
See **imputation, cold-deck**

Community (New Zealand)
An area under the jurisdiction of a Community Council which has a population of at least 200 with an average density of at least one person per acre (approximately 2.5 people per hectare), or has at least 60 dwellings with an average density of at least one dwelling to three acres (1 dwelling to 1.2 hectares). Although a borough possesses some autonomy, it ids still the responsibility of the parent county and is included as part of the administrative county.

content error
See **error, content**

County, Administrative (New Zealand)
Territorial local authority. Consists of a county excluding town districts, boroughs and cities which are located within the overall geographic area, and includes any community, district community or unincorporated township which comes under its control.

County (New Zealand)
A legally and geographically defined area.

Community, District (New Zealand) (formerly known as a County Borough)
An area with a minimum population of 1,500, and is officially part of an administrative county.

County, Geographic (New Zealand)
An areal unit based on the outer boundaries of administrative counties, but including all the areas enclosed by the boundary, whether or not they are administratively separate.

coverage, census
The response rate achieved in the census.

cyclic questions
See **questions, cyclic**

data, demographic
Information on the population and its characteristics such as geographical distribution, age distribution, ethnic distribution and level of education.

data item imputation
See **imputation, data item**

data, socioeconomic
Data which can be used to determine the level of prosperity of the population. For example, the median annual income, the average number of cars per family, ownership of boats, caravans or holiday homes.

data, qualitative
Data which cannot be measured, but which can be categorised.

data, quantitative
Measurements, such as heights, weights and lengths.

de jure population
See **population, de jure**

de facto enumeration
See **enumeration, de facto**

de facto population
See **population, de facto**

demographic analysis
See **analysis, demographic**

demographic data
See **data, demographic.**

deviation, standard
A measure of the **spread** of a data sample. The standard deviation of a data sample,

$$s = \sqrt{\textbf{variance}},$$

where the variance,

$$s^2 = \frac{1}{n-1} \Sigma(x - \bar{x})^2$$

where n = sample size and $\bar{x}$ = sample mean. The variance takes into account the squares of the deviations of the observations from the sample mean. The deviations are squared before summation because some observations will be smaller than the sample mean while others will be larger.

Summation of the deviations of the observations from the sample mean will, apart from rounding error, always produce the figure 0.

District Community (New Zealand) (formerly known as a County Borough)
See **Community, District**

Districts (New Zealand)
Designated by the Local Government Commission, and combine counties and interior boroughs into areas which are often neither wholly urban nor wholly rural. Nine such districts had been constituted by the time of the 1981 Census.

Districts, Town (New Zealand)
The smallest local government entities. Under the New Zealand Municipal Corporations Act 1954, they require a minimum population of 500, with an average density of population of at least one person to the acre (approximately 2.5 persons per hectare).

Divisions, Statistical (New Zealand)
Introduced at the 1971 New Zealand Census. Statistically defined areas which show the main population centres of the country. The basic requirement is a minimum population of 75,000 within a relatively compact area, where the population generally shares common economic and community interests. The concept goes beyond that of a conurbation. At the time of the 1981 Census, there were 7 Statistical Divisions, and a significant number of persons within the 7 Divisions were classified as 'rural' residents.

dwelling
Under the Statistics Act 1975, a dwelling is defined as *"any building or structure, whether permanent or temporary, which is wholly or partly used for living purposes"*. Thus the definition includes any shelter in which people are located on census night, whether it be a house, hospital, flat, boarding

house, hotel, motel complex, hospital, prison, hut, tent, car, caravan, ship, which means that even people on an overnight tramp must be located and enumerated.

For census purposes, dwellings are divided into 2 categories: **group-living quarters** (**non-private dwellings**) and **private dwellings.** Group-living quarters cover institutions and other dwellings which have been designed to cater for large groups of individuals or a large number of families, and a private dwelling is any dwelling lived in by one private household and which has its own separate sleeping, cooking and dining facilities.

editing
The detection and identification of errors in a data set.

edits, range
Performed by determining the admissible set of values for each particular variable, and identifying any response which is not in the admissible set.

empirically based residual
See **residual, empirically based**

enumeration, de facto
Enumeration of each individual at the place he or she is on census night.

enumeration, multiple
Occurs when some members of the population counted more than once during a census operation.

equal-weights models
See **models, equal-weights**

error, content
A measure of the inaccuracy of information supplied by the respondents, whether deliberately or unintentionally.

error of closure
See **closure, error of**

error, net census
The difference between the gross undercoverage rate and the rate of erroneous enumerations.

error, nonsampling
All sources of error excluding sampling error, and includes data collection error, coding errors and processing errors.

