

第12次人口會議結果報告書

TWELFTH POPULATION CENSUS CONFERENCE
BEIJING, CHINA

1988. 9

經濟企劃院 調查統計局

National Bureau of Statistics
Economic Planning Board

033523

目 次

I. 第 12 次 人口센서스 會議 概要	3
1. 會議開催	5
2. 會議目的	5
3. 討論議題	5
4. 會議日程	8
II. 參考資料	9
1. 參席者 名單	11
2. 會議日程	21
3. 韓國代表의 基調演說文	29
4. 韓國의 1985 年 人口 및 住宅센서스 確定 結果	39
III. 各國의 Country Report	83
1. A New Processing System for Australia's 1991 Consus	85
2. The 1991 Census of Canada: Determining the Questions to be Asked	101
3. The Utilization of China's 1982 Population Census Data	113
4. The 1990 Population Census in China:Prepations and Basic Ideas	137
5. In-Depth Fertility Survey in China	153
6. Automation and Data Processing of the Hong Kong 1991 Population Census	171

7. Processing of Data on the 1986 Population Census of Fiji	183
8. Approach to the 1991 Census of India	197
9. A Note on the Methodology of the 1990 Population Census of Indonesia	239
10. Proposed Measures for Improving Coverage of the 1990 Population and Housing Census of Malaysia	255
11. The Use of Computerised Technology in the 1990 New Zealand Census of Population and Dwellings	267
12. Some Policy Considerations for Planning the 1991 Census of Housing and Population of Pakistan	281
13. The Population Census of Papua New Guinea with Special Reference to the 1990 Census	311
14. Urbanization Index Based on the Census of Population and Housing Results of Philippines	325
15. Measurement of Economic Activity in Sri Lanka	349
16. Planning of the 1990 Population and Housing Census of Thailand ...	379
17. Status Report on the 1990 Census of Population and Housing of the United States	395
18. Use of Sampling in Conjunction with the Census of Population ...	411
19. Use of Optical Mark Reading Systems in the Census Process	439
20. The Use of Distributed, Low Cost Optical Mark Reading (OMR) for Census Data Collection	459
附錄：中國的人口・經濟統計現況	469

I. 第12次 人口센서스會議概要

I. 第 12 次 人口센서스會議 概要

1. 會議開催

1) 主 催

- 美國 東西센터 人口研究所 (Population Institute, East West Center)
- 中國 中央統計廳 (State Statistical Bureau of China)

2) 日 字 : 1988.9.5 - 9.9 (5 日間)

3) 場 所 : 中國北京市

4) 參加 : 아시아・太平洋地域	17 個國	代表	28 名
國際機構	3 機關	代表	8 名
其 他		代表	8 名

5) 會議開催経緯

- 第 7 次會議 : 1979. 9.24 - 9.29 美國 하와이
- 第 8 次會議 : 1981. 9.28 - 10. 2 韓國 서울
- 第 9 次會議 : 1983. 3. 1 - 3. 5 日本 東京
- 第 10 次會議 : 1985. 4.29 - 5. 3 美國 하와이
- 第 11 次會議 : 1987. 2. 9 - 2.13 濠洲 시드니・캔버라
- 第 12 次會議 : 1988. 9. 5 - 9. 9 中國 北京

2. 會議目的

同 會議는 參加國의 統計機關 責任者 및 人口統計 專門家들간에 相互統計技術協力增進 方案과 人口센서스의 効率的인 實施・分析 및 利用技法에 관한 論議가 目的임.

3. 討論議題

1) 컴퓨터技法 活用關係

- 홍콩의 1991 年 人口센서스 資料處理와 自動化

- 뉴질랜드의 1990년 인구 및住宅센서스에 있어 自動化技法活用
 - 호주의 1991년 인구 및住宅센서스를 위한 새로운 資料處理技法
- 2) 1990年度 센서스 企劃關係
- 인도의 1990-1991년 센서스 接近
 - 美國의 1990년 인구 및住宅센서스 現況報告
 - 파키스탄의 1991년 인구 및住宅센서스 企劃을 위한 政策的 고려
 - 파푸아뉴기니의 센서스
 - 태국의 1990년 인구 및住宅센서스準備
- 3) 中國의 센서스와 標本調查關係
- 中國의 1990년 人口센서스準備 및 基本方向
 - 中國의 1982년 人口센서스 資料分析活用
 - 1987년 1% 人口標準調查結果
 - 中國의 深層出産力調查
- 4) 센서스資料로부터의 人口變數推定關係
- 經濟活動과 人口移動의 測定
 - 人口 및 住宅센서스 結果를 기초로한 都市化指數
 - 센서스資料를 活用한 死亡率推定
 - 幼兒死亡率의 推移
- 5) 센서스資料의 蒐集, 處理 및 活用을 위한 새로운 技法關係
- OMR 시스템의 活用
 - 既開發되고 저렴한 OMR의 利用
 - 센서스資料가 經濟적이고 廣範圍하게 使用되기 위한 技法
 - 마이크로 컴퓨터의 活用關係

6) 기타 센서스의 주요 장점 關係

- 센서스에 있어서 標本調査利用
- 센서스資料의 質的問題
- 캐나다의 1991年 센서스에서의 調査項目
- 말레이시아의 1990年 人口 및 住宅센서스에 있어서 누락을 방지
를 위한 제안
- 인도네시아의 1990年 人口센서스 실시 技法

7) 1985년 센서스 結果關係

- 韓國의 1985年 人口 및 住宅센서스 確定結果
- 피지의 1986年 人口센서스의 資料處理
- 日本의 1985年 및 1990年 人口센서스의 考察

4. 會議日程

日 字	日 程
9. 5	<ul style="list-style-type: none"> ○ 開 會 式 ○ 컴퓨터技法 活用關係 및 1990년 센서스 企劃關係 論文發表 ○ 中國 中央統計廳主催 晚餐會
9. 6	<ul style="list-style-type: none"> ○ 中國의 센서스와 標本調查關係 및 센서스資料로부터의 人口變數推定 關係論文發表 ○ 美國 東西센터 人口研究所 主催 晚餐會
9. 7	<ul style="list-style-type: none"> ○ 박물관 및 萬里長成訪問
9. 8	<ul style="list-style-type: none"> ○ 센서스資料의 蒐集·處理 및 活用을 위한 새로운 技 法關係論文發表 ○ 컴퓨터 關聯會社主催 晚餐會
9. 9	<ul style="list-style-type: none"> ○ 기타 센서스의 主要爭点關係 및 1985年 센서스結果 關係論文發表 ○ 閉會式
9.10 - 12	<ul style="list-style-type: none"> ○ 西安訪問
9.12 - 15	<ul style="list-style-type: none"> ○ 桂林訪問

II. 參 考 資 料

1. 參 席 者 名 單

TWELFTH POPULATION CENSUS CONFERENCE

September 5-9, 1988
Beijing, China

Participant List

Australia

Dr. Mike Giles
Assistant Statistician
Social Branch
Demography, Labour and Social Division
Australian Bureau of Statistics
P.O. Box 10
Belconnen, ACT 2616, Australia

Bangladesh

Mr. Md. Shahadat Hossain
Joint Director
Bureau of Statistics
Ministry of Planning
Bangladesh Secretariat
Dhaka 2, Bangladesh

Canada

Mr. David Bruce Petrie
Assistant Chief Statistician
Statistics Canada
Ottawa, Ontario
Canada K1A 0T6

China

Mr. Sun Jingxin
Deputy Director-General
State Statistical Bureau
38 Yuetan Nanjie
Sanlihe
Beijing, China

Mr. Li Chengrui
Honorary President
National Statistical Society of China
c/o State Statistical Bureau
38 Yuetan Nanjie
Sanlihe
Beijing, China

Mr. Shen Yimin
Deputy Director
Department of Population Statistics
State Statistical Bureau
38 Yuetan Nanjie
Sanlihe
Beijing, China

Twelfth Population Census Conference
Participant List
Page 2

Mr. Wang Shulin
Deputy Director
Department of Population Statistics
State Statistical Bureau
38 Yuetan Nanjie
Sanlihe
Beijing, China

Mrs. Liu Changhong
Statistician
Department of Population Statistics
State Statistical Bureau
38 Yuetan Nanjie
Sanlihe
Beijing, China

Fiji

Mr. Naibuka Bolaqace Navunisaravi
Government Statistician and
Census Commissioner
Bureau of Statistics
P.O. Box 2221
Suva, Fiji

Hong Kong

Mr. Joseph Man-Kong Lee
Assistant Commissioner
Census and Statistics Department
21/F Wanchai Tower I
12 Harbour Road
Wanchai, Hong Kong

India

Mr. V.S. Verma
Registrar General
Office of the Registrar General
Ministry of Home Affairs
2/A Mansingh Road
New Delhi 110011, India

Mr. N. Rama Rao
Assistant Registrar General
Office of the Registrar General
Ministry of Home Affairs
2/A Mansingh Road
New Delhi 110011, India

Twelfth Population Census Conference
Participant List
Page 3

Japan

Mr. Akihiko Ito
Director
Consumer Statistics Division
Statistical Survey Department
Statistics Bureau
Management and Coordination Agency
19-1 Wakamatsu-cho, Shinjuku-ku
Tokyo 162, Japan

Mr. Yuki Miura
Vice-Chairman and Executive Director
Japan Statistical Association
19-1 Wakamatsucho, Shinjuku-ku
Tokyo 162, Japan

Korea

Mr. Kang-Woo Lee
Director-General
National Bureau of Statistics
Economic Planning Board
90 Gyeongun-dong, Jongro-gu
Seoul 110, Korea

Mr. Il-Hyun Kim
Director
Population Statistics Division
National Bureau of Statistics
Economic Planning Board
90 Gyeongun-dong, Jongro-gu
Seoul 110, Korea

Nepal

Mr. Bishwa Raj Regmi
Director-General
Central Bureau of Statistics
Thapathali
Kathmandu, Nepal

New Zealand

Mr. Michael Anthony Moore
Director
Population Census Division
Department of Statistics
Private Bag
Christchurch, New Zealand

Twelfth Population Census Conference
Participant List
Page 4

Pakistan

Mr. Muhammad Aslam Chaudhry
Census Commissioner
Population Census Organization
Statistics Division
Government of Pakistan
Post Office Box 1026
Islamabad, Pakistan

Papua New Guinea

Mr. Nick Suvulo
National Statistician
National Statistical Office
Post Office Wards Strip
Waigani, Papua New Guinea

Philippines

Dr. Nelia R. Marquez
Deputy Administrator
National Census and Statistics Office
Post Office Box 779
Manila, Philippines

Sri Lanka

Mr. Raja Bertram Maligaspe Korale
Director
Department of Census and Statistics
Government of Sri Lanka
6, Albert Crescent
Colombo 7, Sri Lanka

Thailand

Ms. Chintana Pejaranonda
Chief
Population Studies Section
Population Survey Division
National Statistical Office
Lan Luang Road
Bangkok 10100, Thailand

Mr. Nerawit Dhanasakdi
Chief
Population and Housing Statistics Section
Population Survey Division
National Statistical Office
Lan Luang Road
Bangkok 10100, Thailand

Twelfth Population Census Conference
Participant List
Page 5

United States

Ms. Susan Mary Miskura
Chief
Decennial Planning Division
U.S. Bureau of the Census
Washington, D.C. 20233
U.S.A.

Dr. Daniel Melnick
Professional Staff Member
Joint Economic Committee
United States Congress
1537 Longworth House Office Building
Washington, D.C. 20515
U.S.A.

Mr. Philip Neal Fulton
Assistant Chief for Census Programs
Population Division
U.S. Bureau of the Census
Washington, D.C. 20233
U.S.A.

Mr. Glenn Stephen Ferri
International Technical Advisor (Data Processing)
U.S. Bureau of the Census
International Statistical Programs Center
Scuderi Building
Washington, D.C. 20233
U.S.A.

Resource Persons

Dr. William Brass
Professor
Department of Medical Statistics and
Epidemiology
London School of Hygiene and Tropical Medicine
99 Gower Street
London WC1E 6AZ, United Kingdom

Dr. Vijay Kumar Verma
Consultant
105 Park Road
Teddington TW11 0AW
United Kingdom

Twelfth Population Census Conference
Participant List
Page 6

United Nations Population Fund (UNFPA)

Mr. Prod A. Laquian
Deputy Representative
United Nations Population Fund
No. 2 Dongqijie, Sanlitun
Beijing, China

Economic & Social Commission for Asia & the Pacific

Mr. Laurence Henry Lewis
Regional Adviser on Population Censuses and
Surveys
Economic and Social Commission for Asia and
the Pacific
The United Nations Building
Rajadamnern Avenue
Bangkok, Thailand 10200

Computer Representatives

Mr. Wilbert Riley
Vice-President
National Computer Systems
3 Neshimany Interplex, Suite 201
Trevose, Pennsylvania 19047
U.S.A.

Dr. John Sirles Massey
Managing Director
Space-Time Research Pty. Ltd.
668 Burwood Road
Hawthorn East
Victoria 3123, Australia

Mr. Stephen Stewart
Technical Director
DRS Data and Research Services PLC
Sunrise Parkway
Linford Wood
Milton Keynes MK14 6LR
United Kingdom

Twelfth Population Census Conference
Participant List
Page 7

Ms. Liu Wenyan
China Marketing Director
DRS Data and Research Services PLC
Sunrise Parkway
Linford Wood
Milton Keynes MK14 6LR
United Kingdom

Dr. John Morgan
Manager
Enterprise Marketing
IBM World Trade/Asia-South Pacific Area
36 Floor, Connaught Center
1 Connaught Place
Hong Kong

East-West Center

Dr. Lee-Jay Cho
Director, East-West Population Institute and
Vice-President, East-West Center
1777 East-West Road
Honolulu, Hawaii 96848
U.S.A.

Dr. Fred Arnold
Research Associate
East-West Population Institute
1777 East-West Road
Honolulu, Hawaii 96848
U.S.A.

Dr. Linda G. Martin
Research Associate
East-West Population Institute
1777 East-West Road
Honolulu, Hawaii 96848
U.S.A.

Mr. Chi-Hsien Tuan
Research Associate
East-West Population Institute
1777 East-West Road
Honolulu, Hawaii 96848
U.S.A.

Twelfth Population Census Conference
Participant List
Page 8

Ms. Sandra E. Ward
Senior Editor
East-West Population Institute
1777 East-West Road
Honolulu, Hawaii 96848
U.S.A.

Ms. Valerie C. Wong
Program Officer
East-West Population Institute
1777 East-West Road
Honolulu, Hawaii 96848
U.S.A.

International Travel Service

Ms. Iris Suzuki Stickley
Travel Consultant
International Travel Service
220 S. King Street, Suite 1025
Honolulu, Hawaii 96813
U.S.A.

2. 會 議 日 程

State Statistical Bureau
Beijing

East-West Population Institute
Honolulu

TWELFTH POPULATION CENSUS CONFERENCE

September 5-9, 1988

Beijing, China

AGENDA

Monday, September 5, 1988

Morning Session (9:00 a.m.-12:00 noon)

OPENING CEREMONY

Chair: Sun Jingxin

Opening Remarks: Zhang Sai, Director-General, SSB, China

Opening Remarks: Lee-Jay Cho, Vice President, EWC

Conference Arrangements, Du Weiqun, SSB

Valerie C. Wong, EWPI

GROUP PHOTOGRAPH SESSION

Coffee Break

THE 1990 ROUND OF CENSUSES: THE USE OF COMPUTER TECHNOLOGY

Chair: V.S.Verma

Automation and Data Processing of the Hong Kong 1991

Population Census: Some Issues and Opportunities

(Joseph Man-Kong Lee)

The Use of Computerised Technology in the 1991 New Zealand

Census of Population and Dwellings (Michael A. Moore)

Afternoon Session (1:30 p.m.-5:00 p.m.)

THE 1990 ROUND OF CENSUSES: THE USE OF COMPUTER TECHNOLOGY

(Continued)

A New Processing System for Australia's 1991 Census of

Population and Housing (Mike Giles)

PLANS FOR THE 1990 ROUND OF CENSUSES

Chair: Mike Giles

Approach to the 1990-1991 Census of India (V. S. Verma)

Status Report on the 1990 Census of Population and Housing
of the U.S. (Susan Miskura)

Some Policy Considerations for Planning of the 1991
Population and Housing Census of Pakistan
(Muhammad Aslam Chaudhry)

The Census in Papua New Guinea (Nick Suvulo)

Planning for the 1990 Population and Housing Census of
Thailand (Chintana Pejaranonda)

Dinner Reception hosted by the State Statistical Bureau (leave
hotel at 5:45 p.m.)

Tuesday, September 6, 1988

Morning Session (9:00 a.m.-12:00 noon)

CHINA'S POPULATION CENSUSES AND SURVEYS

Chair: Fred Arnold

The 1990 Population Census in China: Preparations and Basic
Ideas (Sun Jingxin)

Exploitation and Utilization of China's 1982 Population
Census Data (Li Chengrui)

The One Percent Population Survey in 1987: Characteristics
and Preliminary Analysis (Shen Yimin)

The In-Depth Fertility Survey in China (Shen Qiuhua)

Afternoon Session (1:30 p.m.-5:00 p.m.)

DEMOGRAPHIC ESTIMATION FROM CENSUS DATA

Chair: B. R. Regmi

Measurement of Economic Activity and Migration
(Raja B. M. Korale)

Urbanization Index Based on the Census of Population and
Housing Results (Nelia Marquez)

Census Data for the Estimation of Mortality (William Brass)

Trends and Differentials in Early Age Mortality: Evidence
from the Post-Census Survey (Shahadat Hossain)

Dinner Reception hosted by the East-West Population Institute
(leave hotel at 5:30 p.m.)

Wednesday, September 7, 1988

FIELD TRIP

Alternative 1: (7:45 a.m.-about 4:00 p.m.)

Palace Museum and Great Wall

Alternative 2: (8:00 a.m.-about 4:00 p.m.)

Tanzhe Temple in the morning and Yuanmingyuan Park in the
afternoon

Evening shopping at the Friendship Store (leave hotel at 6:30 p.m.)

Thursday, September 8, 1988

Morning Session (9:00 a.m.-12:00 noon)

DEMONSTRATION OF COMPUTER EQUIPMENT AND SOFTWARE

Afternoon Session (1:30 p.m.-2:40 p.m.)

ADVANCED TECHNOLOGY FOR COLLECTING, PROCESSING AND
DISSEMINATING CENSUS DATA

Chair: Yuki Miura

The Use of Optical Mark Reading Systems in the Census
Process (Wilbert Riley)

The Use of Distributed, Low-cost OMR for Census Data
Collection (Stephen Stewart)

Group Meeting with State Councillor Li Tieying at the Great
Hall of the People (leave hotel at 2:45 p.m.)

Dinner Reception hosted by DRS and NCS

Friday, September 9, 1988

Morning Session (9:00 a.m.-12:00 noon)

ADVANCED TECHNOLOGY FOR COLLECTING, PROCESSING AND

DISSEMINATING CENSUS DATA (Continued)

Appropriate Technologies for the Economic and Extensive
Dissemination and Use of Census Data (Jack Massey)

The Use of Microcomputers for Census Processing in
Developing Countries (Glenn Ferri)

SPECIAL ISSUES IN CONDUCTING CENSUSES

Chair: Daniel Melnick

The Use of Sampling in the Census (Vijay Verma)

The Quality of Census Data (Yuki Miura)

Determining the Questions to be Asked in the 1991 Census of
Canada (D. Bruce Petrie)

Proposed Measures for Improving Coverage of the 1990
Population and Housing Census of Malaysia (paper
submitted by Khoo Teik Huat)

A Note on the Methodology of the 1990 Population Census of
Indonesia (paper submitted by Azwar Rasjid)

Afternoon Session (1:30 p.m.-5:00 p.m.)

MID-DECADE CENSUSES

Chair: R.B.M. Korale

Final Results of the 1985 Population and Housing Census
(Kang-Woo Lee)

Processing of Data from the 1986 Population Census of Fiji
(Naibuka Navunisaravi)

Special Features of the 1985 and 1990 Population Censuses of
Japan (Akihiko Ito)

FUTURE REGIONAL COOPERATION

Chair: Lee-Jay Cho

Presentation on the Asian and Pacific Population Forum
(Sandra E. Ward)

CLOSING CEREMONY

Chair: Fred Arold

Closing Remarks: Lee-Jay Cho, Vice President, EWC

Closing Remarks: Sun Jingxin, Deputy Director-General, SSB

Dinner Reception hosted by IBM

3. 韓國代表의 基調演說文

Country Statement for the Twelfth
Population Census Conference

September 5 ~ 9, 1988

Kang - Woo Lee
Director - General
National Bureau of Statistics
Economic Planning Board
Seoul
Republic of Korea

Ladies, Gentlemen and Distinguished Participants:

I would like to express my sincere appreciation to the Population Institute of East West Center and to the State Statistical Bureau of the People's Republic of China for their kind invitation for me to participate in this conference. It is a great pleasure to have this opportunity to present to you the results of our Population and Housing Census, with particular emphasis placed on the changing characteristics of the population, households and housing units in the Republic of Korea.

Before I proceed with the detailed results of the census, please allow me to briefly describe our census procedures. The most recent census in Korea was conducted as of November 1, 1985. It was the thirteenth population census taken since 1925 and the fifth housing census since 1960.

The National Bureau of Statistics, a semi-autonomous organization under the Economic Planning Board of the Korean Government, has the overall responsibility for the planning, execution, tabulation and publication of the census.

In order to comprehensively enumerate the entire population in a limited amount of time, we divide the country into enumeration districts using base maps obtained from the National Geographical Institute. The total number of general EDs in the 1985 census was 154,082, with each ED consisting of 60 households on average.

To fulfill the objective of assigning two EDs to each enumerator, approximately 85,000 enumerators were hired on a temporary basis, for 13 days. Among them, 27 per cent were university students and 40 per cent were representatives of neighborhood associations in urban areas and heads of village communities in rural areas. Meanwhile, approximately 40,000 local government officials were appointed as temporary supervisors during the census period.

In the 1985 census, all survey items were covered by the complete enumeration method. The number of items covered were 30, and the population was counted based on the de-jure concept.

As for data preparation, particularly manual editing and coding, 300 temporary female workers with a minimum a high school diploma were recruited for one year. Then computer editing seven times and tabulation were accomplished by the National Bureau of Statistics

Preliminary counts of population, households and housing units by minor administrative units were released in March of 1986. An advance report on the basic characteristics of the population, based on a two percent sample of households, was also published, in November, 1986. Then, final reports for the entire country, along with special reports on internal migration and fertility based on the full count, were published last December 1987.

These census results provide up-to-date benchmark information on the socio-economic and demographic characteristics as well as the living conditions of the population. As we focus on some important characteristics, we discover the several interesting results.

The population growth rate in Korea tended to show a gradual and steady downward trend. Since 1984, the annual growth rate has slowed to a mere one per cent, due to the rapid decline in fertility. The crude birth rate dropped to 16 per thousand population in 1987 from 43 thousand population in 1960.