error, sampling
Error that results from using a probability sample rather than a complete enumeration to estimate a parameter of a finite population.

errors, coding
Errors which occur during coding operation. Include factual errors, interpretation errors and writing errors.

errors, factual
Usually caused by lack of concentration on the part of the coder.

errors, interpretation
Occur when coding schemes are inadequately detailed.

errors, writing
Include transcription and transposition errors, and occur when the handwriting of a respondent is not clearly legible or the coder is not paying sufficient attention. **Transcription errors** occur when one or more digits are misread or miscopied, and **transposition errors** occur when one or more digits are interchanged.

explanatory variables (or independent variables)
See **variables, independent**

factual errors
See **errors, factual**

field maps
See **maps, field**

files, address reference
Files used to associate addresses with their geographic locations.

files, geographic reference
Files used to catalogue the various geographic areas and define their relationships, facilitating an ordered presentation of census data by area.

forward tracing
See **tracing, forward**

frame, sampiing
A complete listing of the members of the population to be sampled, from which the sample is to be drawn.

Geographic County (New Zealand)
See **County, Geographic**

geographic reference files
See **files, geographic reference**

gross undercoverage rate
See **rate, gross undercoverage**

group-living quarters (non-private dwellings)
See **dwellings**

hot deck imputation
See **imputation, hot deck**

household, private
Comprises all persons who live on the same premises and who usually eat one or more meals together daily or, at least, *share the same cooking facilities*. Temporary guests will also be part of the household in which they are staying with. Persons in group living quarters are, for New Zealand census purposes, *not* regarded as living in 'private households'.

imputation
Replacement of missing or erroneous data with data inferred or derived from elsewhere in the questionnaire or from some other source.

imputation, cold-deck
Cold-deck imputation entails using values from some prior distribution to substitute for missing responses. Usually, the distribution used is derived from as similar a population as possible.

imputation, data item
Imputation for item nonresponse, that is imputation for missing value of a variable (as opposed to total nonresponse).

imputation, hot deck
Basically, a record-matching technique in which an incomplete record is compared with a complete record having similar characteristics. The missing field in the incomplete record is then imputed from the value which appears in the corresponding variable (field) in the complete record. Some variations of the hot deck method are:

(1) **Sequential imputation**, where a particular value for a variable of a record is imputed, taking into account the valid variable values in the record, then the next missing variable is imputed, taking into account the valid values,

including the one which has just been imputed, and so on, until all the missing values have been imputed.

(2) **Joint imputation**, where missing values of variables are imputed simultaneously. This method takes into account information in the data record which may be correlated with the value of the variable being imputed. Joint imputation ensures that the incidence of combinations of values will appear in the same proportions as in the population, thus preserving the joint distribution of the variables. Fellegi and Holt (1976) recommend that it be used whenever possible, with sequential imputation used as a default option; that is, used only when joint imputation cannot be achieved.

imputation, total
Imputation for total nonresponse.

independent variables (or explanatory variables)
See **variables, independent**

interpretation errors
See **errors, interpretation**

interquartile range
See **range, interquartile**

interviewer bias
See **bias, interviewer**

logarithmic models
See **models, logarithmic**

mandatory questions
See **questions, mandatory**

maps, field
Maps used by census field staff to ensure that no area of land is either omitted or duplicated during enumeration.

maps, outline
Maps produced to assist those who work with census data in locating the legal and statistical jurisdictions to which the data refer. Include field maps and user maps.

maps, thematic
Used to present the spatial distribution and relative magnitude of a given set of data.

maps, user
Maps defining meshblock/area unit boundaries or presenting statistical data.

mean (arithmetic)
A measure of the **location** of a distribution. The sum of the observations divided by the number of observations. That is,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

where $\sum_{i=1}^{n} X_i$ means the sum of all x's, where each X may have the same value or a different value to the others.

median
A measure of the **location** of a distribution. The middle number of an array of the data of the number of observations is odd. If the number of observations is even, the median is the arithmetic mean of the two middle numbers in the array. The *position* of the median is given by the formula

$$\text{position of median} = \frac{n + 1}{2}$$

where n is the number of observations.

Meshblock
The basic areal unit from which census data is compiled. A meshblock is the smallest land area for which statistical records are kept. In urban areas, the average meshblock may comprise about 50 dwellings and 150-200 persons. In rural areas, meshblocks generally have a population averaging 100-150 persons, although a few may contain no people at all. At the time of the 1981 Census, New Zealand was divided into 33,441 meshblocks.