Consequently, the population of Korea was 40.4 million in 1985, providing a population density of 408 persons per square kilometer. In absolute numbers, our population size ranked 23rd in the world. We also witnessed rapid changes in our socio-economic and demographic characteristics.

The proportion of population that had a higher than middle school education among the population aged 6 years and over increased to 47 per cent for males and 37.3 per cent for females in 1985, a tremendous leap from the 16.5 per cent and 5.3 per cent recorded in 1960, respectively. Meanwhile, the proportion of the population living in the urban area jumped to 65.4 per cent in 1985 from 28 per cent in 1960, mainly due to the migration to urban areas.

There was also a rapid change in the occupational structure. The proportion of people aged 15 years and over employed in the agricultural sector has continuously decreased. It declined from 46 per cent for males and 60 per cent for females in 1970 to 24 per cent and 35 per cent, respectively, in 1986. To the contrary, the percentage of people employed in non agricultural sectors showed an increasing trend.

The census also revealed that the median age of the population increased from 18.7 years in 1960 to 24.5 years in 1985, with the proportion of population aged 65 years and over increasing from 2.9 per cent to 4.3 per cent, clearly showing an aging of the population.

With regard to the household size and composition, it was observed that the rate of increase in the number of households was higher than that of the total population. The annual growth rate of households was 3.2 per cent, compared to the 1.9 per cent for population growth, over the last 25 years. The analysis of households by family type revealed structural changes in the composition of homes.

The proportion of total households comprising three generations or more declined to 14.8 per cent in 1985 from 25.5 per cent in 1966. This implies that the general structure of the Korean family is becoming more typified by the nuclear family rather than the traditional extended family.

As for the housing condition, the housing supply rate for the entire country showed a declining trend from 82.5 per cent in 1960 to 71.7 per cent in 1985. The situation was found to be more serious in the urban areas where the rate was only 59.4 per cent. However, the rate in the rural areas showed an increasing trend from 88.9 per cent in 1960 to 95.5 per cent in 1985.

As mentioned earlier, there are some distinctive features in the 1985 Population and Housing Census. We adopted the complete enumeration method, abandoning the sample survey method. Also, a number of new items were added, such as the religion and the average floor space of housing units.

Meanwhile, workloads for enumerators were markedly reduced from 150 households over eight days during the 1980 census to 120 households over ten days in 1985. I would also like to make mention of the fact that we waged a strong publicity campaign through the mass media to promote interest and cooperation among the public.

However, as we are all well aware, even with a greatly increased budget and manpower, an accurate count of the entire population and the determination of its characteristics within a limited time span are extremely difficult to calculate. Any census conducted in an ever mobile society is usually subjected to many errors, including coverage and content inaccuracies.

With the experience obtained in the last census, we are now preparing for the 1990 census with the following considerations in mind.

First, we plan to introduce the Optical Mark Reading system, or OMR system, to speed up data processing as well as to reduce coding and key punching errors. Our experience with this system on a survey of 60,000 households in 1987 was encouraging, and accordingly, a few pretests will be carried out on a trial basis for the 1990 census. Second, in order to meet the demands of diverse users, we will carry out an in-depth analysis of the census results, perhaps jointly with universities and research institutes.

Third, we believe additional effort should be made to recruit more qualified enumerators. This would mean that more funds would be required for the enumerators' compensation, while recruiting must be more carefully planned.

Fourth, audio-visual aids will be utilized to improve the effectiveness of the training materials, and fifth, the introduction of a mesh code for each enumeration district will be made in order to facilitate the preparation of various statistical geographical maps.

In closing, I wish once again extend my gratitude for this opportunity provided me here today. It was indeed a pleasure and an honor to speak before you. I look forward to maintaining our relationship of close cooperation with the Institute and also with the statistical bureaus of all countries represented here at this most meaningful conference.

Thank you.

4. 韓國의 ~~경제~~ ~~개발~~ ~~정책~~
~~세~~ ~~정~~ ~~비~~ ~~교~~ ~~류~~ ~~수~~ ~~입~~ ~~출~~ ~~구~~ ~~분~~ ~~배~~

Twelfth Population Census Conference
September 5-9, 1988
Beijing, People's Republic of China

Final Results of the 1985
Population and Housing Census
in the Republic of Korea.

National Bureau of Statistics
Economic Planning Board
Seoul
Republic of Korea

CONTENTS

I. Introduction	45
II. Trends in Population Size	47
III. Changes in Characteristics of Korean Population	50
3.1 Education	50
3.2 Economic Activity Status and Occupational Structure	52
3.3 Urbanization	53
3.4 Child Mortality	54
3.5 Marriage Pattern of Women	55
IV. Trends in Fertility	58
4.1 Changes in CBR and TFR	58
4.2 Age-specific and Marital Fertility Rates	60
4.3 Births by Birth Order	61
4.4 Mean Number of Children Ever Born	62
V. Household Sizes and Composition.....	64
VI. Housing Conditions	66
Tables and Appendix	67

I . Introduction

It is a well established practice in Korea to undertake a full-scale population and housing census every 5 years. The most recent census, the 1985 Population and Housing Census, was carried out as of 1 November 1985. This was the thirteenth census of population and the fifth census of housing in the series of census taking.

The National Bureau of Statistics, which is a semi-autonomous organization under the Economic Planning Board of the Korean government, has the overall responsibility for the planning, execution, tabulation and publication of the census.

Legal basis for the census is provided by the Statistics Law in which population and housing censuses are named as Government Designated Statistics Numbers 1 and 2. Moreover, a provision of the Census Decree of Korea stipulates that a full-scale census is to be taken in years ending in zero and a simplified census in years ending in five. However, the distinction between the full-scale census and the simplified one disappeared in recent years, as the scope of the simplified census has expanded to meet the ever-increasing demand for census information.

In order to enumerate all population in a limited time, the country was divided into enumeration districts(EDs) using census base maps prepared by the National Bureau of Statistics. The total number of general EDs in the 1985 census was 152,368 with each ED consisting of 60 households on average.

With a view to assign two EDs to one enumerator, about 85,000 enumerators were employed on a temporary basis for the census. Among them about 27 per cent were university students and 40 per cent were from heads of quasiadministrative neighbourhood units in urban areas and heads of village communities in rural areas. Meanwhile, approximately 40,000 census officials from the local administrative hierarchy worked as supervisors.

In the 1985 census, all survey items were covered by the complete enumeration method. The number of items covered were 30 items. Among them, 16 items were population characteristics such as name, age, sex, relationship to household head, religion, place of birth, place of residence one year ago, place of residence five years ago, school attendance, type of economic activity industry, occupation, marital status, number of children surviving and children dead. The other 14 items housing characteristics such as type of living quarters, type of household, tenure, etc.. Compared with the items in the 1980 census, the items on the origin of the family name, religion and floor area of a housing unit are newly appeared in the 1985 census. The population was counted on the basis of the de-jure concept.

As for the data preparation, particularly for manual editing and coding about 300 female workers with the minimum qualification of high school diplomas were recruited on a temporary basis for one year. However, data entry on to magnetic disks for complete questionnaires was executed on a contract basis with private business data centers.

Preliminary counts of population, households and housing units by minor administrative units were released in March 1986. An Advance Report on the basis of two per cent sample of households was also published in November 1986. Then, a Report for the Whole Country, Provincial Reports for 13 provinces and Special Reports for Internal Migration and Fertility on the basis of the entire count tabulation were published in December 1987.

These census results have provided us basic data for establishing various range of national policies, monitoring and evaluating the development programs. The census also provides the most extensive and concrete information on the characteristics of the population and living conditions of all the people as well as on size, structure and growth of population.

Thus, fundamental changes in its social and economic structures occurred during the last two and half decades are examined in this paper on the basis of census results.

II. Trends in Population Size

Korea is located among the major historical powers of Asia, and has on several occasions suffered from foreign invasion and domination. For instance there was the Imperial Japan's forcible annexation in 1910 lasting 36 years and the Korean War in 1950 bringing about the tragic division of the country. Nevertheless, it has succeeded in maintaining and preserving its language, culturally and linguistically homogenous.

The total area of the Korean peninsular is comparable to that of the British Isles, Rumania or New Zealand, which covers around 220,840 square Km. Since the Korean war, the Republic of Korea(South Korea) was established in the southern part of the peninsular occupying an area of 99,000 square Km, which is nearly as large as that of Jordan, Netherlands or Hungary. However, the major portion of the country is mountains(around 70 per cent) which makes its in this respect comparable to that of Finland or Sweden. Only 23.5 per cent of the land is arable, with plains mostly located in the west.

According to the released results of the 1985 population census, the total population of Korea is 40.4 millions giving a total density of 408 persons per square Km. In general, its population size ranks 23rd among the world, and stands 3rd in the world with respect to population density excluding those city countries like Hongkong and Singapore.

Until recently, Korea was predominantly an agrarian country characterized by intensive use of human labor. In the past two and half decades, however, the country has experienced a profound change which was significantly reflected in its socio-economic and demographic characteristics.

The process of very rapid growth and change was made possible through

the government's policy with a will to overcome the wartime destruction of assets. Since the national economic development plan was launched in 1962, efforts have been made for the take-off from traditional agrarian society to modern industrialized country emphasizing some leading sectors such as light industries in 1960s and heavy industries in 1970s. Thus, the economy has grown at an average annual rate of 8.3 per cent during the last 20 years.

During the enforcement of the development plans, the population factor (human resources) in addition to agriculture, technology, capital formation and foreign exchange expansion has received a special attention. The trend in population size of Korea is shown in Table 1.

In a modern sense, the first Korean census was undertaken in 1925. Subsequently, until the end of World War II, five censuses covering all of Korea were conducted almost every five years. Since the establishment of the Republic of Korea in 1948, eight censuses have been conducted. The partition of the peninsula in 1945 made it difficult to follow the trend in the total Korean population. Thus, while the data for the pre-1945 period relate to the entire Korean peninsula, the data for the post-1945 period relate only to South Korea.

The annual rate of population growth for the intercensal periods indicates that the country's population suffered a great fluctuation due to a number of factors such as the homecoming rush of overseas Koreans after the 1945 liberation, the influx of in-migration of war refugees from North Korea, and the post-war baby-boom and the rapid economic development in the recent years.

Prior to 1945, it is seen that the rate of population growth was slow but ununiform largely due to outmigration to Manchuria and Japan. In addition, high mortality which prevailed during that period had undoubtedly accounted for the low rate of natural increase of the Korean population.

Between 1944 and 1949, the population of South Korea increased at an

extremely high annual growth rate of about 4.7 per cent. Such abnormal growth rate is believed to be resulted from the return of Korean to the homeland from an overseas following the liberation of the country in 1945. It was reported from UN analysis (1959) that out of the total increase of 4,200 thousands during this intercensal period, the repatriates from abroad were 1,220 thousands and the refugees from North Korea were 1,835 thousands respectively. The proportion which was attributed to the natural increase of the population marked only one quarter of the total increase.

In contrast, the 1955 census revealed that the annual rate of population growth between 1949 and 1955 was only 1.2 per cent. This low rate of population growth might have resulted from the disruption of the Korean War which claimed a large number of war casualties and, to a lesser extent, from temporary drop in the birth rate. According to the UN analysis (1959), it was estimated that losses also an equivalent number of gain of refugees from the north (996 thousands) during the war time. This fact leads to the conclusion that the slow growth rate was largely due to the low birth rate in this period.

The 1960 and 1966 census showed a relatively high growth rate for the corresponding intercensal periods, at about 2.5 per cent per annum. The major reason for this reverse trend may be attributed to the post-war baby-boom coupled with a significant decline in the death rate due to improved public health measures and better medical facilities.

Since that time, the rate of population growth tended to show a gradual and steady downward trend. In recent years, the growth rate was halted and stands at around 1.0 per cent per annum. In the mean time, the crude birth rate dropped to 16 per thousand population. It suggests that such a decline has taken place at a much faster pace and in a very short duration of time in contrast to what had happened in western societies.

III. Changes in Characteristics of Korean Population

3.1 Education

It was reported that human resources represent the main contributing factor in the process of rapid socio-economic development in Korea.

After the complete destruction of assets resulting from the Korean War, the economic policy has been dedicated to rapid growth in the non-traditional sectors, relying heavily on external capital and technology. Accordingly, government efforts have been directed toward increasing the size of qualified labor force through rapid expansion of the basic and technical education.

Besides the importance outlined as above, many studies have demonstrated the important role of education in the modernization process and it has resulted in lowering the level of fertility. Thus, information relating to the educational level of the population has significant bearing on many aspects that account for sustaining fertility differences. Accordingly, it is postulated that the woman which has high level of educational attainment should be characterized with smaller number of children born.

The educational system of Korea was radically transformed immediately after the Second World War. A new system based on the United States experience was established by the Education Law of 1949. Today as of 1987, the regular school system consists of six years of primary schooling, three years of middle school, three years of high school, two-years of junior college and four to six years of college or university education.

Table 2 presents enrollment ratios in the various levels of the formal education system from 1966. In case of primary school enrollments, it was not changed during the last two decades standing at 100 percent. It is true since the Education Law specifies primary education as compulsory and free.

However, enrollment ratios in other levels of education show an increasing

trend both males and females. Enrollment ratios in middle school level increased from 50.9 per cent for males and 33 per cent for females in 1966 to 99.4 per cent and 98.7 per cent respectively in 1985, implying that all middle school-going-aged population today enrolls at middle school. Enrollment ratios in high school also increased rapidly from 35 per cent for males and 20 per cent for females in 1966 per cent to 74.4 per cent and 62.2 per cent respectively in 1985.

The most significant change was witnessed in enrollments in colleges and more level. It increased from 14 per cent for males and 5.1 per cent for females in 1966 to 46.6 per cent and 26.4 per cent respectively in 1985.

Corresponding to the change in enrollment ratios outlined as above, the level of education attained by the Korean population aged 6 years and over have drastically changed during the last two decades, as is evident from Table 3.

It is seen from the table that the proportion of those who attained the primary school level only and those who did not go to school was reduced to 13.4 per cent and 6.2 per cent respectively for males as of 1985 from 27 per cent and 30 per cent in 1960. The corresponding proportion for females is 19.8 per cent and 15.2 per cent as of 1985 from 25.1 per cent and 50.3 per cent in 1960.

On the contrary, the proportion of educated population with more than primary school level rapidly increased from 16.5 per cent for males and 5.3 for females in 1960 to 47 per cent for males and 37.3 per cent for females in 1985 respectively.

Although there has been a rapid growth in those attending school in term of the absolute number, the proportion of it among population aged 6 years and over did not show much variation over time taking about 34 per cent for males and 29 per cent for females.

3.2 Economic Activity Status & Occupation Structure

Economically active population is defined as all persons 15 years and over who did any work for pay of profit or want to work and are capable of working. Thus, persons not economically active are those who are engaged in household duties in their own house, those who are attending school or those who are unable to work.

The total labor force in Korea was observed to increase from about 7.5 millions in 1960 to about 16 millions by 1986. The annual rate of growth was about 3 percent during the corresponding period which is higher than that of total population. To some extent, the size of the labor force was influenced by developments in the social and economic fields which might change the participation rates in certain groups.

In Korea, traditional oriental values restricting the role of women and limited employment opportunities had limited women from participating actively in economic and social activities. Since 1960, however, considerable changes have steadily occurred in the status and role of women in Korea.

Economically active population participation rates for both males and females among the population aged 15 and over are shown in table 4. Against the expectation, a rising trend in participation rates is not observed during 1970-86. The participation rates are rather stable for both sexes at 73 per cent for males and 38 per cent for females respectively in Korea. The reason may be either due to the problems in data capturing or due to the change in occupational structure. The lower age limit in the definition of economically active population increased from 14 years to 15 years from 1983.

The change in occupational structure is shown in the same table. The proportion of persons employed in professional, technical & related jobs and in clerical & related jobs shows an increasing trend. Meanwhile, those employed in agricultural works have continuously decreased. It decreased from 46 per

cent for males and 60 per cent for females in 1970 to 24 per cent and 35 per cent respectively in 1986, implying that the economy of Korea appears to be evolving towards a modern structure. It is, however, noteworthy to mention that females are still predominantly engaged in agricultural works which comprise about 35 per cent of total employment in 1986.

3.3 Urbanization

It is one of striking features that there has been a rapid urbanization in Korea. It is known that the rapid increase in urban areas in Korea has been mainly attributed to the following two factors.

Firstly, from 1955 to the mid-1960s adverse economic conditions and high population pressure that prevailed in the countryside caused the rapid growth of urban areas. Secondly, from the latter part of the 1960s the pattern of urbanization appears to be following the experience of the industrialized nations in the sense that urbanization has been proceeding in close association with the nation's and economic development.

Analysis by two components for urban growth as shown in table 5 reveals that fifty-two per cent of total urban population growth during the 1960-1966 period was attributable to net migration and reclassification of boundary with that of the 1966-70 period composing 75 per cent. Turning to the 1970s, and 1980s the percentage of the component of migration still recorded high. It was 58 per cent during the period of 1970-75, 52 per cent during the period of 1975-80 and 53 per cent during the period of 1980-85 respectively. It is thus said that net migration is still the most important determinant for urban population growth.

As a result, the urban population increased by about 3.8 times from 7 million in 1960 to 26 million in 1985. Meanwhile, the total population of Korea increased by 1.6 times and rural population decreased by 22 per cent for the

corresponding period.

Consequently the share of urban population to the total population increased from 28.0 per cent in 1960 to 41.1 per cent in 1970, and then, to 65.4 per cent in 1985. In particular, the size of the capital city of Seoul increased by about 4 times from 2.4 million in 1960 to 9.6 million in 1985. The more worsering matter is that a majority of urban population are concentrated in large cities. The 1985 population census results revealed that the proportion living in six large cities were 69.4 per cent out of total urban population.

Therefore, it is needless to mention that population concentration is regarded as a threat to national security and balanced regional development. In addition, such rapid growth in cities necessarily causes such induced urban problems as the shortages of housing, educational, transportation, and other facilities, as well as a variety of pollutions as nowadays witnessed in Korea.

If we mention about future course of urbanization in Korea, the pace of urban population growth will be downed although the absolute number continue to increase due to natural increase in cities. Especially, it is expected that the speed of population growth in Seoul and six large cities will be halted. It is, therefore, seen that population growth in medium and small cities will be accelerated for the time being

3.4 Child Mortality

The level of mortality in general and of infants in particular has been found in several studies to have a significant effect on fertility.

Usually, child mortality affects fertility through psychological, biological and /or social aspects. In other words, there will be volitional response (replacement effect and insurance effect), shortening of breast feeding and changes of norms and attitudes about family size, which are related to child death. Therefore, in countries characterized with a high child mortality couples tend to re-

produce more children than they actually want in order to insure against child death or to replace immediately the child who has died. It also represents an important index of socio-economic and health status of the society.

The different levels of child mortality over time in Korea are presented in Table 6. It is seen from this table that all indicators in child mortality showed a decreasing trend. The infant mortality rate decreased rapidly from 68 per 1000 live births in 1960 to 15 in 1987. This decreasing pattern was thus reflected in the decline of the proportion of children dead among children born from women aged 15-49. The proportion of children dead was only 2.7 per cent in 1987 compared to 12.4 per cent in 1966.

If we also look into data by women's side, although there is no time series data on the proportion of ever-married woman aged 15-49 who have experienced child death, the 1987 Special Demographic Survey revealed that only 5.3 per cent of married women have experienced child loss.

It is, thus, noteworthy that the level of child mortality in Korea stands now at a very low level. This fact can imply that Korea has succeeded in creating a sound health condition for children enough to generate a favoring climate conducive to changes in traditional norms and attitudes about family size.

3.5 Marriage Pattern of Women

Marriage influences the reproductive behavior of individuals and determines a society's fertility level in many ways. It determines what proportion of women ultimately engage in reproduction, when reproduction starts, and how long on average a woman is exposed to childbearing. Especially, the proportion of currently married women is significant in understanding the level and structure of fertility in Korea since most births occur within wedlock.

With reference to Korea, it was revealed that age at marriage had increased continuously since 1930 upto 1960 (singulate mean age at marriage for females

is 16.8 years in 1930 and 21.6 in 1960). Therefore, overall drop in fertility in this period prior to the introduction of modern birth control methods was mainly attributed to the decrease in the proportion married among potentially fertile women.

The increasing trend in age at marriage is also seen in years even after 1960. It increased by 3.1 years to 24.7 in 1985 from 21.6 in 1960. This current level can be comparable to that of China (24.1 years observed in a community of suburban Beijing as of 1982) and Japan (24.1 years as of 1975). At the same time, the age at which marriages begin (around 18.5 to 19.0 years) was found to be the highest among Asian and the Pacific countries.

The reasons underlying the increase in age at marriage in Korea are understandable due to the following considerations. The rapid increase in employment opportunities for unmarried female, the difficulties facing young men in securing employment and establishing a household have contributed to increase age at marriage in Korea to a very high level.

In addition, the requirements of military service to all eligible young men aged 20 for around 3 years made them postpone their marriages until completion of the military service. In a sense, this fact is important not only raising the age at marriage but also providing the opportunities even to illiterate men to have an access to a new value system, since recruitment socio-economic characteristics.

Such a change in age at marriage eventually brought about the change in proportion of currently married women of reproductive age as is shown in Table 7.

As would be expected, the proportion of married women decreased in the age groups below 30 and increased slightly at ages 40-49. The proportional decrease at younger ages is mostly explained by the rising age at marriage and the increase at older ages by the decreasing risk of mortality. It is thus seen

that the proportion of currently married women among those females aged 15-49 shows no variation in recent time due to the offset each other between the decrease at younger ages and the increase at older ages.

IV. Trends in Fertility

4.1 Changes in CBR and TFR

Three standard measures of fertility in Korea since 1960 were presented in the Table 8: crude birth rate, total fertility rate and estimated number of births. The data were derived from two sources. The data for 1960 to 1979 were from the results of population census using the so-called own children method, while the data for 1980 to 1987 were from the so-called Continuous Demographic Survey which are conducted every month by the National Bureau of Statistics.

As shown in this table, the estimated crude birth rate (CBR) suggests that there has been a downward trend except in 1979 to the extent that it reached a level of 16 per thousand population in 1987. With an already low level of mortality around 13 in 1960 and 6 in 1985, the rate of population increase has consequently dropped from 2.88 per cent during 1955-1960 to 1.0 per cent in 1985. However, the fertility decline in Korea was not uniform. It has some fluctuation of sharp decline or slow decline.

By periods, the crude birth rate for 1955-60 reached a level of around 40 with a peak point around in 1960, which may be attributed to the baby boom that resulted from the return of soldiers enlisted in the Korean war. However, these high birth rate which began in the mid-1950s lasted only until 1962 when it started a dramatic decline during the next 5 to 6 years to reach 32 in 1967. But during the subsequent 5 years until 1972, the trend showed a temporary slackening of the rate of decline, maybe due to the threshold effect.

After the stationary status between 1967 and 1972, the period 1972-76 experienced a resumption of the steep decline in the crude birth rate to reach 22 in 1976. Between 1976 and 1981 the fertility level continued to drop but at a much slower pace, reaching a CBR of 21 in 1983. Meanwhile, turning to the mid-1980s there is an evidence that another sharp decline began to start from 1982 to reach a CBR of 16 in 1987.

On the other hand, the estimated total fertility rate (TFR) also showed a pattern of decline comparable to that of the crude birth rate. The TFR was around 6 in 1960 and it decreased rapidly to reach the level of 4.5 in 1967. But for next 5 years to 1972, the level of 4.5 continued. Then, the TFR began to drop rapidly starting 1973 until 1976 to reach a TFR of 3 in 1976. As it were in the crude birth rate, the TFR continued the decline between 1976 and 1981 but a much slower pace, reaching a TFR of 2.6 in 1981.

Turning to the mid-1980s, another sharp decline began to start from 1982 to reach a TFR of 1.6 in 1987. This fact is very interesting in a sense that the level of a TFR of 1.6 is below replacement level. And it will be a question to encounter when and to what level the decline will be finished.

It is also noteworthy to mention about the fluctuation in the TFR during 1977-79 when it was 2.94 in 1977 plummeted to 2.74 in 1978 and again increased to 2.83 in 1979. The unexpected low fertility in 1978 is attributed to the horse years. Since there is a common belief that children especially born in the horse year are most likely to be unlucky in their future life, therefore, that year tended to show a lower level of fertility than that of the adjacent years because couples planned to have their births either before or after that years. In the lunar calendar system of Far-Eastern countries of Asia, every year has a symbol characterized by one of twelve animals (mouse, cow, tiger, rabbit, dragon, etc.) which is given once in every 12 year cycle. According to this system, the year of 1978 was given the symbol of a horse. The fear of having children who would be unfortunate in the future life encouraged women to become effective users of contraceptives which consequently resulted in the lower level of fertility in 1978 compared to that of 1979. This fact can substantiate the widespread use of birth control in Korean society and the importance of attitudes rather than the availability of contraceptives.

4.2 Age - specific and Marital Fertility Rates

The age specific fertility rates for the period of 1960 - 87 are shown in Table 9. As shown in this table, it reveals that there has been a continuous decline of fertility in all age groups during the last decade except in 1979. Looking at the fertility level of the various age groups reveals that the decline was of a negligible in the age category 15 - 19 years. However, fertility of females in the age group 20 - 24 years showed a significant decline.

The most important feature of fertility decline during this period was manifested in the crucial decline in the reproductive performance of females in the age category 25 - 29 years as well as 30 - 34 years. But if we look at the speed of decline for these three age groups, it was not uniformly declined over time. The speed of decline for these age groups showed the same pattern of decline as that of CBR and TFR, which can divide the overall period into five periods: 1960 - 67, 1967 - 72, 1973 - 76, 1977 - 81 and 1982 to present.

It is also noted from the table that the most fertile age group has always been those women aged 25 - 29 years. The second most fertile group was 30 - 34 until 1972; thereafter it was the 20 - 24 age group. This clearly shows that after 1972 child-bearing activities have tended to concentrate among younger age women. In other words, married women have begun to have their children and finish their childbearing sooner, mostly before reaching the age of 30 years. Thus it is the women aged 20 - 29 years who are responsible for the largest proportion of births occurring every year in recent time; 47.8 per cent in 1960; 58.3 per cent in 1972; 69.4 per cent in 1977; 76.7 per cent in 1980; 85.3 per cent 1987. If we consider the women aged 30 - 34 years, the proportion of births occurring among women aged 20 - 34 increases to 97.3 per cent in 1987 from 70.7 in 1960. This would indicate that in Korea overall fertility trends have by and large been determined by the reproductive behaviour of married women aged 20 to 34 years.

Meanwhile, fertility of females at the ages beyond 35 years tended to taper off quite rapidly after reaching a peak in the age interval 25 - 29 years. The proportion of births occurring among ages 35 and more is almost negligible taking only 2.1 percent as of 1987.

However, if we consider only marital fertility rate, which refers to rate for currently married women only the picture is somewhat different from that of age - specific birth rates as shown in Table 10. Except in the last intercensal period, birth rates for married women at ages 20 - 24 remained virtually constant and at ages 15 - 19 has been somewhat erratic as shown in the table. The trend in fertility at ages 25 - 29 has been gradual but slow. Birth rates for currently married women age 30 and over fell steadily throughout the period.

The reason birth rates at ages 20 - 24 fell for all women but remained relatively constant for currently married women is that age - specific proportion married fell at these ages. The proportion currently married fell substantially at ages 20 - 24, and moderately at ages 25 - 29. Also the reason birth rates at ages 15 - 19 has been erratic is that the proportions married at these ages are too small.

4.3 Births by Birth Order

Another dimension for analyzing fertility, in addition to the age of mother, is the order of birth of the child. Order of birth refers to the number of children born alive to the mother, including the present child. The simplest way of analyzing births classified by order of birth consists of calculating the proportional distribution of the births by order. Such a percentage distribution of births by order is much less affected by under registration of births than are rates and can be computed without use of population data.

The percentage distribution of births in a year by birth order for the period of 1970 - 87 is shown in Table 11. It is first noted from this table that

there has been a change in the structure of birth orders. The proportion of births for parity 3 and over have recorded continuous and substantial decline roughly corresponding to the declines in fertility rates of women aged 30 years and over. In the case of parity 1, the proportion of births has been rapidly increasing to reach 54 percent in 1987. In the case of parity 2, the proportion of births for this parity has also shown a continuous increase but at a lower rate than that of parity 1. Thus, the proportion of births for parity 1 and 2 accounts for the majority of total births born in a year to reach 92 percent. It is, therefore, clear that the decline in the birth rate has largely been due to a decline in the proportion of births of higher orders

4.4 Mean Number of Children Ever Born

Information on the mean number of children ever-born per ever-married woman collected through population censuses is presented in Table 12. Although the number of children ever-born does not tell us anything about current fertility because we do not know when each child was born, these data seem to present a reasonable indicator of the prevailing pattern of fertility especially in a stable population

It is from the table that between 1966 and 1987 there has been a steady decrease in the average number of children ever-born from 3.8 in 1966 to 2.3 in 1987 among married women aged 15-49. The analysis of the mean number of children ever born by age on mother shows the different rate of decrease. The declining rate is higher among older aged women and is almost stable in the younger women aged 15-24. This implies that fertility behavior of younger married women has not changed markedly irrespective of time change.

It is, however, worthwhile to mention that the decline of cohort fertility has been less dramatic than that of period fertility. For cohorts of women who reached ages 45-49 in 1987, the peak childbearing years below age 30 occurred before 1970, when marital fertility was still very high. Thus their comple-

ted family size at ages 45 - 49 does not show nearly as dramatic a decline as the TFR based on the current age - specific birth rates. One could say that the children ever born figures show a great deal of inertia, since they embody, for older women, a good deal of high - fertility experience that occurred prior to the onset of the marital fertility decline in the 1960s and 1970s.

Thus, women who were aged 45-49 years that were those deemed to have completed their childbearing, had still 3.7 children even in 1987. This fact can suggest that women in younger ages who did not complete their childbearing at present is important in determining the future level of fertility.

V. Household Sizes and Composition

Human activities are by and large carried out in groups and the primary human group is the family or the household. It is also through the family that the population is perpetuated and cultural continuity is maintained from one generation to the next. In addition, the household tends to occupy a single housing unit and is therefore the most relevant demographic concept for use in analysis of housing trends and planning for future needs. The family or the household, and not the individual, are the primary unit of consumption used in various marketing and the cost of living studies.

It is observed from Table 13 that the total number of households in Korea has increased from 4,378 thousand in 1960 to 9,598 thousand in 1985, or by about 219 per cent in 25 years. The intercensal average annual rates of increase in the number of households grew faster than that of total population especially since 1966. That is to say, it is observed that the annual growth rate of the households recorded 3.2 per cent compared to 1.95 per cent of the total population on an average during the last 25 years.

This may be explained largely by a higher rate of growth of the marriageable population, a very high rate of urbanization and a higher rate of increase of single member households, all of which accelerated the rate of household formation during the period. In addition to the high growth rate of household formation, it is also found that there was a declining trend in the average size of households. It decreased from 5.56 persons in 1960 to 4.16 persons in 1985.

The analysis of households by family type also revealed the structural changes of the composition. In the Korean population censuses since 1960, the ordinary households were classified by the number of generations as shown in Table 14.

It is observed in this table that the proportion of one generation households showed a gradual increasing trend from 5.5 per cent in 1966 to 9.6 per cent

in 1985. Meanwhile, the two generation households formed about 68 per cent of all households showing no variations in different years. The three generation households, though constituting the second largest proportion, recorded a decline in their relative share from 23.0 percent in 1966 to 14.4 per cent in 1985. In case of four generation households, it also recorded a decline to reach only 0.4 per cent in 1985 from 2.5 per cent in 1966. On the other hand, there is another striking feature. There was a rapid increase in the absolute number as well as in the proportion of one person households. The annual growth rate of one person households marked an annual growth rate of 7.9 per cent during the last 25 years to share 7 per cent out of total households in 1985 from 2.3 percent in 1960.

It may be concluded from the above analysis that the general structure of the Korean family is typified by the simple conjugal family rather than the traditional extended family today. In other words, the self-sufficient economic unit that was identified with the traditional Korean family has ceased to exist.

V. Housing Conditions

The number of housing units in the country as a whole as well as in urban and rural areas in the period of 1960-85 is shown in Table 15. It is observed that there was an increase in the number of housing units in the country from 3,494 thousand in 1960 to 6,271 thousand in 1985. In other words, the number of housing units increased by about 2.4 per cent annually for the last 25 years. It is also noted that it was in the urban areas that the housing units showed a rapid increase during the period, while in the rural areas there was a decline.

The housing supply ratio is also shown in the same table. It has been estimated that the housing supply ratio for the whole country showed a declining trend from 82.5 per cent in 1960 to 71.7 per cent in 1985. The situation is found to be more serious in the urban areas where the housing supply ratio was only 59.4 per cent in 1985.

The serious housing situation in Korea has been caused by a number of factors. The major cause is that the gradual dissolution of the traditional large family system and the large-scale migration of the rural population to urban centers created an unusually high demand for housing, particularly in the urban areas. Also, the rising prices of land and construction materials in recent times is one of obstacles in supplying new housing units

Another aspect of the housing problem is the demand created newly by the increasing population. Given a certain backlog in housing, it is also necessary to provide the future increase in population with suitable housing if an increase in the backlog is to be prevented. The perennial requirements of houses for the future annual increments in population thus form another aspect of the problem.

TABLES and APPENDIX

1. Census Population & Estimated CBR and CDR : 1920-85
2. School Enrollment Ratios by Sex : 1966-85
3. Percentage Distribution of the Population aged 6 years & over by Educational Attainment and Sex : 1960-85
4. Economic Active Population Participation Rates and Percentage Distribution of Total Employed by Occupation : 1970-86
5. Indicators for Urbanization in Korea : 1960-85
6. Indicators for Child Death
7. Proportion of Currently Married Women
8. Number of Births, CBR and TFR : 1960-87
9. Age Specific Fertility Rates
10. Age Specific Marital Fertility Rates
11. Percentage Distribution of Births by Birth Order : 1970-87
12. Mean Number of Children Ever Born
13. Number of Households, Average Size and Annual Increase Rate : 1960-85
14. Percentage Distribution of Ordinary Households by Household Type : 1960-85
15. Households and Housing Stock

Appendix 1 : Censuses of Korea

Appendix 2 : Questionnaire for 1985 Population and Housing Census

Table 1. Census Population & Estimated CBR and CDR : 1920 - 85.

Census Year	Census Pop. ^{a)}	Annual Growth Rate (%)	Period	CBR (%)	CDR (%)	NI (%)
			1919 - 15 ^{b)}	45.3	33.7	11.6
			1916 - 20 ^{b)}	47.5	31.6	15.9
1925	19,020,030	1.44	1921 - 25 ^{b)}	48.0	29.5	18.5
1930	20,438,108	1.66	1926 - 30 ^{b)}	55.9	26.4	19.5
1935	22,208,102	1.17	1931 - 35 ^{b)}	45.5	23.3	22.2
1940	23,547,465		1936 - 40 ^{b)}	43.3	21.4	21.9
1944	25,133,352	1.41	1941 - 45 ^{b)}	42.1	19.5	22.6
	(15,944,000) ^{d)}					
1949	20,166,756	4.70	1946 - 50 ^{b)}	39.9	15.8	24.1
1955	21,502,386	1.20	1951 - 55 ^{b)}	41.0	14.3	26.7
1960	24,954,290	2.84	1951 - 60 ^{b)}	43.0	12.8	30.2
1966	29,159,640	2.67	1956 - 65 ^{c)}	37.0	10.0	27.0
			1966 - ^{c)}	31.9	8.6	23.3
1970	31,435,252	1.88	1970 ^{c)}	29.5	9.8	19.7
1975	34,678,992	1.96	1975 ^{c)}	24.6	7.3	17.3
1980	37,406,815	1.49	1980 ^{c)}	23.4	6.7	16.7
1985	40,432,397	1.57	1985 ^{c)}	19.7	6.2	13.5

Note : Data before 1945 are for all Korea ; data for subsequent years are for South Korea only.

Sources : a) Various Census Results

b) Choe, 1967 cited from Cho, 1982

c) National Bureau of Statistics, 1986

d) Republic of Korea only, Hong, 1978

Table 2 : School Enrollment Ratios by Sex : 1966 - 85

Unit : %

	Primary	Middle	High	College & more
1966				
Total	96.6	42.3	27.5	9.7
Male	98.1	50.9	35.0	14.0
Female	95.1	33.0	19.6	5.1
1970				
Total	102.4	56.1	30.5	10.1
Male	102.9	65.1	36.7	14.6
Female	101.9	46.5	24.1	5.5
1975				
Total	103.2	75.2	43.6	12.2
Male	103.0	80.8	51.1	17.5
Female	103.4	67.0	35.8	6.7
1980				
Total	101.0	94.6	68.5	16.4
Male	100.7	96.4	74.4	22.5
Female	101.4	92.6	62.2	10.0
1985				
Total	102.0	99.1	86.4	37.0
Male	101.7	99.4	88.0	46.6
Female	102.3	98.7	84.8	26.4

Note :The enrollment ratio is the proportion of students enrolled among school -age population in different levels of school.

Source : Various population census reports.

Table 3 : Percentage Distribution of the Population aged 6 years & over
by Educational Attainment and Sex : 1960 - 85

Unit : %

	1960	1966	1970	1975	1980	1985
Males						
Attending	26.7	30.6	33.2	33.5	33.7	33.4
Primary	26.9	27.8	25.8	24.2	18.0	13.4
Middle	14.0	17.7	11.3	14.1	14.4	13.6
High			10.7	12.4	20.1	26.5
College & more	2.5	4.1	4.2	4.7	5.4	6.9
Never attended	29.9	19.8	14.8	11.1	8.4	6.2
Females						
Attending	19.2	24.7	27.3	27.7	29.1	29.0
Primary	25.1	31.2	32.3	31.9	25.5	19.8
Middle	5.0	9.0	7.8	11.4	14.3	15.4
High			4.7	7.4	12.6	19.1
College & more	0.3	0.7	1.1	1.5	1.9	2.8
Never attended	50.3	34.4	26.8	20.0	16.6	15.2

Source : Various population census reports

Table 4 : Economic Active Population Participation Rates and Percentage
Distribution of Total Employed by Occupation : 1970 - 86

	1970 ¹⁾		1975 ¹⁾		1980 ¹⁾		1983 ²⁾		1986 ²⁾	
	M	F	M	F	M	F	M	F	M	F
Economically active population participation rates (%)	72.5	37.6	77.8	45.7	72.4	38.4	71.8	33.9	74.4	39.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1	3.8	2.1	4.1	2.1	5.2	3.5	5.7	4.2	6.3	5.0
2	1.4	0.1	1.3	0.1	1.6	0.04	1.3	0.06	1.4	0.04
3	7.6	2.8	8.5	4.0	10.0	8.6	10.2	9.3	12.2	9.9
4	10.5	9.6	11.1	9.5	12.3	11.6	13.4	17.3	13.7	17.0
5	4.5	10.8	4.6	9.3	5.4	9.9	6.1	12.8	6.1	14.2
6	46.4	59.8	43.0	53.8	32.5	46.4	29.7	39.8	23.8	35.2
7	25.7	14.8	27.4	16.2	32.9	19.9	33.6	16.6	36.5	18.6

- Note : 1) Professional, technical & related workers
2) Administrative & managerial workers
3) Clerical & related workers
4) Sales workers
5) Service workers
6) Agricultural & related workers
7) Production & related workers, transport equipment operators & laborers

Source : 1) NBOS, Population Census Reports for 1970, 1975 and 1980
2) NBOS, First and Second Employment Structure Survey Reports

Table 5 : Indicators for Urbanization in Korea ; 1960 - 85

(Number in thousand, %)

	1960	1966	1970	1975	1980	1985
1. Total pop.	24,989	29,193	31,466	34,707	37,436	40,467
2. Urban pop.	6,997	9,805	12,953	16,793	21,434	26,458
3. Seoul pop.	2,445	3,803	5,536	6,890	8,364	9,646
4. Six large cities pop.	5,231	7,328	10,062	12,567	15,592	18,354
2 / 1	28.0	33.6	41.2	48.4	57.3	65.4
3 / 1	9.8	13.0	17.6	19.9	22.3	23.8
3 / 2	35.0	38.8	42.7	41.0	39.0	36.5
4 / 1	20.9	25.1	32.0	36.2	41.7	45.4
4 / 2	74.8	74.7	77.7	74.8	72.7	69.4
Annual pop. growth rate						
1	2.70	1.89	1.98	1.50	1.57	
2	5.96	7.21	5.33	3.98	4.30	
3	7.87	9.84	4.47	3.89	2.89	
4	5.95	8.25	4.55	4.33	3.31	
Contribution by net migration including reclassification of boundary change for urban growth						
	52	75	58	52	53	

Source : Various census report

Table 6 : Indicators for Child Death

	IMR (%)	% of children dead among children born from women aged 15 - 49	% of married women aged 15 - 49 who experienced childloss
1960	67.7	—	—
1966	55.5	12.4	—
1970	45.4	10.7	—
1975	37.2	6.3	—
1980	30.5	3.9	—
1985	20.0	3.2	—
1987	15.0	2.7	5.3

Source : 1) Various census reports

2) 1987 Special Demographic Survey results

Table 7 : Proportion of Currently Married Women

Unit : %

	1960	1966	1970	1975	1980	1985
15 - 19	5.83	3.81	2.84	2.59	1.63	0.85
20 - 24	55.72	47.66	42.29	37.22	33.75	27.79
25 - 29	92.08	89.80	88.38	86.79	84.94	80.88
30 - 34	91.61	93.90	94.56	94.43	94.30	93.27
35 - 39	87.93	89.21	91.95	93.22	93.42	93.29
40 - 44	82.05	82.74	84.81	88.23	90.06	90.41
45 - 49	72.73	75.24	76.85	79.09	83.11	85.88
15 - 49	67.4	64.2	62.9	59.5	59.8	60.5
SMAM	21.6	22.8	23.3	23.7	24.1	24.7

Source : Various census reports

Table 8 : Number of Births, CBR and TFR ; 1960 - 87

	Estimated No. of Birth (000)	CBR (%)	T F R
1960 ²⁾	1,060	42.64	5.98
1961 ²⁾	1,103	42.83	6.02
1962	1,104	41.65	5.85
1963	1,140	41.82	5.89
1964	1,035	36.99	5.23
1965	1,028	35.81	4.95
1966	1,003	34.07	4.84
1967	956	31.73	4.50
1968	1,010	32.75	4.66
1969	994	31.68	4.54
1970	998	30.96	4.47
1971	1,028	31.25	4.54
1972	1,020	30.45	4.41
1973	953	27.94	4.01
1974	918	26.46	3.75
1975	847	24.02	3.33
1976	796	22.21	3.00
1977	808	22.19	2.94
1978	785	21.25	2.74
1979	851	22.67	2.83
1980	854	22.41	2.71
1981 ³⁾	855	22.50	2.64
1982	840	20.90	2.37
1983	791	19.80	2.09
1984	693	17.10	1.80
1985	674	16.40	1.70
1986	658	15.80	1.62
1987	651	15.50	1.55

Source : 1) Cho, The Determinants of Fertility in the Republic of Korea, 1982, p.38

2) For 1961 - 80, NBOS, The Levels and Trends of Fertility, 1984

3) For 1981 - 87, NBOS, Estimates obtained from civil registration and Continuous Demographic Survey, 1987

Table 9 : Age Specific Fertility Rates

	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1960	35.0	249.0	323.0	273.0	204.0	96.0	16.0
1961	36.0	247.8	330.3	276.0	203.0	95.9	13.6
1962	31.0	238.5	332.3	267.5	195.2	92.4	13.4
1963	28.1	244.2	338.7	273.1	191.0	89.6	12.1
1964	23.1	217.2	310.3	243.0	165.9	75.3	10.6
1965	21.5	212.0	309.0	225.7	146.4	64.5	10.3
1966	22.2	212.5	310.0	218.5	136.3	59.5	9.2
1967	19.8	190.0	302.7	208.7	119.8	51.8	8.2
1968	19.8	196.8	319.7	218.0	120.2	49.4	7.3
1969	18.2	187.6	313.8	216.0	119.2	47.1	6.5
1970	16.9	190.4	314.7	211.0	112.5	43.2	6.2
1971	18.1	202.7	320.7	215.3	108.6	38.1	4.6
1972	17.6	196.8	317.4	207.1	102.9	35.4	4.9
1973	16.3	187.2	300.3	178.6	86.1	29.5	4.1
1974	15.4	181.2	291.5	160.5	72.7	24.6	3.3
1975	13.7	163.2	267.9	139.9	59.3	19.4	2.7
1976	13.3	157.5	248.3	116.5	47.3	14.8	2.0
1977	12.8	154.8	253.4	113.7	39.8	11.9	1.7
1978	12.5	152.7	238.5	101.7	31.8	8.7	1.3
1979	12.0	161.4	255.8	101.0	28.1	6.9	1.1
	1						
1980	9.5	154.7	248.4	96.7	25.8	5.9	0.8
1981	8.2	167.3	244.7	84.6	27.2	5.2	0.7
1982	7.8	153.9	222.6	67.4	18.0	3.1	0.6
1983	7.5	134.9	220.4	51.5	18.6	4.2	-
1984	5.7	124.6	172.5	43.6	10.6	2.7	-
1985	5.5	115.5	169.0	42.0	8.9	1.7	-
1986	3.8	102.1	168.6	42.2	6.5	0.6	-
1987	2.4	102.6	161.8	37.3	6.0	0.5	-