Metropolitan Regions (New Zealand)
See **Regions, Metropolitan**

models, equal-weights
Models which effectively assume that, for each data item, the missing values have the same distribution as that of the data supplied by respondents. Values for missing data items are imputed using the distribution of the same data items supplied by the respondents.

models, logarithmic
Frequently used for data sets which are highly skewed, since taking the logarithm of data values will reduce the variation between the data values. Data on incomes is invariably skewed, and usually has the bulk of respondents in the low-medium income categories, with relatively few persons receiving abnormally high incomes. Transforming the data by taking the logarithm of the income values will reduce the skewness of the data set.

models, ratio
Models which require the ratio of two responses to lie between prescribed bounds. The upper and lower bounds are determined by historical observation, subject-matter expertise, and, when feasible, by a sample of responses. Ratio edits can incorporate data from an earlier time frame such as previous censuses or surveys, or from external files such as administrative records.

models, regression
Models which use regression methods to estimate the missing data item from the other data items (variables). Can only be used to impute quantitative variables (variables which have numerical values).

net undercoverage rate (net census error)
See **rate, net undercoverage**

New Zealand
Relates solely to 'geographic New Zealand' and excludes Tokelau, but includes Kermadec and Campbell Islands.

non-private dwellings (group-living quarters)
See **dwellings**

nonsampling error
See **error, nonsampling**

outline maps
See **maps, outline**

overenumeration (multiple enumeration)
Occurs when some members of the population counted more than once during a census operation.

population, de facto
The de facto census population of a particular area includes all individuals in that area on Census night, regardless of whether or not it is their normal place of residence.

population, de jure
The normal population of a locality. The de jure census population of a particular area includes all residents of that area, wherever they may happen to be at the time of the census. Overseas residents who were temporarily in the country at the time of the census are excluded from the de jure population counts.

Population, Rural (New Zealand)
Consists of all people in rural areas (i.e. all the people not residing in population centres of 1,000 or more).

Population, Shipboard (New Zealand)
The persons enumerated on board vessels in New Zealand waters on census night. The enumerations are included in populations of Statistical Areas and Divisions, North and South Islands and Total New Zealand, but are not included in the populations of other areal subdivisions.

population, target
The population in which the data collector is interested. In the case of a census, the target population is the entire population; i.e. every individual.

Population, Total (De Facto Population)
The actual population counted at the places where they were enumerated on census night.

predictor variables (independent variables)
See **variables, independent**

private dwellings
See **dwellings**

private household
See **household, private**

purposive sample
See **sample, purposive**

quantitative data
See **data, quantitative**

qualitative data
See **data, qualitative**

questions, ad hoc
Questions which are generally included in a census questionnaire on a one-time basis, in order to provide national and sub-national statistics for some particular purpose.

questions, cyclic
Those questions which are regarded as being unnecessary for inclusion in every census, but which need to be asked at regular intervals (such as every 10 years).

questions, mandatory
Those questions which must be asked in every census, according to legal statute. The mandatory questions contained in New Zealand censuses cover the following subjects: name, address, sex, and ethnic origin of every occupant of the dwelling and particulars of the dwelling as to location, number of rooms, ownership, and number of occupants on Census Night.

questions, standard
Those questions which are regarded as being necessary for inclusion in every census, because they are considered to be of national value. Examples of topics covered by standard questions are occupation, employment status, number of children born, and rent paid.

random sampling
See **sampling, random**

range
A measure of the **spread** of a distribution. The difference between the two extreme observations.

i.e. range = largest observation - smallest observation

range edits
See **edits, range**

range, interquartile
A measure of the **spread** of a distribution. The difference between the upper and lower quartiles, that is

Q3 - Q1,

where the quartiles are the three values which divide the data into four equal parts. The **first quartile** is usually known as the **lower quartile**, and it divides the first quarter of the data from the upper three quarters. The **second quartile** is another name for the **median**, and the **upper quartile**, or **third quartile**, separates the top quarter from the bottom three quarters. The shorthand notation Q1 is often used to represent the lower quartile, and Q3 for the upper quartile. The *position* of the lower quartile, Q1 is given by

$$\frac{1}{4}(n + 1)$$

and the *position* of the upper quartile, Q3 is given by

$$\frac{3}{4}(n + 1).$$

rate, gross undercoverage
The estimate of the proportion of the national population missed by the census.

rate, net undercoverage
The difference between the gross undercoverage rate and the rate of erroneous enumerations.

ratio models
See **models, ratio**

Region (New Zealand)
Abbreviation of the term **Local Government Region**. At the time of the 1981 New Zealand Census, all territorial local authority areas with the exception of the Chatham Islands County were jointly covered by 22 regions. A region does not include extra-county islands and shipboard populations.