Source : Same as table 8

Table 10 : Age Specific Marital Fertility Rates

	1960	1966	1970	1975	1980	1985
15-19	600	582.7	595.1	529.0	528.8	823.5
20-24	447	445.9	450.2	438.5	458.4	413.8
25-29	351	345.2	356.1	308.7	292.4	209.0
30-34	298	232.7	223.1	148.2	102.5	45.0
35-39	232	148.2	122.3	63.6	27.6	8.5
40-44	117	70.2	52.8	22.0	6.6	1.9
45-49	22	12.0	8.1	3.4	1.0	0.5

Note : The rates were calculated from ASFR shown in table 3.2 divided by proportions currently married from various census result

Table 11 : Percentage Distribution of Births by Birth Order ; 1970 - 87

Unit : %

	1	2	3	4	5	6 +
1970	25.0	21.8	18.8	14.3	9.6	10.3
1971	25.7	22.3	18.7	14.0	9.2	10.1
1972	26.7	23.7	18.6	13.2	8.6	9.2
1973	28.0	24.5	18.1	12.3	8.2	8.9
1974	31.2	25.5	17.8	11.1	7.0	7.4
1975	34.0	27.6	17.3	9.4	5.7	6.0
1976	34.5	29.3	17.4	8.8	4.9	5.1
1977	35.9	30.9	17.2	7.9	4.1	4.1
1978	38.5	31.7	16.8	7.1	3.3	3.2
1979	38.5	33.3	17.1	6.4	2.6	2.3
1980	39.4	31.6	16.8	6.7	2.9	2.7
1981	40.9	33.8	16.5	5.5	2.0	1.4
1982	41.8	35.7	14.8	4.8	1.7	1.1
1983	44.8	38.5	11.1	3.6	1.3	0.7
1984	49.5	38.1	8.4	2.6	0.9	0.5
1985	51.5	38.0	7.4	2.0	0.7	0.4
1986	52.9	37.8	6.8	1.7	0.6	0.2
1987	54.1	38.0	5.7	1.6	0.3	0.2

Note : The data for 1970 - 86 were derived from the vital registration system, while the data for 1987 were from the Continuous Demographic Survey

Table 12 : Mean Number of Children Ever Born

	1966 ¹⁾	1970 ¹⁾	1975 ¹⁾	1980 ¹⁾	1985 ¹⁾	1987 ²⁾
15-19	0.50	0.50	0.48	0.45	0.44	0.42
20-24	1.10	1.03	1.00	0.92	0.86	0.70
25-29	2.31	2.15	1.97	1.75	1.54	1.39
30-34	3.78	3.53	3.18	2.67	2.29	2.11
35-39	4.85	4.55	4.11	3.43	2.80	2.63
40-44	5.53	5.28	4.79	4.17	3.44	3.14
45-49	5.66	5.62	5.23	4.69	4.07	3.70
15-49	3.78	3.70	3.43	2.94	2.52	2.33

Source : 1) Various population census results

2) 1987 Special Demographic Survey

Table 13 : Number of Households, Average Size and Annual Increase Rate :
1960 - 1985

Year	No. of ordinary households	Average size	Total house - holds	Annual household growth rate (%)	Annual population growth rate (%)
1960 ^{a)}	4,362,953	5.56	4,377,973	2.76	2.71
1966 ^{b)}	5,121,610	5.43	5,132,910	3.35	1.90
1970 ^{c)}	5,792,983	5.18	5,856,901	2.89	1.98
1975	6,647,778	5.13	6,754,257	3.37	1.50
1980	7,969,201	4.62	7,992,968	3.72	1.57
1985	9,571,361	4.16	9,593,679		

Notes : a) Unknown households (1,534) were added

b) One person quasi households (64,580)
were included in the number of ordinary households

c) Estimated one person households (216,706) were added in the
number of ordinary households

Table 14 : Percentage Distribution of Ordinary Households by Household Type :
1960 - 1985

Household type	1960	1966	1970	1975	1980	1985
Total	100.0	100.0	100.0	100.0	100.0	100.0
One generation HHs	7.3	5.5	6.5	6.7	8.3	9.6
Married couple		4.4	5.2	4.8	6.0	7.1
Others		1.1	1.3	1.9	2.3	2.5
Two generation HHs	62.6	64.8	67.4	68.9	68.5	67.0
Couple with children		51.8	53.4	53.2	53.0	52.8
A parent with children		7.7	10.2	9.7	9.3	8.6
Couple with their parents or a parent		1.4	1.3	0.5	0.6	0.7
Couple with children & brothers or sisters		2.0	1.9	2.1	2.3	2.3
Others		1.9	0.6	3.4	3.3	2.3
Three generation HHs	26.3	23.0	21.2	19.2	16.5	14.4
Couple with children & their parents		6.6	4.9	2.0	1.9	1.9
Couple with children & a parent		12.4	11.9	8.5	7.9	7.2
Others		4.0	4.5	8.8	6.7	5.4
Four generation HHs	1.6	2.5	1.1	0.9	0.5	0.4
One person HHs	2.3	3.6	3.7	4.2	4.8	6.9
Non - relative HHs		0.8			1.5	1.7

Note : HHs is abbreviation of households

Table 15 : Households and Housing Stock

	1960	1970	1975	1980	1985
Whole					
Houses (A)	3,494	4,409	4,816	5,434	6,271
Households (B)	4,198	5,545	6,293	7,470	8,751
A / B (%)	82.5	79.5	76.5	72.8	71.7
Annual growth rate of A	2.35	1.78	2.44	2.91	
Annual growth rate of B	2.82	2.56	3.49	3.22	
Urban					
Houses (A)	805	1,442	1,851	2,526	3,433
Households (B)	1,209	2,354	3,125	4,362	5,779
A / B (%)	66.6	61.3	59.2	57.9	59.4
Annual growth rate of A	6.00	5.12	6.42	6.33	
Annual growth rate of B	6.89	5.83	6.90	5.79	
Rural					
Houses (A)	2,659	2,987	2,966	2,908	2,838
Households (B)	2,989	3,191	3,168	3,107	2,972
A / B (%)	88.9	93.6	93.6	93.6	95.5
Annual growth rate of A	1.17	-0.14	-0.39	-0.49	
Annual growth rate of B	0.66	-0.14	-0.39	-0.88	

- Note : 1) One person households and non - relative collective households were excluded in the calculation of households
- 2) Vacant housing units were included in the calculation of number of houses.

Ⅲ. 各國의 COUNTRY REPORT

1 . A NEW PROCESSING SYSTEM FOR
AUSTRALIA'S 1991 CENSUS

**EAST WEST POPULATION INSTITUTE
STATE STATISTICAL BUREAU OF CHINA**

**TWELFTH POPULATION CENSUS CONFERENCE
BEIJING, CHINA
5-9 SEPTEMBER 1988**

***A NEW PROCESSING SYSTEM FOR
AUSTRALIA'S 1991 CENSUS***

***Mike Giles
Australian Bureau of Statistics
Canberra, ACT, Australia***

Wed, 31 Aug 1988 A New Processing System for Australia's 1991	91
Government direction on reducing cost of the Census.	91
General strategy for reducing costs.	91
User Consultation.	92
The Input Processing System.	92
The 1986 System.	92
The need to redevelop the system.	93
Optical Mark Recognition.	93
Write-in responses.	94
Computer Assisted Coding (CAC).	95
CAC/Edit/Amend.	96
CAC.	96
Edit/Amend.	97
Quality Control.	97
Cost effectiveness.	98
Job Design.	98
The Output Processing System.	99

1: Government direction on reducing cost of the Census

In 1987 approval was obtained from the government to start planning for at least a minimum population census in 1991. The minimum census would collect only information on age, sex and place of usual residence for the purpose of providing reliable data for the population estimates required for electoral and other purposes.

At the same time, the Government directed that consultation with users about the content of the 1991 Census should proceed but on the basis that the cost of the census is to be a good deal less than the 1986 Census. The Treasurer, the minister responsible for the ABS, was to bring forward a further submission on content and cost of options.

2: General strategy for reducing costs

Following this decision, ABS published an information paper "1991 Census of Population and Housing: Preliminary ABS views on Content and Procedures." The paper proposed the following cost-saving measures:

- . use of respondent marked optical mark recognition (OMR) technology for data entry;
- . increased use of self-coding by respondents; and
- . computer assisted coding of responses requiring office coding.

However, the paper acknowledged that these measures would only go part of the way to reducing costs to the required level (ie about 20%). Two collection options were therefore also proposed.

Option A would entail a full enumeration of all topics included, but with a significant reduction in the number of topics compared to 1986. Data would be available for small areas and small groups in the population but, of course, only for those topics included.

Option B would entail a less severe reduction in the number of topics but some of these topics would only be asked of a sample of the population. Data would be available for small areas and small groups of the population, as for Option A, but only for the smaller number of core topics put to the whole population. Sample estimates for the other topics asked of a sample of the population would be available for large geographic areas and larger groups in the population.

For Option B a short form would be used which contained the smaller number of core topics going to two thirds of the population and a long form which contained the core topics plus the other topics going to the remaining one third of the population.

3: User Consultation

Users were asked to submit proposals to ABS on which of the two options it should pursue. More than 500 written responses were received and consultation took place with all the major users. Predictably perhaps, there was substantial opposition to cuts in the content of the Census and to any reduction in the amount of small area data available.

This put considerable pressure on the need to make as much savings as possible through more efficient census operations, even though this was never likely to allow a substantial extension of the census content.

The consultation and debate with users led to some reassessment of the data content of options A and B as put forward in the preliminary views. However, it was clear that if additional topics were to be considered, they had to have a low processing cost if they were to be added to Option A or to the core component of Option B. Moreover, additional topics with a significant processing cost could only be considered for the sample part of Option B. In the event, the need to accommodate some costly but high priority topics led to the recommendation that the sample be reduced from one in three to one in four.

The plans for the processing of the 1991 Census are currently proceeding on the assumption of the government accepting a recommendation for a modified option B.

4: The Input Processing System

The main change proposed for the 1991 processing (and the main subject of this presentation) relates to the input processing system. Here it is intended to use an optical mark recognition strategy for the data capture of the self coded responses and a computer assisted coding strategy for coding and capture of the remaining responses and for editing and amending.

This combination is considered to be the most cost effective option for input processing and reduces significantly the number of temporary staff required compared with the number needed for processing previous censuses. Both OMR and CAC have the potential to improve the quality of census data by reducing processing errors.

The strategy for implementation also provides early capture of some of the census data for use in timely revisions to population estimates and also early release of some preliminary census counts for those topics captured in the OMR phase.

4.1: The 1986 System

In 1986 the census forms were coded and then transcribed onto computer files by keying direct to disk and edited. When all errors detected by the editing system were corrected by batch processing of amendments keyed from

clerical modifications to edit reports, census data were reformatted to create the Final Unit Record File (FURF) used for output processing.

As for all previous censuses, 1986 input processing was very labour intensive, requiring 1050 staff years of effort. Coding was mostly manual and required 470 staff years; data capture, 325; editing, 95; and other operations and overheads, 160 staff years. Key entry was used for data capture in 1986 whereas previous censuses had used office marked OMR which required even higher numbers of temporary staff.

1986 coding processes involved matching of descriptions provided by respondents against printed indexes and clerically transcribing codes onto the census forms for later key entry of the data. A limited computer assisted coding system was however used successfully for coding occupations.

4.2: The need to redevelop the system

A major objective of the 1991 processing strategy is to automate processes as much as possible to reduce the temporary staff levels required and to reduce the extent of movement of forms through processes, with the aim of reducing the costs and overheads involved.

Manual coding has other inefficiencies besides high labour costs. Large volumes of printed indexes involved considerable printing costs. There were also inefficiencies in referencing these cumbersome volumes and significant costs in manually updating them. There were also clerical errors in the transcription of the codes to the census forms for subsequent key entry and errors in that key entry. Furthermore because of the effort needed in coding and the fact that the process was sequential, all the data was delayed in being captured. There was also extra effort needed in controlling document flow through the specialist coding processes.

4.3: Optical Mark Recognition

An alternative to key entry is the use of OMR whereby respondents answer as many questions as possible by marking the forms with machine readable responses.

To reap the full benefits of OMR, questions on the census form need to be designed to minimise the coding effort and maximise the data captured by this method. One option for topics such as religion and birthplace is to ask respondents to mark the appropriate box from a short list of the most common responses or to write a response in a sink category of "other, please specify". Only the small number of write-in responses to "other" then require office coding.

While it seems that this approach could reduce the coding effort for these

questions to possibly less than 10% of that currently incurred, the possibility of response biases needs further evaluation. These biases would occur if respondents attempt to choose a listed answer rather than write-in a response if their answer does not appear. The experience of some other countries also suggests there may be an adverse reaction to this approach from minority groups claiming that failure to list a specific response for them is discriminatory.

However, on the basis of the OMR testing undertaken so far, respondents completed the forms marginally better than the design used for key entry. Nor was there any adverse reaction to the self coding questions. An average of 6% of responses were not read by the OMR equipment used in the test. Of these, 1.5% were responses which were present but not detected by OMR. The remainder were not stated by respondents. It was also shown that the OMR was more reliable than key entry and gave improved data quality.

The test also showed that the readers could be operated by relatively unskilled staff. This also meant that the machines could more easily be operated on a shift basis. One significant bonus was the removal of the risk of RSI. The test highlighted the most critical factor in the success of using an OMR data capture method. This was the precision needed in printing the forms. The test showed that Australian manufacturers could meet the requirement within reasonable cost bounds.

The use of OMR will require a significant change to the general design of the census form. Each household and each person within the household will require a separate page. Because separate forms involve extra field costs, the present plan is for the pages to be stapled together during the enumeration phase and then to be split and batched prior to reading. The forms for each household will then be stapled again for the preliminary processing and CAC phases.

Passing the forms through the OMR reader aims to capture all the self-coding information. This generates a raw unit record file which is then transferred to the CAC equipment. Preliminary processing is then undertaken which involves the balancing of census collector record book counts against the number of person records and household records.

This is followed by the imputation of age, sex and marital status where there are non-responses to these topics. If invalid values are generated, these are flagged for attention during the CAC/edit/amend stage.

Two files are then generated. The preliminary data file is used to produce preliminary output of counts of self coding characteristics. A data file is also created for the CAC/edit/amend processing.

4.4: Write-in responses

There are several possible ways to process the write-in responses on the 1991 census forms. They could be office coded as in previous Australian censuses but with the codes marked onto reserved areas of the forms and then the forms read by OMR equipment. This system was used as early as the 1971 Malaysian Census.

Another alternative is the use of optical character recognition whereby respondents answer self coding questions by marking a response as with OMR and then writing block character responses in specified spaces. The whole form is then read by OCR equipment and all answers are then ready for further computer processing. The write in responses could be automatically coded by the computer with clerical intervention only for difficult cases. Automatic coding, as distinct from computer assisted coding, has been used by ABS in systems for processing data from administrative systems but further research and development would be required to use this approach for census questions. There does not appear to be any international experience with OCR coupled to automatic coding, although there is a lot of interest in both techniques.

Initial investigations of OCR technology indicated some considerable doubt on its ability to read with acceptable accuracy the extremely diverse range of written responses likely from the public. Current applications of OCR technology tend to involve the reading of characters marked under controlled conditions or reading of numeric information filled in under less controlled conditions. Performance in reading numeric information tends to be significantly better than reading alphabetic characters. Even if a small proportion of the characters provided on the census forms proved unreadable, the correct action needed would be costly. Further OCR equipment is much more expensive than OMR equipment and because the throughput rate tends to be lower, more equipment would be needed.

For these reasons, OCR was not considered to be a viable option for 1991. However, because of the potential for capturing written responses for automatic computer coding, the technology will be monitored for possible use in future censuses.

4.5: Computer Assisted Coding (CAC)

Another option is to use key in the alpha characters and to have the computer automatically allocate a code. However, initial investigations indicated that automatic coding required very sophisticated techniques to replicate clerical decisions and would require the full keying in of write-in responses. In effect, the system would require a matching algorithm to determine a range of possible codes and an artificial intelligence algorithm to make the selection of a code from the range. ABS is looking at this technique in the coding of cause of death and is hopeful of applying some of the work currently being done by

the US National Centre for Health Statistics. If optical character reading had been possible in the Census the task of developing an automatic coding system may have been seriously contemplated. As it was, the system was ruled out in the initial planning phase as too costly and a high risk.

However, an alternative to automatic computer coding is computer assisted coding (CAC). This technique was used successfully in 1986 for the coding of occupations and it was felt that it could be extended in 1991 to all write-in responses. Two French censuses and the 1986 New Zealand census had adopted the approach. It was therefore decided to adopt CAC for the write-in responses and to use OMR for the initial capture of the self coding responses instead of key entry as in the New Zealand case.

4.6: CAC/Edit/Amend

A further feature of the New Zealand system was the use of interactive editing and amendment as part of the CAC process. In the 1986 Australian Census amendments to records which failed edit checks of the captured data was made by batch processing of keyed edit reports amended by clerks who referred to the census forms before changing any codes. An alternative to this system and to the CAC/edit/amend system was an entirely automatic edit/imputation process. However, this was rejected because the development costs were considered to be too high. For 1991 it was therefore decided to adopt the CAC/edit/amend approach but using substantial automatic edit resolution.

4.6.1: CAC

The proposed approach to CAC will be for the clerk to select a batch of forms, log on to the CAC system and identify the batch of forms to be processed. The system will then locate the corresponding data file containing the OMR captured self coded data, attach it to the clerk's session and then commence the processing.

The system will display the dwelling identifier of the form being processed and the clerk will locate the form relating to that identifier. The system will then present one by one the topics needing attention. These include topics requiring coding, topics for which no mark has been read and which require verification for genuine non-response or correction for data capture problems. Also displayed for correction will be self coded topics for which multiple marks have been captured.

For topics requiring coding ie those with write-in responses, the clerk will be requested to enter the respondents' description (or part thereof) for the topic. The system will take the description and search an index file for all matches on each word entered. A table will be constructed to give a list of those index entries which contain one or more of the entered words,

together with the number of words which were matched. The index entries with the highest number of words matched will then be displayed for the clerk to make a selection.

The clerk will be able to also examine all possible matches which were found. The selected match will be identified to the system by the clerk and the data record for the person will be updated with the corresponding code by the system automatically.

4.6.2: **Edit/Amend**

Once all coding has been completed for a person, intra record editing is performed. The editing strategy being adopted incorporates as much automatic correction as possible while ensuring sufficient clerical intervention to fix data capture problems. This is achieved by requiring the clerk to determine whether edit failures are due to incorrectly captured data or inconsistencies reported by respondents.

In cases where a capture problem has occurred the clerk will be required to correct the data. If edit failure is due to a reporting inconsistency, the clerk will invoke an automatic correction process by pressing a predetermined key. This process will either flag genuine non-responses, set one of the inconsistent fields to a predetermined dump code or in the case where 'age' is inconsistent against a range of topics, input an appropriate value.

In developing the automatic resolution rules, topics which take precedence over others will be specified. Precedence is determined by one of two ways. First, a write-in response is considered to be more reliable than the answer to a related tick box question if the two responses are inconsistent. Second, a variable is considered to take precedence over the inconsistent variables if the variable in question is consistent with the majority of a predetermined series of related variables for consistency checking purposes. The automatic resolution rules will be based on the precedence specified and will dump code the fields considered to be less reliable.

When the end of the batch of work is reached, inter record editing is carried out. This is done in batch mode and edit failure listings will be returned to the clerk for amendment.

4.6.3: **Quality Control**

A system of quality control will be used involving an independent check on the quality of the coding process. Codes for checking will be selected using a predetermined sampling plan.

The work of coders who have low coding error rates will be sampled using a small sampling fraction for monitoring coding performance and measurement of average outgoing error rate. On the other hand, the work of

coders with high error rates will be sampled with a large (close to 1) sampling fraction.

Using the same coding procedures as described earlier the quality control clerk codes the items presented on the screen. To ensure independence of the coding operations, quality control clerks will not be given information on the codes assigned by the CAC clerk.

A batch job is then submitted to match the results of the two coding processes and any discrepancies are reported. The list of discrepancies will be examined by a different quality control clerk who will determine the correct code and correct, where appropriate, the data file. Error rates will be derived for management information and used to measure average outgoing data quality and identify retraining needs.

After the data is cleared by the processes described, it is ready for transfer to the Mainframe computer for output processing.

4.6.4: Cost effectiveness

With the introduction of CAC there will be an increase in equipment costs, but a decrease in staff costs because of higher CAC throughput rates than manual rates and consequently a decrease in office accommodation requirements. Cost efficiency of CAC is in the main determined by the extent to which the increase in equipment costs is more than offset by the decrease in salary and rental costs.

The cost of processing is of course dependent on the content of the census and viability of CAC will vary with different census contents. As the final census content has yet to be determined, assessments of cost efficiency has been calculated for several content scenarios.

Cost effectiveness varies according to the number of shifts used. It is also effected by the length of time taken to complete the processing and whether the processing is centralised or decentralised. Bearing in mind other factors such as the availability of staff to work multi-shifts and the need to produce the output in a timely fashion, it has been determined that a OMR/CAC/edit/amend system, using a centralised 1.5 shift operation and processing the census over 12 months is the most cost effective option for 1991. This doesn't take into account the cost efficiency implied in the better quality data due to OMR/CAC.

4.7: Job Design

As part of a general government move towards more rewarding (and ultimately more efficient) job design, the 1991 Census will pay particular attention to this issue. For example, in 1991 each CAC clerk will be responsible

for the processing work of the entire collectors district assigned to the clerk. Clerks will be responsible for clearing the CD through the system, including the coding of all topics using CAC and the resolving of edit failures.

5: The Output Processing System

As indicated at the beginning of this paper, the output system for 1991 will not be significantly different to 1986. The product range will include statistical publications, broadly divided into a standard series and a special analytical series. It will also include a number of magnetic tape files for use on mainframe computers plus a series of products on floppy disk with spreadsheet facilities. In addition, a CD-ROM will be available which includes software enabling maps and tables to be produced. Data on microfiche will also be released. In addition, there will be a series of technical papers and information papers relating to the census and to census data.

That concludes my presentation. Thank you.

2 . THE 1991 CENSUS OF CANADA: DETERMINING
THE QUESTIONS TO BE ASKED

The 1991 Census of Canada: Determining the
Questions to be Asked

Notes for presentation by D. Bruce Petrie,
Assistant Chief Statistician of Canada, to the
Twelfth Population Census Conference, organized
by the East-West Centre and the State Statistical
Bureau of China

Beijing
September 5-9, 1988

The 1991 Census of Canada: Determining the Questions to be Asked

It is a pleasure for me to participate in this Conference, and to have the opportunity to discuss with you the work we have been undertaking in Canada to prepare for our 1991 Census of Population. As requested by our Conference organizers, I will focus particularly on our efforts to develop the content, or questions to be asked in the census.

To set the stage for my remarks on what we have been doing to develop the census content, I will first outline the role of the census in the Canadian statistical system, since this is an especially important factor in influencing decisions on census questions. I will also summarize the needs and pressures for changes relative to our recent censuses, and note the constraints we must face in introducing these changes. No doubt both the pressures for change and the constraints will be familiar to most of you.