Regions, Metropolitan (New Zealand)
The four combined main urban areas of Auckland and Wellington respectively form the Auckland and Wellington Metropolitan Regions.

regression models
See **models, regression**

residual, empirically based
A residual which is obtained from observation, rather than from a formula.

Reverse Record Check (RRC)
An evaluation programme in which a sample of the population is drawn from a frame created several years prior to the census, traced forward to the time of the census, and matched to the census.

retrospective tracing
See **tracing, retrospective**

Rural Areas (New Zealand)
See **Areas, Rural**

Rural Population (New Zealand)
See **Population, Rural**

sample, purposive
A sample designed with the emphasis on the range of the relevant characteristics of the population being included, rather than the distribution matching that of the wider population.

sampling, cluster
Usually applied when it is desired to achieve considerable savings in terms of time and travelling expenses. Every member of a randomly selected group or **cluster** of the sampling units is selected. For instance, every member of a class of third-form students in each of a few secondary schools could be interviewed to determine the number of households with colour television sets, spa pools, or microwave ovens. Cluster sampling avoids the problem of constructing a frame for the entire population, and is extremely quick to conduct, since the units in a cluster are adjacent, and hence the explanation as to why the survey is being conducted, and the instructions to the respondents need only be given once to each selected cluster. However, the accuracy of the estimates obtained is reduced, since a group of 40 students studied in each of 5 specified schools will not usually be as representative of the entire population as would a simple random sample of 200 students. Moreover, although each member of a cluster is interviewed, the responses obtained from members of a cluster may well be affected by the opinions of other members of that cluster, and hence a reduced number of independent responses will be obtained.

sampling error
See **error, sampling**

sampling frame
See **frame, sampling**

sampling proportional to size
Sampling procedure in which the size of the sample selected in each strata is proportional to the total sample size.

sampling, random
Sampling procedure in which the sampling units are selected impartially. Random number tables or an appropriate computer-generated list of random numbers usually employed.

sampling units
See **units, sampling**

sampling with replacement
See **with replacement, sampling**

sampling, stratified random
A sampling design which first divides the population into **homogeneous strata**, and then draws independent random samples from these individual subpopulations. Members of such a subpopulation or strata will be similar with respect to the characteristic in question. Because stratification takes advantage of the known homogeneity of the subpopulations, only relatively small samples are necessary to estimate the characteristic for each subpopulation. An estimate for the whole population is then easily obtained by combining these individual estimates.

sampling, systematic
A sampling procedure in which the units are selected nonrandomly, the aim being to ensure that the selected units are spread evenly throughout the sampling frame. The periodicity of the selected units is calculated as follows:

$$k = N/n,$$

where k is the periodicity, N is the population size, and n is the desired sample size.

After one unit is selected from the first k units of the sampling frame, every following kth unit is automatically selected. It is possible to introduce an element of randomness by randomly electing the first unit, but the selection of the first unit will determine the selection of the remaining sample units.

For example, if the population size is 6,000 and the required sample size is 120, then after the initial unit is selected, every 50th unit thereafter is selected, since

$$k = 6000/120 = 50.$$

If the sample frame to be used consists of a list, such as a telephone directory, a stack of file cards or a computer file, or a roster of names, systematic will obviously be very simple and convenient to operate. However, care must be taken to ensure that the list order is not relevant to the characteristic being studied. If the list contains hidden periodicities, then a seriously nonrepresentative sample could be obtained. For example, a classroom of children could be asked to group themselves into pairs, and one from each pair is asked for the names of the pair. After recording such a list, it could be decided to allocate one teaching method to the first member of each pair, and to allocate another method to the remaining member. The validity of the result of such an experiment could be questioned, particularly if, as is often the case, the child who volunteered the names of his/her pair is the more dominant member of the pair.

However, if the list order is not relevant to the characteristic in question, then systematic sampling may be regarded as an approximation to simple random sampling.

sampling without replacement
See **without replacement, sampling**

Shipboard Population (New Zealand)
See **Population, Shipboard**

socio-economic data
See **data, socioeconomic**

standard deviation
See **deviation, standard**

standard questions
See **questions, standard**

Statistical Area (New Zealand)
See **Area, Statistical**

Statistical Divisions (New Zealand)
See **Divisions, Statistical**

stratified random sampling
See **sampling, stratified random**

systematic sampling
See **sampling, systematic**

target population
See **population, target**

testing
Terms such as **pilot test** and **pretest** have no agreed meaning, and are often used interchangeably. We will adhere to the following interpretation of their terms adopted by the New Zealand Department of Statistics:

> **dress rehearsal**
> The final field test(s) with the questionnaire and the processing of the data thus collected to test the coding, editing, tabulation and publication procedures
>
> **field testing**
> A general term used to cover the overall exercise of testing concepts, questionnaire wording and layout, procedures, etc. of a survey or census.
>
> **pilot testing**
> The overall process of developing the census questionnaires and field operations (such as delivery, collection and initial checking of the questionnaires).

pretesting
The process of developing specific aspects of the census, such as particular questions asked or procedures (methods) used.

thematic maps
See **maps, thematic**

total imputation
See **imputation, total**

Total Population (De Facto Population)
See **Population, Total**

Town Districts (New Zealand)
See **Districts, Town (New Zealand)**

Township (New Zealand)
A non-administrative centre which has no legal boundaries. Townships include the nucleus of a self-contained community (e.g. post office, school, hall and several shops).

tracing, forward
Uses a sample from a previous census, a sample from immigration records, and a sample of missed people from the census, but unlike the RRC, the tracing begins at the beginning of the period

tracing, retrospective
Tracing after the current census has been completed.

underenumeration
Occurs when some members of the population are not counted during a census operation.

Urban Areas (New Zealand)
See **Areas, Urban**

units, sampling
Individual members of the population.

Urban Area
See **Area, Urban**
user maps
See **maps, user**

variables, independent (explanatory, predictor)
The data variables which are used for the basis of predicting the unknown variable.

with replacement, sampling
Sampling procedure in which, after selection, the person may be re-selected at any later stage.

without replacement, sampling
Sampling procedure in which every member of the population in question has an known probability of being selected in the sample, but once selected, he or she is not eligible for re-selection. This technique is usually referred to as **simple random sampling**. Because measuring a sampling unit (i.e., recording their response) more than once does not contribute new information, sampling without replacement will collect more information about the population in question than will sampling with replacement.

writing errors
See **errors, writing**

REFERENCES

A Guide to the Content of Questionnaires for the 1981 Census of Population and Dwellings of New Zealand, 1982, 7, 9, 11-15 (Department of Statistics, Wellington).

Bailar, Barbara A. Comment on "Estimating the Population in a Census Year 1980 and Beyond" by Ericksen, E.E. and Kadane, J.B. (1985) *Journal of the American Statistical Association March 1985, 80, 389, Theory and Methods,* 109-114 (American Statistical Association).

Bailar, J.C. and Bailar, B.A. "Comparison of Two Procedures for Imputing Missing Survey Values." *Proceedings of the Section on Survey Research Methods*, 1978, 65-81 (American Statistical Association).

Barabba, V.P., Mason, R.O. and Mitroff, I.I. "Federal Statistics in a Complex Environment: The Case of the 1980 Census" *The American Statistician*, August 1983, 37, *3*, 203-212 (American Statistical Association).

Boreham, J. "Present Position and Potential Developments: Some Personal Views on Official Statistics". *J.R.Statist.Soc. A*, 1984, 147, Part 2, 174-185.

Bounpane P.A. and Jones, T.A. "Automation of the 1990 U.S. Census" *Chance - New Directions for Statistics and Computing*, 1988, 1, *1*, 28-35. (Springer-Verlag New York Inc.)

Brant, J.D. and Chalk, S.M. "The Use of Automatic Editing in the 1981 Census" *Journal of the Royal Statistical Society* (A), 1985,*148*, Part 2, 126-146.

Butz, W.P. "The Future of Administrative Records in the Census Bureau's Demographic Activities." *Proceedings of the Section on Survey Research Methods*, 1984, 61-63 (American Statistical Association).

Cartwright, D.W., Levine, B. and Buckler, W.L. "An Update on Establishment Reporting Issues: Practical Considerations." *Proceedings of the Section on Survey Research Methods,* 1983, 481-486 (American Statistical Association).

Census of New Zealand, 1916 Householder's Schedule

Census of Population and Dwellings 1971 - The New Zealand People Department of Statistics pp 137-139.

Chapman, D.W. "A Survey of Nonresponse Imputation Procedures" *The Southern Region Education Board Summer Research Conference on Statistics 1973* (Westat, Inc).

Childers, D.R. and Hogan, H. "Census Experimental Match Studies" *Proceedings of the Section on Survey Research*, 1983,173-176 (American Statistical Association).

Cho, L.J. and Hearn, R.L. *Censuses of Asia and the Pacific: 1980 Round*, 1984, 1-12, 31-40, 63-80, 137-148, 241-263, 270, 281-282, 349-354. (East-West Population Institute East-West Centre, Honolulu, Hawaii).