In describing the work we have been doing to develop the questionnaire content for 1991, I will outline in particular the extensive program of user consultation which we have undertaken, the testing program which is in progress and the process by which a final decision will be reached on the questions to be asked.

Role of the Census in the Canadian Statistical System

As in most countries, our census of population is a major component of our statistical system. In Canada, however, I believe that the census of population has evolved to play a much more substantial role than is the case in many, if not most, other countries.

As in many countries, our Constitution calls for a census every ten years, for the familiar reason of determining appropriate representation by population in Parliament. Other federal and provincial legislation requires census data for a broad range of other important purposes.

But ever since the first census in 1666 in what is now Canada, the census of population has been used to provide information on matters going far beyond mere counts and distribution of the population.

Since 1951, a census of population has been carried out every five years, a periodicity now required by law under the Statistics Act. The range of information collected in the census, especially in the mid-decade program, also increased over time. The number of questions asked in the census reached its zenith in 1971, but the census of 1981 and 1986 were only slightly less comprehensive in the range of information collected.

The result of this evolution has been to make the census of population an increasingly important component of our statistical system, and has led to great dependence on the census as a source of social and socio-economic information at both national and local levels. Because the census source has been there, we have not developed some of the survey and administrative data sources available in many other countries, nor could we readily do so in the short term.

The importance of the census in Canada is illustrated by the events leading up to the 1986 quinquennial census. As many of you will recall, a decision to cancel this census was taken by the government as an expenditure-reduction measure, but the program was reinstated in the face of sustained and widespread pressure from data users who made abundantly clear their need for census data.

For 1991, then, we must carry out a census to provide the basic population data required under our Constitution. And, our planning assumption is that we must continue to collect the wide range of additional data which Canadian data users expect from the census. Our dilemma is to determine exactly what that information should be, knowing that demands and expectations far exceed our capacity to deliver.

The Need for Change

In approaching our planning for 1991 questionnaire content, we recognized from the outset that significant changes from previous censuses probably would be necessary. Our 1986 census was, by design -- with a view to minimizing costs -- in large measure a repeat of the questions asked in 1981. That census, in turn, was little different in concept from that of 1971.

On the other hand, social and economic conditions in Canada have changed significantly in the past two decades. The social issues and concerns which preoccupy Canadians today, and which will be of interest in the coming decade, are in many ways far different from those of the fifties, sixties and seventies.

Our strategy, then, in addressing content issues for 1991, was to set out to identify the areas where change was needed, either in terms of questions dealing with new topics or with modifications to existing questions which would yield information more relevant to the needs of the nineties.

Constraints

There are, however, as all of you are well aware, serious constraints which face those of us who set out to make changes in census taking, whether in content or in operational procedures.

Not the least of these constraints, of course, is the budget. We at Statistics Canada have, of necessity, developed some rather innovative approaches to financing our census program -- ranging from cost sharing with other federal departments to sharply increased prices for census data -- but we still are faced with severe budget restraint.

Making changes to questionnaire content is expensive. The changes must be incorporated throughout the system ranging from enumerator training to final tabulation and output programs. The changes must be tested, and often modified and re-tested. Such testing is expensive. We have found from bitter experience -- as recently as 1986 -- that failure to test adequately can be even more expensive, leading to wasted resources, adverse respondent reaction and data user frustration.

Time is also a major constraint. With a five-year census cycle, we have found it very difficult, especially during the past several years of budget restraint, to devote adequate time, resources and attention to long-term research and development on both content and operational issues. As a result, we have had to accept the fact that in planning the content for 1991, there are some changes which, however desirable they may prove to be, we simply will not have time to implement for 1991. Consequently, we are making a conscious effort to devote increased attention to post-1991 content issues, so as to be in a better position to deal with them in 1996 or 2001.

Another familiar constraint on census takers is respondent tolerance. We believe that we have little, if any, room to increase the number of questions on our census. Hence, in planning for 1991, we have taken the position that if we are to add new questions, we must delete some of the existing ones. In our discussions with data users, however, we have been singularly unsuccessful in obtaining suggestions, let alone consensus, on what questions might be dropped.

The need to maintain continuity or comparability in census output from one census to the next also constrains change, especially when it comes to modifying existing questions. It has been our experience that many users resist change, even when modifications would produce data clearly more relevant to current needs, and even when the comparability in previous census results is more apparent than real.

Finally, an important constraint on the time and attention we can devote to content change in the next Canadian census is our need to improve coverage, i.e., to reduce the rate of underenumeration. In the 1986 census, we experienced an undercoverage rate estimated to be in excess of 3 per cent, some 50 per cent higher than the level of the past two censuses. Our top priority for 1991 is to improve coverage, and any changes to census content must be considered in light of this objective.

Determining User Needs

Recognizing both the need for change, and the constraints facing us, we have made a concerted effort to determine the needs of census data users. Borrowing some ideas from our American and Australian colleagues, and building on our own past experience, we launched a very extensive program of user consultation, much broader in scope and objectives than any we have conducted in the past.

Our primary objective was to canvass existing and potential users of census data to determine as precisely as possible their current and future information requirements. At the same time, we wanted to start building a widespread awareness of the forthcoming census, and to establish across the country a network of contacts upon whom we could call for support during the taking of the census. Achievement of these latter objectives, we hope, will make a significant contribution to our efforts to improve coverage in 1991.

Traditionally, in planning for the census, we have consulted major data users in all levels of government and in the private sector. This time, we broadened the process to involve the general public, inviting any interested citizen or group to meet with us and express their views.

We held a total of some 88 meetings throughout the country, involving about 1,150 participants. The meetings were advertised in local newspapers, and where possible we tried to obtain radio coverage and T.V. as well. People who registered in advance were provided, if they so requested, two background documents on the census -- 1991 Census: A Blueprint for Discussion, and The 1991 Census Content Guide Book.

About half the participants at the meetings were representatives from various levels of government. The academic and library communities accounted for about 15 per cent, while the remainder were drawn from private industry, ethnocultural associations, religious and social service organizations and special interest groups. Private citizens represented about 6 per cent of the total. (A more detailed breakdown of participants is provided in Table 1, attached.)

A careful record was kept of topics discussed and comments or recommendations made at the meetings. More than 3,600 comments, both oral and written, were recorded and subsequently analysed. (A breakdown of comments by topic is provided in Table 2.) More than 150 written briefs on content issues have been received, about one third of which were presented at the consultation meetings.

What have we learned from the process? Not surprisingly, nothing startling. Based on our continuing contact with major data users, we had to begin with a fairly good idea of the issues

of concern to our users, and I will summarize these in a moment. What we did gain, however, was a much better appreciation of the relative priorities of smaller data users, and a fuller understanding of the uses, or potential uses, of our data. We also, I think, established the contacts we were seeking across the country. And, not unimportantly, we exposed many data users to the demands and concerns of others, so that when we ultimately settle on the questions to be asked, those whose recommendations have not been followed will have a better understanding of why they have not been. We fully intend to provide further opportunities for discussion with these interested groups before we reach the final stages of content determination.

Our user consultation process has confirmed the continuing demand for the variables which traditionally have been included in the census in Canada: age, sex, marital status, family status, language, education, ethnicity, religion, fertility, labour force status, industry of employment, occupation, income, mobility and housing. It also confirmed that existing questions fall short of meeting current information requirements in several sensitive areas:

- (1) family status, where existing questions do not enable us to reflect adequately the living arrangements common in today's society (e.g. consensual unions, joint custody arrangements for children of divorced or separated parents);
- (2) language, where the bilingual nature of our country gives rise to complex and difficult questions about language transfer, assimilation, and other minority language issues;
- (3) ethnicity, where our existing questions seem to satisfy hardly anyone, but where the promotion of multiculturalism, employment equity and other such programs has led to more and more demands for data on the numbers and characteristics of Canadians of different ethnic and cultural backgrounds; of particular concern is information about Canada's Native Peoples, where, despite special efforts, the 1986 census failed to provide much-wanted better data; and
- (4) labour force participation, where our traditional questions, based on the concept of the market economy, do not reflect the activities of persons doing housework, child-rearing or volunteer community work.

Testing the Options

Based on our user consultation and previous research, we have been conducting what must be described as a very limited amount of testing of new questionnaire formulations to address the issues I have just described.

Concurrent with our consultation program, we conducted two small tests (3-4,000 households each) involving new or

revised questions on the topics of marital status, fertility, nuptiality, unpaid work, work history, income, language, ethnic origin, ethnic identity, race or colour and mobility.

Based on these tests, and on the user consultation exercise, we have developed what we are calling our National Census Test, which will be carried out next November. It will involve some 32,000 households, and is designed so as to enable us to produce weighted estimates for the various questions. The questions included in the test undoubtedly will not all be included in the census in 1991, but they represent at this point our "best guess" as to the wording of questions which may ultimately be included.

One thing is for certain: we will not include questions in 1991 which have not been tested in the National Census Test or have not been proven in previous Census or major survey programs.

Reaching a Decision

In Canada, the final decision on what questions will be included in the census is made by Cabinet. We, through our Minister, will present to Cabinet late next year options for consideration, based on our user consultation and testing programs.

In developing those options, we must weigh the evidence carefully, taking into account the factors I mentioned earlier: cost, respondent acceptance, the need for continuity/comparability, new or emerging requirements, the availability of data from alternative sources (actual or potential), and the overriding need to improve our coverage in 1991.

For each potential question, or group of questions, we will state as clearly as possible the costs and the benefits: who needs the data, for what purpose, and what will be the incremental cost of collecting and processing the information.

Experience has shown us that a census with the minimum required content, i.e., the questions needed to provide a complete and accurate enumeration, costs about 80% of the amount needed to conduct a "full-scale" census equivalent in scope to those carried out in 1981 and 1986.

This fact helps demonstrate the tremendous marginal benefits to be had through the addition of questions to the minimum "head-count", and our data users have made it very clear that they need and want the additional information. We are looking forward to the challenge of meeting as many of these needs as we can, and I hope that by learning more about your experience in this regard we can do a better job in Canada.

TABLE 1

1991 CENSUS CONSULTATION MEETINGS

Representation by Organization Type

Organization	Number	Percentage
Federal Government Department	194	16.9
Crown Corporation	8	0.7
Federal Government Agency	15	1.3
Provincial Government	232	20.2
Provincial Advisory Group	5	0.4
Regional Government	53	4.6
Municipal Government	96	8.4
Aboriginal Group	18	1.6
Private Industry - General	25	2.2
Private Industry - Financial	6	0.5
Private Industry - Manufacturing	8	0.7
Private Industry - Retail	6	0.5
Private Industry - Services	6	0.5
Private Industry - Transport	--	---
Private Industry - Small Enterprise	4	0.3
Private Industry - Consultants	26	2.3
Labour Unions	5	0.4
Ethnocultural Associations	47	4.1
Special Interest Groups	49	4.3
Religious Organizations ¹	13	1.1
Social Service Organizations ¹	58	5.1
Media - General	2	0.2
Media - Press	26	2.3
Media - Radio	20	1.7
Media - Television	3	0.3
Education - General	9	0.8
Education - Researchers	72	6.3
Education - School Boards	16	1.4
Libraries	55	4.8
Private Citizens	69	6.0
Total	1146	100.0

TABLE 2

Topic	Number	Percentage
Aboriginal	158	4.3
Basic Demographic	133	3.6
Education	165	4.5
Ethno-cultural	545	14.9
Family	163	4.5
Nuptiality & Fertility	217	5.9
General	207	5.7
Health	164	4.5
Households	122	3.3
Housing & Shelter Costs	406	11.1
Income	130	3.6
Labour Force	217	5.9
Language	231	6.3
Mobility & Place of Work	279	7.6
Non-Content	275	7.5
Other	245	6.7
Total	3657	100.0

3. THE UTILIZATION OF CHINA'S 1982 POPULATION CENSUS DATA

Paper for the Twelfth
Population Census Conference
5-9 September 1988
Beijing, China

THE UTILIZATION OF CHINA'S
1982 POPULATION CENSUS DATA

Li Chengrui
Former Director-General
State Statistical Bureau
People's Republic of China

THE UTILIZATION OF CHINA'S 1982
POPULATION CENSUS DATA

Li Chengrui

The aim of a population census is to provide data to meet the maximum benefits of the society. When we are preparing the 1990 round population census, a review of the utilization of China's 1982 population census data, its achievements and the problems that occurred, may be helpful to improve the future census work in this respect.

During the third population census in China in 1982, a great amount of accurate data was collected from its one billion people. The utilization of the data have been going on since the data were published. The work began in October 1982 when the first batch of brief manually tabulated data were released, and it was expanded in an all-round way after the end of 1983 when detailed data of a 10 percent sample tabulated by computers were made public. The work has further developed since July 1985 when all the data processed by computers were made available.

China conducted its first census in 1953 and its second census in 1964. Compared with the previous censuses, the third census included more items with the data processed by computers. Therefore, the utilization of the data from the 1982 census are more profound in depth and width. In exploiting and utilizing the latest information, computers have been used and new demographic and statistical methods popularly applied in other countries have been adopted in line with China's actual conditions. We have also received support and assistance from some international organizations,

foreign countries and personages in our work.

Among the users of the information from the third census are government departments at all levels, public organizations, institutions, enterprises and research institutes. In devising guiding principles for population control and social and economic development in the last three years of the period covered by the sixth five-year plan (1981-1985), the government departments made an analysis of the social conditions by using the data from the census, and later they also based the drafting of the seventh five-year plan (1986-1990) on the analysis of the data. The census has provided accurate statistics for a projection on the birth peak in the 1986-1990 period and the ageing process of the population, as well as detailed data on population distributions by industry, occupation and education. This has enabled the central authorities to take notice of the problems beforehand and make relevant policies towards them.

Local governments, social organizations, institutions and enterprises have used information from the third census and got good results in dealing with various problems, such as family planning, aged people, children, education, insurance, employment, production planning and urban construction. However, quite a lot of units are not familiar with the utilization of the information from the census. To help solve this problem, a Collection of Examples in the Application of Data from the Third Population Census was published in 1985 by relevant departments in the provinces of Guangdong, Hunan, Hubei, Henan and Guangxi. Later, a handbook entitled Examples of Application of Population Census Data was jointly compiled by the Population Census Office under the State Council and the

Population Statistics Department of the State Statistical Bureau and published by the Sichuan Provincial Social Sciences Academy Publishing House in 1987. This has helped the utilization of the census information.

To help improve the study and analysis of the information from the census, the Population Census Office under the State Council and the Population Statistics Department of the State Statistical Bureau have co-sponsored training courses on "population and economy" and "population forecasts" as well as three large symposiums. The first symposium was held in Kunming in December 1983, the second, the International Seminar on 1982 Population Census in China, was held in Beijing in March 1984, and the third was held in Chongqing in October 1985. Statistical officials, statisticians and demographers from 21 countries and the United Nations attended the 1984 Beijing International Seminar, which was funded by the United Nations. Some 40 papers presented at the Seminar were compiled into a book entitled Census of One Billion People in Chinese and English, which was distributed to all the participants. The Economic Information & Agency in Hong Kong published an English edition of the book in 1987 for distribution in and outside China. At the other two symposiums, more than 100 papers were delivered and later a selected collection of the papers -- Analysis of the Data from the Third Population Census in China -- was published by the China Finance and Economy Publishing House in February 1987. Demographic institutions in China have organized a number of conferences to analyse the data from the census and many Chinese demographers have published a large number of papers and works on their analysis. Prof. Hu Huanyong, a demographer-geographer who is well-known for his work of China's first population distribution map published in 1935, compiled a book entitled China's Demographic Geography based mainly on

information from the 1982 census with the help from Prof. Zhang Shanyu. The work, published by the Huadong (East China) Normal University Publishing House in 1984, gave a push to the development of China's demographic geography. The Editorial Board of the Atlas of Aged People in the People's Republic of China, with Dr. He Huide as its editor-in-chief, compiled an Atlas of Aged People in the People's Republic of China (published by the Atlas Publishing House) in 1986 based on data from the third census. The atlas, with 46 maps, texts and figures, gives an overall picture of China's aged population. The work is a result of research on the ageing process as well as a tool for the further study of this issue. The Research Center of Economic and Technological Development under the State Council has organized a large group of experts to compile a highly academic and practical work entitled Population and Employment in China in 2000, which is the first volume of China in 2000 Series. Jointly sponsored by the State Education Commission, the State Family Planning Commission and the Population Census Office under the State Council, over 800 scholars and experts have joined in the compiling of a series entitled China's Population with Pro. Sun Jingzhi as the editor-in-chief and based on data from population censuses. The 32 - volume series, totalling 12 million Chinese characters, is being published by the China Finance and Economy Publishing House, and ten volumes have already come off the press. The entire project will be completed in 1989, to be followed by editions in foreign languages for distribution abroad.

Apart from the above-mentioned efforts in exploiting and utilizing the information from the census, the Population Census Office under the State Council and the Population Statistics Department of the State Statistical

Bureau have, in cooperation with other departments, carried out three major basic projects in the exploitation and utilization, namely, The Population Atlas of China, the Regional Model Life Tables of China, and the software for the Demographic Analysis System.

I. The Population Atlas of China

The compilation of an atlas of China's population was a project that late Chinese Premier Zhou Enlai and some veteran Chinese scientists had wished to fulfil, yet due to the lack of information from census they could not carry out the work in their lifetime. With abundant information from the 1982 census, conditions are ready for the compilation. The Population Atlas of China Population was published in August 1987 after three years' painstaking efforts by the editorial board of the atlas, which is composed of experts from various fields and guided by the Population Census Office of the State Council and the Geography Institute of the Chinese Academy of Sciences. The Chinese edition of the atlas is published by the China Statistics Publishing House, and the English edition, jointly published by the China Statistics Publishing House and the Oxford University Press in Britain.

The Atlas is mainly based on the information from the 1982 census, supplemented by data from the 1953 and 1964 censuses, historical documents, and censuses in Taiwan and Hong Kong as well as surveys in Macao in recent years. The Atlas shows in a comprehensive and systematic way the natural, social and economic characteristics of the Chinese population and its geographic distribution. The 137 maps in the Atlas are divided into eight groups, namely, the introduction (10 maps), population distribution (9

maps), nationalities (13 maps), sexes and ages (15 maps), population changes (15 maps), educational attainment (12 maps), employment status (41 maps), and family, marriage and fertility (22 maps). Also included in the atlas are texts totalling 30,000 Chinese characters, and figures of 17 major population items by counties and cities. The Atlas is of quarto size and in colour printing. The scale of the national maps is 1:8,000,000 or 1:16,000,000 and that of the regional maps, 1:500,000 or 1:750,000.

Advanced techniques of cartography have been used in preparing the Atlas, with 80 percent of the maps drawn by computers. A special computer software has been developed for the compilation, which covers extracting and checking of data from the census tabulations, establishment of data sets and files for cartography, calculation of population indices, statistical analysis, cartography classifying, digitizing and the whole process of map-making. In this way, the data from the census and other data are turned into explicit graphic forms that can be easily comprehended as a whole, thus greatly facilitating the utilization of the census results. In this giant project, 150 tapes from the third census have been processed and turned into maps, and over 1,000 items of population indices calculated with the results printed on 50,000 pages.

On February 3, 1988, the Atlas won high praises from an appraisal committee composed of noted Chinese demographers, statisticians, geographers, map-makers and computer experts. The Speciality Committee of Cartography and Geographic Information System of the China Geography Society praised the Atlas as "excellent work of cartography". At the 13th Conference of the International Cartography Society held in Mexico in 1987,

experts from various countries also highly praised the atlas. D.R.B. Tayler, a noted Canadian scholar, said the Atlas was the best example of population census cartography ever found in the world. And A.J. Coale, a demographer from the United States, believed that the Atlas was a high-level summary of the social, economic and demographic characteristics of the largest population in the world.

Yet there are also some shortcomings in the Atlas. For instance, the Atlas lacks maps on population migration since 1949, which is due to the fact that no data on population migration had been collected from the three censuses. Another defect in the Atlas is that it lacks maps on dynamic comparisons of some population characteristics, for there are not enough data available from the first two censuses, which covered less items than the third one and whose data were processed manually.

II. Regional Model Life Tables of China

Model life tables are necessary means for the study of population phenomena and laws governing population development, and population projections. For quite a long time, Chinese life tables were virtually not available. The life tables constructed by United Nations experts in the 1950s and the regional life tables constructed by Prof. A. J. Coale and his colleagues in the 1960s failed to get source life tables from China, and the few data on mortality in Taiwan and Hong Kong were not enough to represent the population situations in the vast Chinese territory. As a result, large errors are not avoidable in making forecasts on and statistical analysis of the Chinese population based on indirect estimates, in defining the effects of mortality on the sex and age structure of the

Chinese population, and in the study of related social and economic problems in China. The 1982 census for the first time collected reliable and comprehensive mortality data of the Chinese population, thus providing necessary conditions for the making of Chinese model life tables. In February 1984, the Population Census Leading Group of the State Council and the State Statistical Bureau set up an editorial board composed of experts from various fields for the compilation of China's regional model life tables, and assigned the technical work to the Demography Institute of Xi'an Jiaotong University and the Computer Center of the Sichuan Provincial Planning Commission. The compilation of the life tables was completed in September 1987 through joint efforts of all departments concerned.

The compilation of the life tables was divided into four stages: 1. the collection and verification of source data, 2. the classification and analysis of source life tables, 3. the establishment of models and identification of parameters, and 4, the construction of model life tables. In making the source life tables, the county is taken as the basic unit, but counties with a very small population are grouped together provided that they are in the same vicinity and have similar social and economic conditions and lifestyles. Each of such groups has a population of at least 300,000 , and usually the populations are around 500,000. In this way the mortality for different age groups are more accurately represented. Through efforts by the population census offices and computer centers of all the provinces, autonomous regions and municipalities, 3,136 source life tables were built, with half for the male and half for the female. The data on mortality come mainly from the 1982 census, yet supplemented by data from a large-scale survey on deaths from cancer conducted by public health

departments between 1973 and 1975, as well as data from some sample surveys since 1957. As a matter of fact, these data are different in quality. Even for the data from the 1982 census, which are reliable as a whole, there are differences between localities in terms of quality. We have classified the source data into four groups: very reliable, fairly reliable, reliable and not reliable by direct comparison and other methods that are generally accepted internationally. The sorting and checking of the source data have ensured the reliability of the data in the life tables. In the source life tables for the male, the integer of the lowest life expectancy at birth is 43 years and the highest, 73, in the source life tables for the female, the lowest is 43 while the highest is 78 (the national life expectancy at birth for the male is 66.4 years, and for the female, 69.4). By strict discrimination, 963 source life tables were selected for the male and the same number selected for the female. During the project, which is of a large scale rarely seen in Chinese history, large quantities of analyses and calculations have been conducted on computers, and cooperations have been carried out between various departments and branches. Meanwhile, mathematical models of high accuracy and suitable for China have been established by drawing on international experiences with improvements made in line with China's actual conditions. On March 17, 1987, the editorial board organized a discussion and appraisal of the methods and the initial model life tables. Then revisions were made according to opinions and criticisms from the discussion and appraisal. On December 26, 1987, the final work was presented to an appraisal committee composed of noted demographers, statisticians and computer software experts. The committee says, "The source data have been sorted and selected to ensure a high reliability, therefore the final life tables are practicable," "the methods used in the compilation of the life tables are of a rigorous

scientific approach," and "this is an excellent, advanced project."