Cohen, I. B. "Florence Nightingale", *Scientific American,* March 1984, 98-107.

Cook, L.W. "Statistics: At What Price?" New Zealand Statistical Association Conference, Christchurch August 1987. *New Zealand Statistical Association (Inc) Newsletter,* April 1988,*17* (NZSA, Wellington).

Cowan, C.D. and Fay, R.E. "Estimates of Undercount in the 1980 Census" *Proceedings of the Section on Survey Research*, 1984, 566-571 (American Statistical Association).

David, M. and Triest, R. "The CPS Hot Deck: An Evaluation Using IRS Records" *Proceedings of the Section on Survey Research Methods*, 1983, 421-426 (American Statistical Association).

DeMaio, T.J. Learning from Interviewers. *Proceedings of the Section on Survey Research Methods*, 1983, 669-674 (American Statistical Association).

Diffendal, G. J., Isaki, C.T. and Malec, D. "Some Small Area Adjustment Methodologies Applied to the 1980 Census" *Proceedings of the Business and Economic Statistics Section*, 1983, 164-167. American Statistical Association.

Dixon, W.J. "Analysis of Extreme Values" *Annals of Mathematical Statistics* 1950, *21*, 488-506.

Dixon, W.J. "Processing Data for Outliers" *Biometrics* , 1953, 9,*1* , 74-89.

Dodge, R.W. "Using Record Checks." *Proceedings of the Section on Survey Research Methods*, 1983, 680-685 (American Statistical Association).

Domesday 1086-1986: An exhibition to celebrate the 900th anniversary of Domesday Book (1986,1-64 (Public Record Office).

Encyclopaedia Britannica (USA Edition), 1970, *5*, 167-168.

Encyclopaedia Britannica 9th Edition, 1875, *5*, 334, 335.

Encyclopaedia Britannica 11th edition, 1911, *V*, 662-664.

Encyclopaedia of Statistical Sciences, 1982, *I,* 399-400.

Ericksen, E.P. and Kadane, J.B. "Estimating the Population in a Census Year 1980 and Beyond" *Journal of the American Statistical Association March 1985*, 80, *389*, Theory and Methods, 1985, 98-109 (American Statistical Association).

Ericksen, E.P. and Kadane, J.B. "Using Administrative Lists to Estimate Census Omissions: An Example" *Proceedings of the Section on Survey Research Methods,* 1983, 361-365. (American Statistical Association.)

Fan, M.C., Sutt, M.L. and Thompson, J.H. "Evaluation of the 1980 Census Precanvass Coverage Improvement Operations" *Proceedings of the Section on Survey Research* , 1984, 519-524 (American Statistical Association).

Fay R. and Cowan C. "Missing Data Problems in Coverage Evaluation Studies" *Proceedings of the Section on Survey Research*, 1983, 158-163 (American Statistical Association).

Fay, R.F. Comment on "Estimating the Population in a Census Year 1980 and Beyond" by Ericksen, E.E. and Kadane, J.B. *Journal of the American Statistical Association* , March 1985, 80, *389*, Theory and Methods, 114-116 (American Statistical Association).

Federal Statistical Office, West Germany. *"Problems Experienced During the Preliminary , Main and Follow-up Surveys - Problems of the 1983 Census in the Federal Republic of Germany",* 1984, 1-17.

Fellegi, I.P. and Holt, D. (1976). "A Systematic Approach to Automatic Edit and Imputation" *Journal of the American Statistical Association,* March 1976, 71, *353* , 17-35.

Fellegi, I.P. Comment on "Estimating the Population in a Census Year 1980 and Beyond" by Ericksen, E.E. and Kadane, J.B.

Journal of the American Statistical Association March 1985, 80, *389*, Theory and Methods, 116-119. (American Statistical Association).

Fowler, F.J. and Mangione, T.W. "Standardised Survey Interviewing". *Proceedings of the Section on Survey Research Methods,* 1983, 782-787 (American Statistical Association).

Greenberg, B. and Surdi, Rita "A Flexible and Interactive Edit and Imputation System for Ratio Edits" *Proceedings of the Section on Survey Research Methods,* 1984, 421-426 (American Statistical Association).

Grubbs, F.E. "Procedures for Detecting Outlying Observations in Samples" *Technometrics* , 1969,2,*1*, 1-21.

Grubbs, F.E. "Sample Criteria for Testing Outlying Observations in Samples" *Annals of Mathematical Statistics* 21, 1950, 27-58.

Hansen, M.H. Comment on "Estimating the Population in a Census Year 1980 and Beyond" by Ericksen and Kadane *Journal of the American Statistical Association,* March 1985, 80, *389*, Theory and Methods, 119-122. (American Statistical Association).

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. *Sample Survey Methods and Theory*, Vol. II, 1953 (New York: John Wiley and Sons).

Hinkins, S.M. Matrix Sampling and the Related Imputation of Corporate Income Tax Returns. *Proceedings of the Section on Survey Research Methods*, 1984, 427-433. (American Statistical Association.)

Hogan, H. "The Forward Trace Study: Its Purpose and Design" *Proceedings of the Section on Survey Research Methods*, 1983, 168-172 (American Statistical Association).

Icon IICT The Newsletter of the Institute of Information and Communications Technologies, 1989, 3 (CSIRO Australia).

In-house Publication. *"1986 Census of Population and Dwellings Fact Finding Group E Sampling"*, 1982, 9 (Department of Statistics, Christchurch).

International Encyclopaedia of Statistics pp 42, 43.

Jabin, T.B. "Goals for Statistical Uses of Administrative Records: The Next Ten Years." *Proceedings of the Section on Survey Research Methods*, 1983, 66-75 (American Statistical Association).

Johnson, R.A. and Woltman, H.F. "Evaluating Census Data Quality Using Intensive Interviews: A Comparison of U.S. Census Bureau Methods and Rasch Methods" *Sociological Methodology,* 1987,*17*, 185-203 (The American Sociological Association).

Kaiser, J. "The Effectiveness of Hot-Deck Procedures in Small Samples" *Proceedings of the Section on Survey Research Methods*, 1983, 523-528 (American Statistical Association).

Keeley, Catherine and Thompson, J. "The 1980 Census Nonhousehold Sources Program" *Proceedings of the Section on Survey Research Methods* , 1984, 531-536 (American Statistical Association).

Keesing's Record of World Events, June 1987, Volume XXXIII, 35209-35210 (Longman).

Khawaja, M.A. "Evaluation of the Quality of Demographic Data" *The Population of New Zealand* Appendix II, 1982, 19. ESCAP, Bangkok, unpublished.

Lewthwaite, G. "The Population of Aotearoa: Its Number and Distribution" *New Zealand Geographical*, 1950, 6, *1*, 36-39.

Leyes, J. "The Use of Individual Administrative Records for Social Statistical Purposes in Canada." *Proceedings of the Section on Survey Research Methods*, 1983, 624-626 (American Statistical Association).

Lillard, L., Smith, J.P. and Welch "What do we Really Know about Wages: The Importance of Non-Reporting and Census Imputation", 1982 *The Rand Corporation* 1700 Main St, Santan Monica, CA 90406.

Little, R.J.A and Samuhel, M.E. "Alternative Models for CPS Income Imputation" *Proceedings of the Section on Survey Research Methods*, 1983, 415-420. (American Statistical Association.)

Lueck, M., Harahush, T. and Sledge, G. "The Prelist Recanvass and Local Review Coverage Improvement Operations" *Proceedings of the Section on Survey Research Methods*, 1984, 537-540 (American Statistical Association).

Mitroff, I.I., Mason, R.O. and Barabba, V.P. *"The 1980 Census: Policymaking Amid Turbulence"* ,1982 (Lexington).

National Data Book and Guide to Sources "Statistical Abstract of the United States 1988" 108th Edition,1,7 (U.S. Department of Commerce Bureau of the Census).

Nelson, Dawn D. "Informal Testing" *Proceedings of the Section on Survey Research Methods*, 1983, 665-668 (American Statistical Association).

New Zealand Census of Population and Dwellings 1951 General Report, 1956, *VIII*, 9-10 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings 1971 The New Zealand People 1971, 1976, *12,* 137 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings 1976 Internal Migration, 1981, *11,* 152-155, 157-159, 161-163 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings, 1976 Sub-Enumerators Reference Manual, 1975, 9, 157-158 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings, 1981 Population Perspectives '81, 1985, Volume 12 (General Report), 21-22, 54,112,187-188 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings 1981 Enumerator's Handbook , 1980, 7, 8, 13, 14, 38-40, 97 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings 1981 Sub-Enumerators Reference Manual, 1980, 8-9, 16-18, 69 (Department of Statistics).

New Zealand Census of Population and Dwellings 1986 Enumerator's Handbook. 1985, Department of Statistics, 11-12.

New Zealand Census of Population and Dwellings, Census '86, Questionnaire Content and Submissions, 1985, Series D Report 1, 10-11 (Department of Statistics, Wellington).