China is a vast country with a giant population composed of many nationalities, and its social and economic development is uneven in different parts. In classified analysis, the death modes of China are divided into five regional groups, taking into account the similarity of death modes and proximity of localities. The first group is Southwest China; the second group is Central China and East China; The third group is North China; the fourth group is Northeast China; the fifth group is Xinjiang. In the generated model life tables for each group -- 360 in all, with half for the male and half for the female, the life expectancy at birth ranges from 40 to 75 years. The curves of the five groups of model life tables are to be found at the end of this article. From the curves the difference between various groups can clearly be seen. For example, to compare the second group (Central China and East China) with the fifth group (Xinjiang), in the second group, the death probability for babies and children is low, while for the middle aged and old people is comparatively high, meanwhile in the fifth group, the death probability for babies and children is comparatively high, while for the middle aged, especially for the old people is low. To compare the third group (North China) with the fourth group (Northeast China), the death probability for babies and children is comparatively low in both groups, but in the fourth group the death probability for the middle aged people is comparatively high, while for the old people is comparatively low, meanwhile the death probability for the middle aged people is comparatively low, while for the old people is comparatively high.

Preparations are now under way for the publication of the Regional

Model Life Tables of China, with both Chinese and English editions to be distributed in and outside China. To be included in the book is a detailed explanation of the methods used in compiling the life tables. To facilitate its use in practical work, a computer application Software for the model life tables has been made.

The compilation of the Regional Model Life Tables of China has been based on rich current information yet on poorer dynamic information. This defect will be amended after the fourth population census to be carried out in 1990. Improvements will also be made according to opinions to be solicited from scholars at home and abroad when the book reaches its readers.

III. Computer software for the Demographic Analysis System

Computer softwares have been extensively used internationally in demographic analysis. But demographic analysis softwares from abroad, such as those developed by demography institutions of the United Nations and others by foreign research institutes, are not entirely suitable for application in China. Some Chinese demographers have also made some softwares for demographic analysis and population forecasts in the past few years, but they failed to form a complete system. In January 1986, the Population Census Office of the State Council commissioned the task of developing a computer Software for the demographic analysis system to the Sichuan Provincial Population Census Office and the Computer centre of the Sichuan Provincial Planning Commission. The task was completed in January 1988, and approved by an appraisal committee on April 16, 1988. The committee says, "The Demographic Analysis Software is a success, and this committee agrees

to accept it. The committee plans to popularize the software in demographic research and in demographic statistics." In May 1986, the Population Statistics Department of the State Statistics Bureau called a meeting to recommend the software to all provinces, autonomous regions and municipalities.

The demographic analysis system software has been developed by drawing on international experience and according to the requirements in China. The whole system comprises two major parts: the data bank system and the analysis system. The data bank system is composed of a population data bank and an economic data bank which include major indices. The analysis system includes six parts: 1. statistics of population reproduction, 2. population control and population planning, 3. population and economy, 4. calculation of indices of demographic statistics, 5. curve fitting and multivariate statistical analysis, and 6. management and maintenance of data files.

Major characteristics of this software are as follows:

1) The whole software system constitutes seven modules, which are linked by programs and form an independent unit each. The analysis system includes the analysis of frequently quoted indices of population, thus greatly facilitating demographic analysis.

2) Man-machine interaction is realized by Chinese-character prompting on the screen, and some of the calculation results are shown in explicit graphic forms. Audio signals are given when the program is running, lest the operator should mistake it as "deadlock" happening.

3) Data for analysis come from two sources: the data files and the data bank, and the two sources can be converted. Such a design helps to reduce the number of data files, raise the operating speed and save internal storage.

4) What an operator needs to do is to follow the directions on the screen, select the right function, and fill in necessary parameters, and the operation will begin. At the exit of the program, the operator may have another run of the same function or select other functions. Such a design greatly facilitates the work of the operator.

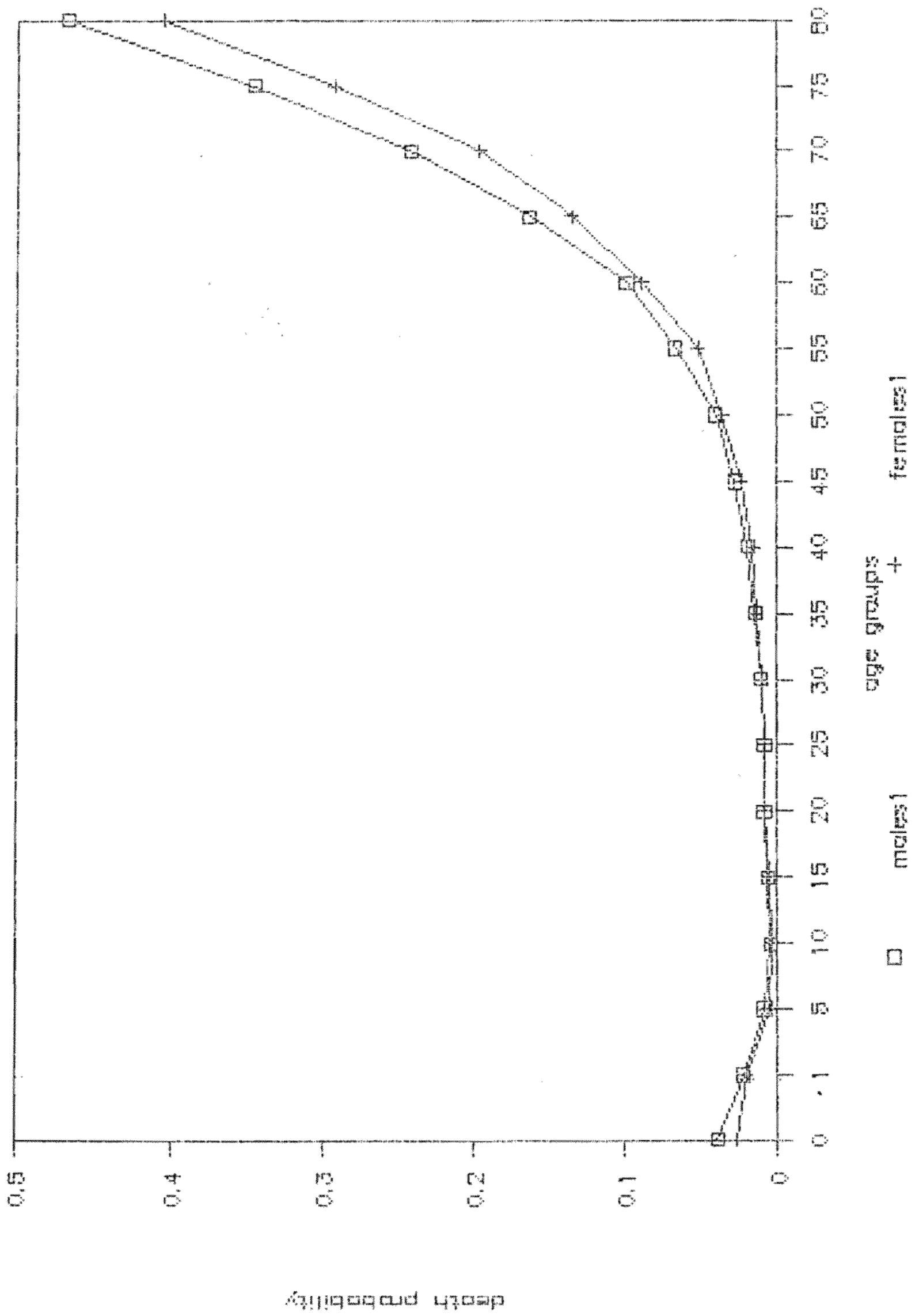
The above-mentioned characteristics show that the software is of a clear structure and easy to read, and even non-professional operators can use it. In fact the software has also some shortcomings and further efforts need to be made to improve it.

Generally speaking, the exploitation of the information from the third census is a great progress compared with the work on the previous two censuses. The Population Atlas of China and the Regional Model Life Tables of China in particular have contributed to the development of world demography, since they have filled the blanks in the international study of China's population. On the other hand, the exploitation has also a lot of shortcomings: 1) The work had not been listed in the plan before the census began, it was difficult to organize resources for the work in the late stage of the census, and the demographic analysis system software came out too late. 2) The published materials of the census have not included enough detailed information, thus hindering comprehensive exploitation of

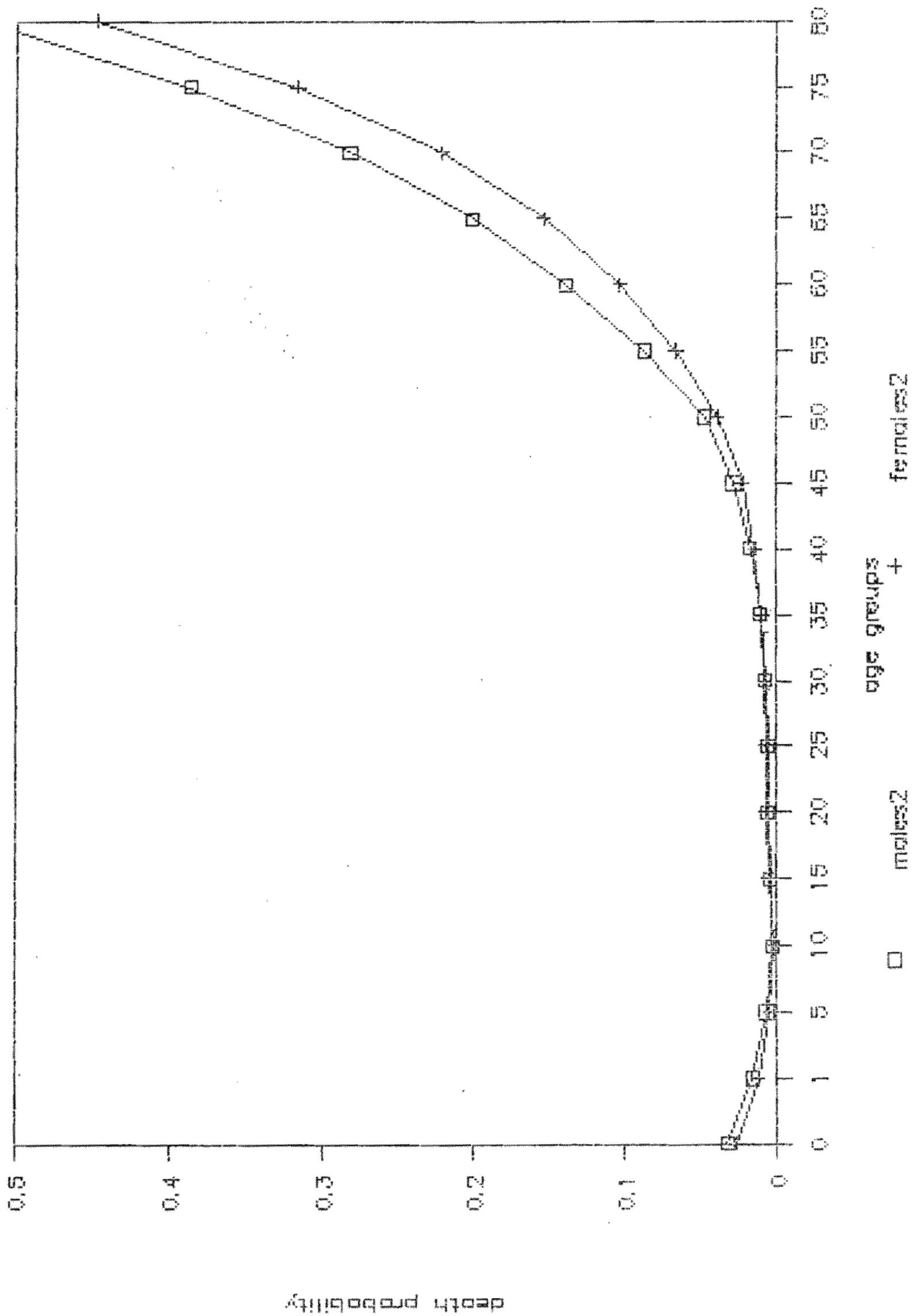
the information. 3) Government departments and the society as a whole are not familiar with the use of the census information, especially in using computers to analyse the census information. 4) Not enough international cooperation has been carried out in the exploitation. All this proves that much has to be done in the exploitation of the rich and reliable information from the census of a population of one billion, which has cost a huge quantity of both human and financial resources.

We hope that in the fourth population census to be carried out in 1990, the exploitation will be listed in the plan at the very beginning (including the revision of the atlas and the life tables and the improvement of the demographic analysis system software), more detailed information will be published (including the summary data of all provinces, autonomous regions and municipalities from the third census), more efforts will be made in personnel training to enable more people to use computers in the information exploitation, and more international cooperations will be carried out according to needs and possibilities.

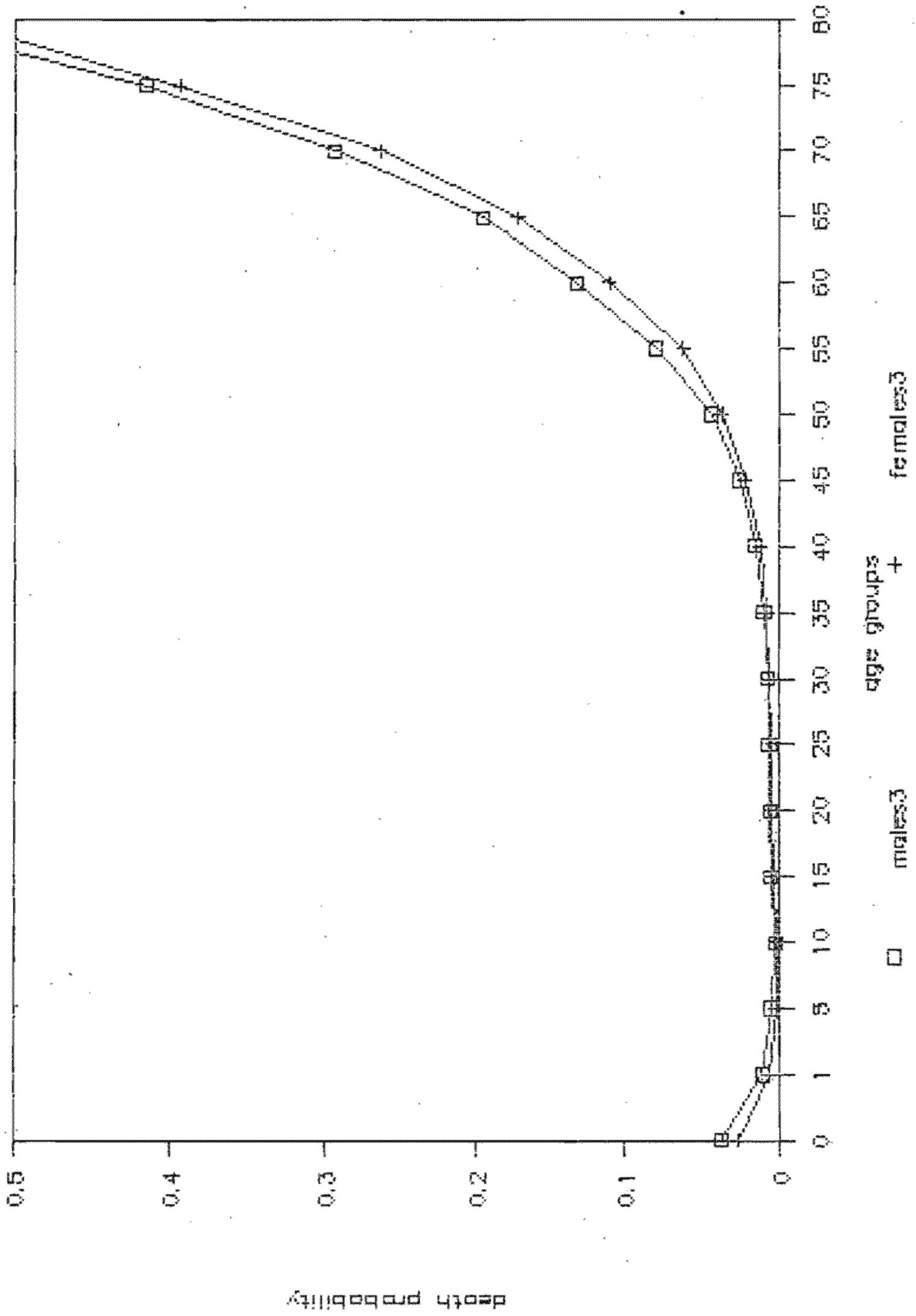
first pattern



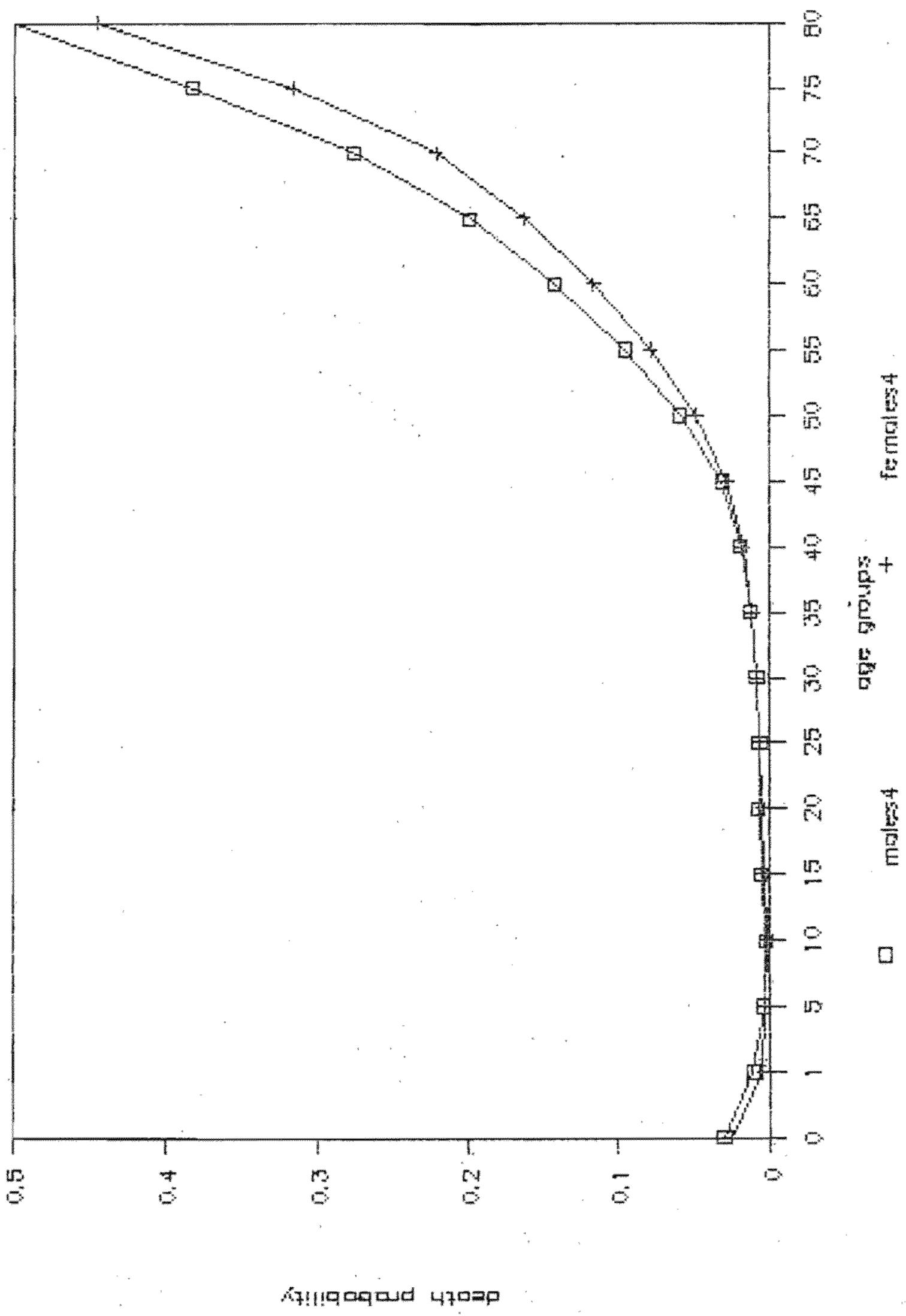
second pattern



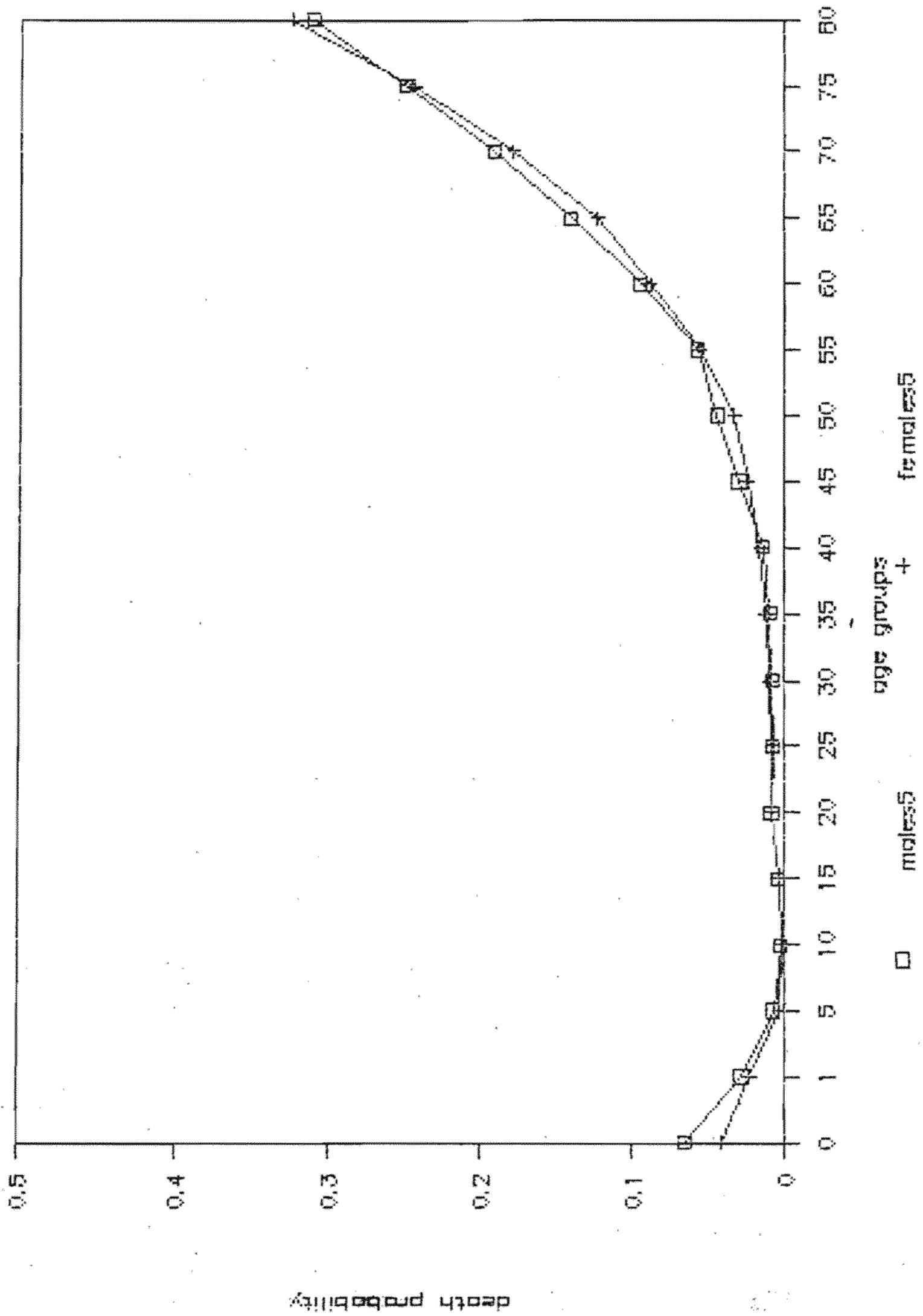
third pattern



fourth pattern



fifth pattern



4 . THE 1990 POPULATION CENSUS IN CHINA:
PREPARATIONS AND BASIC IDEAS

Paper for the Twelfth
Population Census Conference
5-9 September 1988
Beijing, China

THE 1990 POPULATION CENSUS IN CHINA:

PREPARATIONS AND BASIC IDEAS

Sun Jingxin

Deputy Director-General
State Statistical Bureau
People's Republic of China

THE 1990 POPULATION CENSUS IN CHINA:

PREPARATIONS AND BASIC IDEAS

China is a country with the largest population in the world, and the question of population has always been a major issue for the Chinese government and the Chinese people. To gain a clear picture of the changes of the number, distribution, migration, structure and quality of China's population since the country began to implement the policy of reform and opening to the outside world, and to provide detailed and reliable demographic data for working out relevant policies and plans, the Chinese government has now decided to conduct the fourth national population census in 1990, following the previous successful national population census in 1982 and the 1% population survey in 1987.

In order to prepare for the coming census, the State Council set up in August 1988 a Leading Group of the Fourth National Population Census and its working body, the Population Census Office of the State Council, and has issued the Instructions on the Planning of the National Population Census in 1990. Today, preparations for the census are going on nationwide.

Obviously, China is faced with a lot of difficulties in this census owing to the huge population of over one billion, economic and cultural backwardness, difficult accessibility to some areas and unevenness of regional development. It makes the population census more difficult that the population movement has increased rapidly and that the change of people's economic activities has been accelerated as a result of the implementation of the reform and open policy in recent years. Can China, in face of these difficulties, achieve new and great success in the coming,

census on the basis of the third national population census in 1982? This is a question of great concern and interest of the people at home and abroad.

I. To develop China's population census with distinguish Chinese characteristics on the basis of experience of success.

Although China's population census is faced with many difficulties, it has outstanding advantages and strong points in view of the country's reality. They are mainly demonstrated by the fact that China has governments at various levels, which enjoy high prestige among the people, that it has a highly centralized and unified socialist system, that the Chinese people have relatively high political consciousness, that the country has a well-developed household registration system, that there is a contingent of assiduous and hard working officials, that the masses love their motherland, support government work and take an active part in the population census, and, in particular, that the country has experience of success in the third national population census.

The reason why the third national population census succeeded is that the country had given play to the advantages and strong points of its society, overcome its shortcomings and difficulties, proceeded from its realities, and achieved success in past cases with Chinese characteristics.

--- Setting up strong census working bodies in the light of Chinese real conditions. To conduct census over a population of one billion needs strong leadership. Fulfilling a task in which the broad masses are

involved, the government will exercise centralized leadership with the coordination of the various departments concerned. During the third national population census, census leading bodies were established by the governments at various levels, with a leading official of relevant government taking the chairmanship and relevant departmental official taking part in the leadership. Thus, the census was listed as an important item on the agenda while all the important problems about the census were discussed and settled by governments at different levels. This ensured the effective leadership by the government over the census.

By drawing on the experience in this regard, the Leading Group of the Fourth National Population Census under the State Council will be formed with a state councillor as its head and with leading officials from departments of planning, statistics, household registration, civil affairs, propaganda, finance, labour, personnel affairs, public health, commerce, agriculture, etc. as its members. In locality, a deputy provincial governor, deputy prefectural commissioner, deputy county magistrate and deputy township head will be the head of the census leading group at the corresponding level. The working body, the population census office, will be composed mainly of personnel from the governmental departments of statistics, household registration and family planning, with some experts and scholars as advisors to the office.

In the course of the third national population census we first held pilot censuses in a "snowballing" way, and then worked out detailed rules suited to China's concrete conditions, while organizers of the census were trained in the practice of operation. This is our learn-from-practice experience, called "learn swimming by swimming", which was applied and

enriched to some extent in the 1% population survey in 1987. For the purpose of finding out an accurate population size of China in the 1990 census, we took a pilot census in Baoding Prefecture of Hebei Province in April this year, and now we have launched such pilot projects throughout the country. Meanwhile, a series of tests on special items are also underway, including the identification of occupations, industries, urban and rural residence classification of agricultural and non-agricultural activities, etc. of the employed people. More tests at different administrative levels are expected next year in order to train census personnel. These tests will help to formulate a census program suitable to the Chinese conditions and to find out a set of concrete measures for the implementation of the program.

----- Mobilizing all the people so as to win their support and cooperation. With a set of concrete measures being available, how can we implement them? Experience in the third national population census teaches us that the support of one-billion Chinese people and their close cooperation are essential to our success in that operation. Likewise, China's census operation in 1990 will be carried out under the unified leadership of the governments at various levels. In addition, extensive publicity will be given through different means at various levels and census staff at grass-roots levels will be mobilized to give people vivid, oral publicity. Through such publicity, we will make the people realize that to register faithfully in the census is every citizen's glory duty to the country.

----- Making the census a success by concentrating on the improvement

of the quality of census data. The central question of the population census lies in highly accurate census data. Experience in the third national population census has taught us that errors may emerge in three aspects, namely, the census design, the field enumeration and the data processing. Therefore, in the course of the 1990 census, we, on the basis of pilot censuses, will examine and revise the whole scheme repeatedly and eliminate design errors; reduce errors in the field operation by training census staff strictly and mobilizing the masses to make registration faithfully; and cut errors in data processing through the quality guarantee and quality control measures. In one word, we will urge everyone to attach importance to the data quality for the 1990 census.

II. Tentative ideas on the 1990 population census

The 1990 census, compared with the 1982 census, faces new conditions and many new problems, and we should take them seriously and solve them.

In the eight years from the 1982 census to 1990, great changes will have taken place in the economic and political life of the Chinese people. As a result, some new problems have emerged:

----- Along with the change of the distribution system and supply system, the cloth coupon has been abolished throughout the country, and the system of distributing the grain ration according to the number of family members in rural areas has also changed. This has lessened the residents' initiative in household registration, and increased the number of people who fail to make household registration, and aggravate the situation in which people make household registration in one place but live

in another place. So, how to gain the exact total number of the Chinese population will be a problem we have to endeavour to solve in the 1990 census.

----- Along with the deep-going of reforms and the implementation of the policies of enlivening the domestic economy and opening to the outside world, the population movement has increased, the inter-regional flow has been more frequent, and the floating population in urban areas has increased by a big margin. This new situation has added difficulties to the future census and posed a pressing task for us to make clear the floating population.

----- Along with the changing way of economic life, the pattern of occupations and industries of the population has become more complicated. Large numbers of agricultural population have been turning into multi-occupational or trans-trade population. These factors have added difficulties to the survey of the economic activities of the population. On the other hand, it is an important task for us in the 1990 census to get a clear picture of the changing economic activities of the Chinese people so as to serve the economic reform.

----- Because of the practice of the system of rewards and penalties in family planning in China, a number of people have become more reluctant to declare faithfully the number of their family members.

In view of these factors, we will regard adapting ourselves to new situations and solving new problems as vital characteristics of the 1990

census, and we will give particular attention to them in working out census scheme and in its implementation.

We shall adhere to the principle of "smaller quantity, better quality" in deciding the items of the census. To make comparison an easy job, we will keep the 19 items included in the 1982 population census, that is, name, relationship to the household head, sex, age, nationality, household registration, education, industry, occupation, inactive population, marital status, births given by one woman and survival children, births of women at child-bearing age in the year, household category, address, number of household members, and number of births and deaths in the household. Moreover, to meet the new demands for demographic information, some new items will be added, including housing status of urban residents, migration, temporary residents in large- and medium-sized cities, etc.

The census forms will have two types: long form and short form. The long forms which cover more items will be used of sample survey, while the short form will be used for the entire Chinese population.

Now the scheme for survey is under experimentation and will be decided after repeated examination.

As to the enumeration we plan to send enumerators to visit households instead of having residents enumerate at the enumeration centre. It was estimated that in the 1982 census about 80% of the enumerated population were enumerated at the enumeration centres, about 15%, enumerated through household visits by census staff, and about 5% enumerated at crop-fields,

aboard ships, in tents and at construction sites by enumerators.

To make enumerations at the enumeration centres is not as good as that through household visits, because residents are reluctant to tell their family secrets at the enumeration centres. In comparison, it is much easier for enumerators to get accurate data by having talks with residents and observing their conditions at their households. In fact, we made enumerations through household visits by enumerators in the 1% population survey in 1987. Facts show that this way helps improve the quality of the survey and is more acceptable to the residents. Hence, we will use this way as our chief method for the 1990 census.

We plan to make a big improvement in data tabulation. The 1982 census was of high quality and the data processing was a success. However, because of the lack of experience, we only worked out 79 tables as a result of little consideration of detailed tabulations. A lot of important information was thus lost. This has brought about a great deal of inconvenience to the use of information and especially to the data analysis and studies. It also affected the utilization of the data and the social effects of that census, although we later retabulated the data of 1% of the households in 140 tables with seven volumes. Following this lesson, we paid greater attention to improving the tabulation and about 280 kinds of tables were designed during the 1987 survey. This is a considerable progress in both quality and quantity of tabulations as compared with that for the 1982 census. The design of tabulation scheme for the 1990 census is underway. According to preliminary estimation, we will design some 500 tables in 16 volumes. Meanwhile, we will emphasize the practical results of the

tabulation in use, not only increasing its quality but raising its effectiveness, so as to make the population census bring about more social benefits.

So far, we have carried out the survey over the shifting of rural labour forces, worked out a criterion for the division between the agricultural and non-agricultural population, and drawn up a new plan, which has already been approved by the government and will be carried out in the 1990 census.

Recently, we are organizing pilot projects for getting a clear picture of the urban floating population. Based on the experience in such pilot projects, we will put forward an enumeration program for the urban floating population. This program, if approved by the government, will be implemented in the 1990 census.

For a better use of household registration data, the census offices at various levels will help the household registration agencies make a thorough checking before the census begins. The improved registration system will play a better role in the population census.

In the course of preparations for the 1990 census, we will attach great importance to publicity. We will particularly publicize the points that the population census will not affect any of residents' rights and duties; that all those who have failed to apply for their household registration are welcome to register this time, and their liability will not be pursued; and that no personal and household registration data will be used for the purpose other than census data compilation. Thus, people

will be free from worries about reporting the exact numbers of their families.

As to data processing, we will draw on the experience in processing the data of the 1987 1% survey, and will collect and compile data at prefectural, provincial and central levels respectively. This will be done in the following steps: in prefectures or cities of the prefectural level (also called centre cities), census data will be entered with a county as one unit, and county tables and prefecture or city tables will be compiled and worked out; then, the province tables, at provincial level; finally, the national tables, at central level. This processing pattern has three advantages:

1. Time will be saved by using the parallel data entry;
2. Errors at any level can be found in time through feedbacks and corrected at local levels; and
3. The time for the tabulation is comparatively short and the data can be used promptly at all levels.

This way of data processing was adopted in the 1987 1% population survey and was proved successful. It suits China's conditions of large population and backward communications, and helps guarantee the quality of data entries.

During the 1982 population census, we drew on experience and practice of other countries and created our own methods according to China's actual

conditions. In the 1990 census, we will earnestly learn advanced experience and working methods of other countries, improve our work in the light of China's conditions, and improve the population census with Chinese characteristics. We will try our best to achieve fresh and greater success, upgrade China's population census, and make our due contributions to the world population censuses.

5 . IN-DEPTH FERTILITY SURVEY IN CHINA

Paper for the Twelfth
Population Census Conference
5-9 September 1988
Beijing, China

IN-DEPTH FERTILITY SURVEY IN CHINA

Shen Qiuhua

and

Chen Runtian

Department of Population Statistics

State Statistical Bureau

People's Republic of China

IN-DEPTH FERTILITY SURVEY IN CHINA

by Shen Qihua and Chen Runtian

The State Statistical Bureau conducted two phases of in-depth fertility survey respectively in 1985 and 1987 and got fruitful results. The first phase was conducted in the two provinces of Hebei and Shaanxi and in Shanghai municipality, and the second phase, in the five provinces of Liaoning, Shandong, Guangdong, Guizhou and Gansu and in Beijing Municipality. The reports on related data from the first phase survey have been published, and the preliminary report on the second phase survey has also been made public. These reports have provided important information for the study of fertility situation in China.

I. Background of the Survey

Since the founding of New China in 1949, remarkable changes have taken place in the marital and fertility status of the Chinese women as a result of the great changes in the country's social and economic conditions. China's birth rates kept dropping during the 1970s but have fluctuated since 1980, ending the continuous decline in the past decade. What are the major changes in fertility in China? What are the factors contributing to these changes? And what is its trend of development in the future? The answers to these questions are of great importance to the study of China's population, the formulation of population policies and population planning. With approval from the Chinese Government, therefore, the State Statistical Bureau reached an agreement with the International Statistic Institute on an in-depth fertility survey to be conducted in China, which was also part

of the World Fertility Survey (WFS) program.

II. Organization of the Survey

As any survey calls for strict leadership and organization, leading organs were set up at the national, provincial and county levels for the in-depth fertility survey, with their functions clearly defined.

The national office was set up in the Department of Population Statistics of the State Statistical Bureau and in the charge of the Project Director, which was under the leadership of the Bureau. The national office's task was to raise funds for the survey, to work out the survey program and relevant documents and distribute them, to train key members of the national and provincial offices, to give guidance to the provincial offices in the selection of enumerators, and to coordinate the processing and study of survey data.

The provincial offices were set up in Population Statistics Divisions of the provincial statistical bureaus of the target provinces. They were under the leadership of the provincial bureaus, with an Assistant Project Director taking charge in each provincial office. The provincial offices were responsible for the organization and guidance of the provincial survey, the selection of field workers, the distribution of relevant documents, the training of interviewers, the organization of fieldwork, the quality inspection, and analysis and studies.

The county-level offices were set up in the statistical bureaus of the sampled counties, with the bureau director taking charge. The county

offices saw to the dispatch of field workers, the quality inspection and the management of fieldwork.

Survey offices at all levels had a clear definition of functions, with the superior supervising, inspecting and providing technical guidance for the subordinate while the subordinate responsible for the fulfilment of the task assigned.

III. Contents and Methods of the Survey

Since its establishment in 1972, the WFS helped conduct in-depth fertility surveys in various countries. The experience of other countries in this regard is conducive to the survey in China. As a matter of fact, the questionnaires employed in the in-depth fertility survey in China were jointly designed by experts from the State Statistical Bureau and the International Statistical Institute with reference to the core questionnaire of the WFS and in line with China's actual conditions. Used in the survey were three kinds of questionnaires: the individual questionnaire, the household questionnaire, and the community questionnaire.

The individual questionnaire, to be filled by ever married women under age 50, is the most important one of the three. It provides information on the basic characteristics of a respondent, her family and her personal histories of marriage, fertility and conception, as well as her knowledge and use of contraceptives, fertility preference and the status of her husband.

The household questionnaire, administered to the sampled household,

provided information on the economic status of the household which is needed in correlation analysis of the fertility level. Items in the questionnaire include the family members' age, sex and marital status, family income, housing conditions, and its possession of modern household appliances. This questionnaire is also used to identify eligible women for individual survey.

The community questionnaire is directed at the sampled villagers' committees -- the grassroots community in the rural areas. Its items cover information on the community, such as production, income, public transport facilities, public utilities, education, medical care, and family planning services.

As the in-depth fertility survey covers a wide variety of items, a nation-wide survey would be very costly and difficult to organize, therefore seven provinces and two cities were selected for the survey, which represented different regions. Hebei is chosen as representative of coastal provinces in North China; Shaanxi, representative of comparatively developed inland provinces; Liaoning, representative of northeast China; Shandong, representative of East China; Gansu and Guizhou, respectively typical of minority nationality areas in Northwest China and southwest China; Guangdong, representative of economically developed provinces in south-eastern coastal areas; Shanghai, representative of large cities; and Beijing, the political and economic centre of the country. These provinces and municipalities account for one third of the country's population.

The size of the sample for each province or municipality have to be

scientifically decided. Judging from the experiences of in-depth fertility survey conducted in more than 60 countries and regions in the world, a sample size should not be less than 2,000 women, otherwise it would be impossible to carry out detailed group analysis of the survey data. Yet if the sample size is too large, the survey would take much more funds and errors would also increase. To ensure that survey data from different provinces are comparable, sample size for big and small provinces should be roughly the same. In the first phase of the survey in 1985, 5,000 eligible women were selected in Hebei, and 4,000 were selected respectively in Shaanxi and Shanghai. In the second phase of the survey in 1987, the size of the sample were enlarged by 50 percent to raise the representativeness of the data, with 7,000 eligible women selected from Beijing and 6,000 respectively in the five provinces. During the two phases of the survey, the households visited amounted to 68,014 and the women interviewed, 52,460.

A multi-stage stratified PPS (probability proportional to size) cluster sampling scheme was adopted in selecting samples. The strata for each province or municipality were made in line with local conditions and according to the geographic, demographic, social and economic characteristics, and primary units were sampled from the strata. To ensure scientific sampling, the selection of primary sampling units (PSUs) for the provincial survey were jointly made by the Department of Population Statistics of the State Statistical Bureau and the provincial statistical bureau, while the selection of second-stage and ultimate sampling units were made by the provincial statistical bureaus with the results reported to the Department of Population Statistics of the State Statistical Bureau for the record.

The fieldwork was carried out by way of household visits. Usually the visits were conducted by a group composed of one supervisor and three interviewers. The manner of interviewing had a direct bearing on cooperation of the respondents and the quality of the data collected. It was of great importance for the interviewer and the respondents to understand and trust each other, and the interviewer was duty-bound to respect the privacy of the respondents. During the interview, efforts were made to persuade other people to keep out of the spot, so as to ensure that the interviewed woman was not constrained in answering questions. The questions were asked in a neutral way, giving no inducing hint to the respondents. As the interviewed woman had to recall some events in the long past in answering many of the questions, the interviewer should patiently wait and explain to her, and make necessary verifications. Usually a visit to a household took 10 minutes and an interview with a woman took 45 to 60 minutes. It took about one month for each province and municipality to complete the fieldwork and the quality inspection.

IV. Quality Control

To eliminate errors and raise the quality of the survey, we exercised examinations at various levels and short-line feedback inspections. The quality control was carried out at three levels -- on-the-spot inspection, county-level office inspection and provincial office inspection.

The on-the-spot inspection was done in a short-line feedback manner to ensure that errors were eliminated before the interviewers left the

surveyed spot. The interviewers were required to completely examine everyday's work, keep a daily record of the examination, and revisit the respondents should any errors, suspicions and omissions were found. Before leaving the surveyed spots, the supervisor made an overall examination and clarification, and return any questionable questionnaire to the interviewer for a revisit and verification.

The county-level inspection was carried out by in-door editing. The inspection should make sure that the number of the samples were correct, no items were omitted, the questionnaires were correctly coded, and there was no inconsistency in logic check. To assess the quality of the interviewers' work, the county statistical bureau also sent out experienced personnel to re-interview some of the sampled women and make a comparison between the interview results.

The provincial office was responsible for the final quality control before data entry. The inspection was conducted in two stages. First, the materials collected were exchanged between counties and 50% of each other's questionnaires were cross-examined. After the work was approved, the provincial office then selected and inspected 10 percent of the questionnaires. The the data entry began only after all the work was found qualified.

The inspections at various levels have raised the quality of the survey data. During the first phase of the survey, Shanghai found and corrected 1,962 errors (accounting for 0.6 percent), Shaanxi, 3,744 errors (one percent), and Hebei, 5,396 errors (4 percent). In the second phase of the survey, the number of errors found was somewhat less than that for the

first phase.

The high response rates also proved that the quality of the survey is satisfactory. The response rate for the household questionnaires was over 96 percent. For the individual questionnaires, the response rate in Shaanxi was 93.4 percent, in Shanghai, 95.2 percent, and in other areas, over 96 percent, with nearly 100 percent in four provinces.

V. Pilot Surveys and Training

Pilot surveys were an important link in the preparations for the survey. The pilot surveys could help find out whether there was any defect in the survey program and whether the program was feasible, and they also provided good opportunities for the training of personnel, as the survey workers could get familiar with the survey program in the pilot surveys.

In the first phase of the survey, three pilot surveys were conducted. The first was carried out in the urban and suburban areas of Shanghai on 200 households in October 1983. This pilot survey helped us revise the individual questionnaire with more items added in it. The second pilot survey, also on 200 households, was conducted in the urban and rural areas of Xi'an City in Shaanxi in March 1984. The aim of this pilot survey was to assess the feasibility of the revised individual questionnaire, which for the first time asked the interviewed women about their sexual life. With these two trials we were quite sure of the feasibility of the survey program. The third and the last one before the survey was a pilot survey on the national level conducted in Hebei's Shijiazhuang City and Huailu County

on over 300 households in October of 1984. Participated in this pilot were experts from State Statistical Bureau and from provincial and municipal statistical bureaux, as well as experts from the No.1 Medical Science University in Shanghai, the Xi'an Medical College, the Chinese Academy of Social Sciences and the Hebei Provincial Family Planning Commission. With this trial the questionnaires were finalized.

In the second phase of the survey, two pilot surveys were conducted. The first was an experimental training of key survey workers at the provincial level, and the second one, an experimental training of interviewers by the provinces and municipalities. The two pilot surveys helped improve the training and raise the competence of survey staff.