New Zealand Census of Population and Dwellings, Census '86, General Information, 1988, Series D General Report 3, 28-33 (Department of Statistics, Wellington).

New Zealand Official Yearbook 1985, pp 71,81. Department of Statistics, Wellington.

O'Brien, R.F. "Relative Coverage in the 1980 Census of Puerto Rico" *Proceedings of the Business and Economic Statistics Section,* 1983, 541-549 (American Statistical Association).

Pool, D.I. "*The Maori Population of New Zealand 1769-1971*", 1977, 40-71

Population Division News Bulletin. *"Availability of Information from 1981 Census of Population and Dwellings"*, September 1983 (Department of Statistics, Christchurch).

Population Division. *"Technical Details of Magnetic Tapes and Associated Charges"*, 1984 (Department of Statistics, Christchurch).

Population Perspectives '81 *New Zealand Census of Population and Dwellings, 1981* (General Report) 1985, 149-151,158-169,186-193 (Department of Statistics, Wellington).

Press Releases Nos 103-210 1957-58 (Department of Statistics, Wellington).

Questionnaire Content and Submissions, New Zealand Census of Population and Dwellings, Census '86, 1985, 119-120 (Department of Statistics, Wellington).

Reed, A.H., Henry, R.J. and Mason, W.B. "Influence of Statistical Method Used on the Resulting Estimate of Normal Range" *Clinical Chemistry* , 1971, *17* , *4* , 275-284.

Results of a Census of the Dominion of New Zealand General Reports, 1925, 93-96 (NZ Census and Statistics Office, Wellington).

Salmond, Clare E. "Data Editing: Methods of Quality Control" *Blue Book Series 11,* 1981, 1-59. The Management Services and Research Unit, Department of Health, Wellington.

Sande, G. "Numerical Edit and Imputation" *International Association for Statistical Computing*, 42nd Session of the International Statistics Institute, 1979.

Scheuren, F. "Design and Estimation for Large Federal Surveys Using Administrative Records." *Proceedings of the Section on Survey Research Methods*, 1983, 377-381 (American Statistical Association).

Schieber, S.J. "A Comparison of Three Alternative Techniques for Allocating Unreported Social Security Income on the Survey of the Low-Income Aged and Disabled" *Proceedings of the Section on Survey Research Methods*, 1978, 212-218 (American Statistical Association).

Schroeren, J.M. Volkszahlung 1983: "Politiker fragen - Burger antworten nicht!" *Umweltmagazin* , 6/82, Berlin 1982, 35-36.

Siegel, J.J., Passel, J.S., Rives, N.W. and Robinson, J.G. "Developmental Estimates of the Coverage of the Population of States in the 1970 Census: Demographic Analysis" *Current Population Reports*, 1977, P-23, 65, Washington D.C. (US Bureau of the Census).

Simkin, C.G.F. *Statistics of New Zealand for the Crown Colony Period; 1840-1852* ,1954, 37 (University of Auckland).

Sledge, G., Harahush, T. and O'Brien, R. "Misclassified/Occupied and H-4 Edit Coverage Operations" *Proceedings of the Section on Survey Research Methods,* 1984, 525-536 (American Statistical Association).

Statistics of New Zealand, 1858 ,1859, iii-vi (New Zealand Government).

Statistics of New Zealand, for 1853, 1854, 1855, and 1856, 1858, iii-vii (New Zealand Government).

Statistics of New Zealand, for 1857, 1858 (New Zealand Government).

Streett, Anitra R. and Smith, W. "Investigating Respondent's Interpretations of Survey Questions". *Proceedings of the Section on Survey Research Methods*, 1983, 675-679 (American Statistical Association) .

Streett, Anitra R. "Unstructured Individual Interviewing" *Proceedings of the Section on Survey Research Methods* 1983, 661-664 (American Statistical Association).

Thomas, Kathryn F. and Whitford, D.C. "Post Office Effectiveness" *Proceedings of the Section on Survey Research Methods,* 1984, 513-518 (American Statistical Association).

Tukey, J.W. Comment on "Estimating the Population in a Census Year 1980 and Beyond" by Ericksen and Kadane *Journal of the American Statistical Association March 1985*, Vol. 80, No. 389, Theory and Methods, 127-128 (American Statistical Association).

In-house report *Plan for 1986 Census of Population and Dwellings*, 26, 27 (Department of Statistics).

US Department of Commerce. *Census '80: Continuing the Factfinder Tradition* , 1980, 69,182-184, 264-279, 343-344 (US Bureau of the Census).