In-depth fertility survey was a huge project involving complicated items and techniques. What is more, it was the first time for many of the items to be surveyed in China. To ensure that the survey was smoothly carried out, it was necessary to have trained competent workers at all levels. For effectiveness' sake, the training was divided into three parts.

1. The training of field staff. The work was carried out at three levels' the national, the provincial and the grassroots levels. The national-level training aimed at enabling the trainees, by learning from the experience of fertility survey in other countries, to devise the survey program and detailed plans and organize the implementation of the program and plans, and to take charge of the personnel training at the provincial level. The provincial training aimed at enabling the trainees to master the sampling techniques and the requirements of the survey, devise the provincial sampling plan, and to take charge of the training of

interviewers and supervisors. The training of interviewers at the grassroots level helped the trainees get familiar with the questionnaires and learn how to fill them correctly, master the interviewing skills, the manner of raising questions, and the ability to check the questionnaires. A total of 478 field workers were trained for the first phase of the survey, and 1,150 were trained for the second phase of the survey.

2. The training of data processing personnel. Training courses were organized for different stages of the data processing, such as the developing of data entry program, data editing, data re-coding, table generation, and data archiving. In each course, experts from the International Statistic Institute were invited to teach, and the trainees went immediately into work right after the course. In this way the training was combined with the work of data processing and the practical work progressed smoothly.

3. The training for the analysis and study of the data. The large amount of information from the in-depth fertility survey was processed by computers, and about 200 summary tables were generated for each province, with each table containing different cross-tabulated variables. To analyse and study the abundant data, special training courses were provided for the personnel -- the preparation of the preliminary reports, the preparation of the national reports and the preparation of the special reports. After the training, the trainees were required to work out analytical reports of high academic level.

VI. Results of the Survey

Painstaking efforts by survey workers in the last few years have brought about fruitful results. As for the first phase of the survey, the preliminary report -- containing some 20,000 words and over 30 tables -- was published in Chinese and English editions in 1986, the national report -- in two volumes, with some 100,000 words and 188 tables -- was published in both Chinese and English in 1987, and the tapes containing the original data have been made available to domestic and foreign users at the Information and Consultancy Service Centre of the State statistical Bureau and at the Research Centre of the International Statistical Institute. Based on the data from the first phase of the survey, Chinese experts and experts from the Research Centre of the International Statistics Institute have jointly written a series of monographs on the interval between the first marriage and the first birth, the interval between breastfeeding and the next birth, the structure of marriage and family, the assessment of the history of contraception, the present status and difference of contraception, and the effectiveness and duration of contraception. The English versions of these monographs are soon to be released by the International Statistical Institute. As for the second phase of the survey, the preliminary report was published in Chinese and English editions in 1988, and the national report and tapes containing the original data are expected to be available in 1989.

VII. Conclusions

1. The in-depth fertility survey was conducted for the first time in China. Although population censuses, sample surveys on population changes

and other fertility surveys in the past had yielded some relevant information on fertility in China, none of them had covered such a wide range of variables as the in-depth fertility survey. The in-depth fertility survey is highly necessary for the study of fertility and related variables in China, the study of China's population development strategy, and the formulation of population policies and population programs. The survey has attracted the attention of demographic institutions in and outside China, which have asked for information on and from the survey.

2. Facts have proved that the two phases of the in-depth fertility survey are successful. The high quality of survey work and of data collected have won praises from experts at home and abroad. During their inspection tour to China, Mr. Gille, former official of the WFS, Ms. Vlassove from the International Development Research Centre of Canada, and Mr. Cleland from the Research Centre of the International Statistical Institute held that the survey was a success, the quality was high, and the data were reliable. Relevant Chinese government departments also believe that the survey reports have reflected the actual conditions in the country.

3. The in-depth fertility survey has not only provided abundant and reliable data, but also helped raise the professional quality of statisticians involved in population surveys and in demographic analysis, thus laying a good foundation for further surveys and studies in the future.

4. The in-depth fertility survey in China is also a good result of international cooperation and academic exchange. The survey has received

assistance from the International Statistical Institute and from some friendly countries. The governments of Norway, Finland and Denmark and the International Development Research Centre of Canada provided necessary equipment and funds for training, and experts from the Research Centre of the International Statistical Institute offered conducive technical consultancy services at different stage of the survey. Thanks to all this, the project was smoothly completed.

6 . AUTOMATION AND DATA PROCESSING OF THE
HONG KONG 1991 POPULATION CENSUS

Automation and Data Processing of the Hong Kong

1991 Population Census

Automation and Data Processing of the Hong Kong
1991 Population Census

Introduction

It is a well established practice in Hong Kong to take population censuses at regular intervals. Full censuses were held in 1961, 1971 and 1981 and sample censuses in 1966, 1976 and 1986. Plan is in hand to conduct the next full census in 1991.

2. The 1991 Census will enumerate all persons present in Hong Kong, including transients, the armed forces and the institutional population. It is intended that the census will be a two-phase operation comprising a 100 per cent simple enumeration of all persons by age and sex and a one-in-seven sample enumeration a broad range of social, economic, and demographic characteristics. Two questionnaires will be used : a short form for the simple enumeration and a long form for the detailed sample inquiry.

3. For the 1991 Census increasing importance is being placed on producing more timely results, improving productivity and maintaining a high level of accuracy. We hope to achieve these goals by enhancing the computerized frame of living quarters, by automating laborious clerical operations and by redesigning the methods of data capture and data dissemination.

The Computerised Living Quarters Frame System

4. In the past two decades, a comprehensive frame of living quarters was maintained manually. The frame was first established for the 1976 By-Census, and was updated for the 1981 Census. For the purpose of conducting the 1986 By-Census, the manual frame, which contained some 1.5 million living quarters records, was converted to computer files in 1985.

5. This computerized Living Quarters Frame System, or 'COM-Q' in short, consists of two parts. The first part is a fully computerized frame of living quarters in permanent buildings and the second part is a frame of segments of temporary structures. For the second part, the frame holds records at the segment level while individual living quarters records are updated manually. The system is maintained on an ICL 2988 mainframe computer.

6. The COM-Q will be enhanced to meet the 1991 Census operational requirements. The system will have the capability for selecting sample, organising enumerator blocks, preparing work assignments and printing addresses on letters to households.

7. For the 1991 Census, a mail-out enumerator collect approach will be employed for the simple enumeration. The COM-Q will serve as a mail register as well as a house listing for enumerator assignments. Its functions and outputs will be suitably amended to meet these requirements. The system will also be enhanced to support the formation of enumerator blocks and preparation of work assignments for the two phase operation.

8. With the increasing use of the frame for household surveys, more options in the method of sample selection have to be provided to meet post-census requirements. Options to be added will include multi-stage sampling and sampling for small areas.

Enumerator Information System

9. The 1991 Census will require some 17 000 temporary enumeration staff for the field operation. To allow for drop-outs in various stages, a total of 20 000 teachers and senior form students will have to be recruited for training. Suitable candidates applying for enumeration work will be selected based on specific criteria. Selected candidates will then be required to attend training classes at the end of which a test will be conducted to select suitable candidates for employment. Payment of an honorarium will be made to each enumerator who successfully completes his assignment.

10. These tasks were performed manually in previous censuses. For saving cost, improving accuracy and achieving better project management, it is proposed to computerise these procedures. The main functions of the Enumerator Information System are as follows :

(1) Formation of database

A database containing the necessary information of the candidates applying for enumeration work will be set up. Suitable candidates will be selected by reference to specific criteria.

(2) Training and selection of candidates

Selected candidates will be informed to attend training classes. The system will generate the formation of training classes according to prescribed criteria and prepare letters to invite candidates to the training. After the training, examination results and trainers' recommendations will be input into the system which will automatically select the suitable candidates for employment.

(3) Deployment of enumeration staff

The enumeration staff will be organised into different divisions and deployed to different field centres by reference to their place of residence. The name of the enumerator will be printed on his assignment list. It will also be printed on letters sent to households before the field operation to help respondents identify the enumerator when he visits them. Any amendments due to the sudden withdrawal of enumerators can be made within a very short period of time with the aid of this computer system.

(4) Payment of honarium

A magnetic tape containing the necessary information of the enumeration staff who complete their assignments will be processed for payment.

(5) Public enquiries

The Enumerator Information System will be linked up with the Computerized Living Quarters Frame System to support an on-line enquiry service whereby householders can check the identity of the enumerators over the telephone during the field operation.

11. It is proposed that a mainframe computer will be installed for the implementation of this system (and the Field Control System discussed below). A user-friendly data base management software will be employed. The system is expected to be simple and flexible enough to be operated by clerical staff.

Field Control System

12. In previous censuses, the control of the field operation was decentralised. Teams of enumeration staff were stationed in temporary field centres in all districts. Enumerators were under the supervision of the chief enumerators, who in turn were supervised by the district supervisors. Each regional supervisor took charge of his team of district supervisors and he reported to the headquarters the enumeration work progress of his districts.

13. To exercise a tight control over the field operation, prompt information on the progress and quality of enumeration work is essential. In previous censuses, it took more than two days for such data to reach the headquarters. This is considered ineffective. For the 1991 Census it is proposed to set up a computer system (the Field Control System), to monitor the work of the enumeration staff. This is a network system where two micro-computers will be installed in each field centre (total 60 centres) to be linked up with a host mainframe computer at the Headquarters through public electronic communication service. Daily reports by enumerators on work completed will be input into the micro-computers. The system will consolidate the data at each level of supervision and produce prompt reports and charts for individual teams, districts, regions and the territory for monitoring the progress of the operation.

14. In addition to the functions for controlling purposes, the system will have the capability for producing a preliminary population count by small geographical area for release after the field operation.

15. The system will also be used to improve communication, through electronic mail, between the headquarters and each field centre. Message received from the public can be re-directed through the network system to the corresponding field centres immediately to facilitate the operation.

Data Capture

16. The key-to-disc method was used in the 1981 Census for data preparation. The volume of data amounted to 325 million key depressions (including verifications), and 64 data processors were engaged for the task for a period of 4 months. The method was cost-effective and, with this approach, census schedules would not require special handling in the field, thus relieving the field supervision of that burden and resulting in better data quality.

17. For the 1991 Census it is planned to input and edit all census schedules within a period of 15 weeks. It is estimated that the volume of data required to be captured will be about 414 million KD, including verifications, if the key-to-disc method is used. The sheer volume of workload calls for careful consideration of the data capture approach for the census. In particular, there will be serious administrative problems in recruiting a large team of data processors to work for a short period of time and to identify projects to absorb the staff after the task is completed. To take advantage of advanced technology for data processing, consideration is being given to using optical character reading/optical mark reading (OCR/OMR) for data entry.

18. For the 1991 Census it is intended to pursue the feasibility of a mixed-approach for data capture. First, data from the short forms will be input by optical readers using a combined OCR/OMR data input form. Second, data from the long forms will be input by the key-to-disc method.

19. Answers to the census questions in the short forms will be transcribed from the schedules to data input forms by enumerators in the field centres. It is expected that there will be no great difficulty imposed on the field operation with this arrangement. Although the use of a separate form for data preparation would require extra work for transcription, this is not too laborious as enumerators will have to code the answers in any case.

20. Either OMR or OCR machine could be used for optical reading. For the 1991 Census OMR machines are not considered for a number of reasons : First, it imposes greater constraint on form design. Second, transcription errors might be higher if answers have to be coded and transferred to OMR input forms which have format different from the census schedules. Third, the OMR approach has limited applications in other systems and it would thus be difficult to utilise the machines after the census.

21. On the other hand, where answers are already precoded on the census schedules, it would be easier to use optical mark codes. For numerical answers, however, the OCR format would be more convenient to handle. Therefore we intend that a combined optical mark/character approach will be adopted for the 1991 Census.

22. For data transcription enumerators are preferred to a team of markers. This would enable the two steps, coding and transcription of answers, to be combined. With this procedure it would be easier for enumerators to check the completeness of answers on the census schedules. Furthermore, the problem of fatigue (if the task is performed by markers) could be avoided.

23. For the long forms coding and transcription of answers is a more complex task and would be subject to a high risk of error. On the other hand, using the census schedules as OCR/OMR input forms will be inconvenient in the field since extra care will be needed for form handling. For these reasons, the key-to-disc method will still be used for data preparation for the long forms.

24. The feasibility of the above approach will be studied in a pre-test to be conducted in February 1989 and in a pilot census in November 1989.

Data Dissemination

25. The value of census is greatly enhanced if data are available more quickly, in a format and medium geared to users' needs and in response to special requirements that may emerge. Considerations of timeliness, convenience and flexibility are important in designing programme for the release of census results.

26. To produce census results more quickly, it is intended to use desktop publishing software for preparing some of the census publications. Compared with the traditional typesetting method, this approach will provide publications more quickly because printing can be done mainly by a reprographic process.

27. Information from the detailed census tabulations will need to be extracted, summarised and analysed for publication. To facilitate this process, a network of micro-computers will be set up to process the census data. By downloading the tabulated data from the mainframe computer to the micro-computers, statistical staff can use spreadsheet software to prepare the publications. This will reduce the efforts for retrieval and checking and will result in an earlier release of the data.

28. To meet the wide range of users' needs, it is intended that census data can be made available in different formats and released in different media. For the general public, summary information of a wide interest will be disseminated through press releases and public electronic mail service. For convenience, where the data are needed for further processing, they will be disseminated in computer media such as tapes, diskettes, or compact disc read-only memory; the actual medium to be used will depend on the volume of information and the equipment available to user for handling the data. Users may also receive data in computer printouts if their main requirement is for reference and data retrieval. Census publications are important and will continue to be produced to facilitate a wide circulation of data, particularly for users not having an access to computer, and to present information with commentary and graphic aids.

29. The census data will be archived on optical storage devices to meet special requirements. Such devices will save storage space, ensure data integrity and permanence and facilitate future retrieval.

30. Users may have new data requirements after the release of the census results. To allow for such requirements, a flexible post-census service will be available, through which individual users will be provided with the data they need.

Conclusion

31. The above considerations for automation and data processing of the 1991 Census will be further pursued through more detailed studies of system design. In particular, systems for supporting the field operation will be tested in a pre-test and a pilot census. Although there is a whole range of automation possibilities which can improve efficiency, a word of caution should be added. The systems developed must be simple, because they will be operated by a work force with minimal computer training. They must be fully tested and proven reliable. The equipment must be reasonably priced and should continue to be usable after the completion of the census.

7 . PROCESSING OF DATA ON THE
1986 POPULATION CENSUS OF FIJI

TWELFTH POPULATION CENSUS CONFERENCE

September 5-9, 1988
Beijing, China

PROCESSING OF DATA ON THE
1986 POPULATION CENSUS OF FIJI

Naibuka Navunisaravi
Government Statistician
and Census Commissioner

Suva, Fiji
1 September, 1988

PROCESSING OF DATA ON THE 1986 POPULATION CENSUS OF FIJI

I. INTRODUCTION

- 1.1 The 1986 Population Census of Fiji was taken on 31st August, 1986 where 715,375 persons were enumerated. It was the 11th Census and the first was taken on 4th April, 1881. Decennial censuses were taken up to 1921, then the next one in 1986 and then continued on a decennial basis.
- 1.2 Manual processing of data on population census was the normal practise for previous censuses before the one in 1956. The 1956 census saw the beginning of an era of electronic data processing in Fiji even though facilities for such technology were not in existence in the country at that time. The data for that census was processed at the Australian Bureau of Statistics in Canberra using the old punch-card data capturing system.
- 1.3 The introduction of IBM computers at the Electronic Data Processing (EDP) Census had helped not only the processing of data on the 1966 population census, but also other statistical data for the Bureau of Statistics in Fiji. The punch card system was also used during that year. The data for the 1976 Census of Population was then processed using, for the first time, the magnetic-tape data capture system.

II. THE 1986 POPULATION CENSUS

- 2.1 The 1986 Census was the second in which magnetic tape data capture system was used for processing data. Although data was processed at the EDP Centre, most of the preparation of data were carried out by the Census Staff in the field as well as in the office.

(a) Manual Editing of Questionnaires

- 2.2 All census schedules were received at the census office on the 21st September, three weeks after census day (31st August). The staff of the Census office conducted the manual random checks on all schedules which include :

(i) Questions are completely filled.

(ii) Consistency checks on questions.

A check was also made on population summary on the batch covers. A provisional result of the population count was released in October as a result of a manual count of all population summary on batch-covers.

(b) The Coding System

2.3 After a number of discussions and modifications with the EDP Services, the Census Office submitted a list of numeric codes to the computer centre as being the final code list of the 1986 Census of Population. The list include :

- (i) E.A. codes (provincial, tikina, E.A. and locality codes)
- (ii) Industry codes (derived from PSIC - the Fiji version of ISIC)
- (iii) Occupation codes (derived from FCDO - the Fiji version of ISCO)
- (iv) Variable range check codes (eg. Male = 1; Female = 2)

(c) Coders and Data Entry Operators

2.4 Coding of the census schedules was planned to commence on 1st of October and to continue to the end of February 1987. Data entry was to commence by mid-October up to the end of February 1987 also. In order to keep to the time schedules on our data processing programme, the following number of workers were then recruited to do the work :

Coders	35
Coding Supervisors -	5
Data Entry Operators -	10
" " Supervisors -	2

2.5 The duties of the coders are to prepare census records for input to the computer for processing in accordance with the instructions of the Census Commissioner. This requires :-

- . clerical edit to ensure consistency as far as possible,
- . determining appropriate codes and classifications using the code list and the FCDO and FISIC,
- . entering appropriate codes on the record.

The duties of the supervisor are :-

- . to be in charge of a team of seven coders,
- . to be responsible for quality control and output,
- . to make detailed checks on a sample of coded documents to ensure that editing is consistent with the rules and that codes are accurately assigned,
- . to correct errors identified during the computer edit.

The conditions of services of the coding staff were as follows :

Duration of service: 14 October to 28 February 1987

Rates of pay: Supervisor n.a.
Operator \$81.62 per week

Working hours: A six hour shift, Monday to Saturday

Conditions: As for coders

Coding work requires close concentration and a high degree of accuracy. It should be noted that coding throughout breeds boredom. Under these conditions, performances tend to decline as the work progress, production is purposely slowed in order to prolong temporary employment.

An incentive bonus was introduced at the coding stage, where Coders/Supervisors received an extra four cents for every person coded beyond their weekly quota.

- 2.6 The coding work started on 20th October 1986 and by 15th December 1986 53% of the schedules were completed and by the 29/1/87 all schedules were coded of all records coded, 25% of them were coded at bonus rates.

The overall results were :

- (i) Coding cost was cut
- (ii) Timetable was met
- (iii) Good quality work was achieved
- (iv) Staff morale was excellent and
- (v) No staff turnover existed

- 2.7 The duties of data entry operators are to enter coded data to the computer and to maintain satisfactory rates of working and levels of accuracy in accordance with the instructions of the Census Commissioner.

The work of data entry operators basically are :

- . First enter batch record (EA number, Batch number, Total persons and total Households);
- . Individual records of persons in the batch are then entered.

The duties of the supervisor are :

- . to take charge of a shift,
- . to be responsible for standards of work,
- . to perform such other tasks relating to file handling, backing up and error correction as may be assigned.

The conditions of services of the data entry staff were as follows:

Duration of service: 14 October to 23 February 1987

Rates of pay: Supervisor n.a.
Operator \$81.62 per week

Working hours: A six hour shift, Monday to Saturday

Conditions: As for coders

(Note : Since the work is temporary and selection is by aptitude and not educational attainment, the same rate is proposed for all coders and for all data entry operators irrespective of qualification or previous experience.)

- 2.8. With 6 terminals, 5 key operators and a supervisor data entry was started on 14/11/86 on a six hourly shift per day from Monday to Saturday. By 15/12/86, 56% of all records were entered onto the main frame. The data entry was done on-line at the Census Office and the work was completed on 19/2/87, 2 weeks ahead of schedule.

(d) Automatic Processing at the EDP Centre
(1) Problems

- 2.9 Unfortunately the census is processed once every 5 years and during this time, hardware and personnel change. Therefore the project is treated differently each time and the personnel are required to start from scratch.
- 2.10 The development of software should be relatively straight forward with the only problem areas being the complex logic required for the edit/impute software and the large number of output tables required. Both these areas gave considerable problems.
- 2.11 There was a high turnover of EDP staff within the project. Although a range of hardware was used, the project requires 1/2 a Vax 11/750 and 1 RAS1 disk for 6-8 months. These resources were often not available and constant file backup/restore operations and machine load balancing were undertaken.

(ii) System Overview

- 2.12 The completed census forms are sent in batches to Statistics for collation and checking. Each batch is identified by its Province, Tikina and Batch number. Each batch is input to the computer system through terminals located at the Census Office, Ratu Sukuna House.
- 2.13 After a batch has been entered, the software checks the structure of a batch to ensure that all households and persons have been entered and that no data has been entered twice. Valid batches are merged into a Tikina file. If a batch has invalid entries the entire batch is passed back for correction.
- 2.14 When all batches for a particular Tikina are complete the Tikina file is passed through the Edit/Impute program to further validate the data and impute certain fields. If errors are detected the file is passed back for correction otherwise it is appended to the census master file.
- 2.16 A program is run against the master file to create the SAS dataset required for the reporting software. SAS procedures are run as required.

(iii) Run Time Environment

- 2.17 The run time application runs in a directory structure under USER-DISK5:[CENSUS]. The structure is split into subdirectories for batches, commands, data and runtime software.

The CDD node used is CDD\$TOP.CENSUS.

All software and data is accessible from any node in the cluster.

- 2.18 When the census users login, they get placed into a menu which has the following options :

- Run Data Entry program
- Run Structural Edit program
- Run Edit/Impute program
- Maintenance of Tikina or batch information in the DMF
- Print DMF information
- Maintenance of location information
- Print location information
- Exit from Census menu

These facilities will be discussed at a later stage.

(iv) Data Entry Software

2.20 The data entry program runs as an installed image and its source is located in [CENSUS.SOURCE]DATCH_HEADER.COB. The program accepts data from the screen, checks the key (province, tikina, EA number, batch) against the Batch Maintenance File (DMF), checks the location against the Locality File (LOC) and writes the entries out to an output files.

2.21 These output files are RMS relative files and have the name of [CENSUS.DATCHES]CENwwxyyyz.DAT where ww is the province, xx is the tikina, yy is the EA number and z is the batch.

(vi) Structural Edit Software

2.22 The structural edit checks the batches for internal consistency and writes the valid batches to a RMS indexed file. The source is located in [CENSUS.SOURCE]SDTEDIT.COB.

The tikina file has the form [CENSUS.DATA]TIKwwxx.TIK where ww is the province and xx is the tikina.

(vii) Edit/Impute Software

2.23 The Edit/Impute software reads a tikina file and validates each filed in turn. Depending upon certain criteria the program may modify (impute) the data in that field before it writes the data to the census master file [CENSUS]CENSUS_MASTER.DAT. The source is located in [CENSUS.SOURCE]SDTEDIT.COB.

(viii) DMF and LOC files

2.24 The Batch Maintenance File (DMF) contains a list of all batches and the status of each batch. The software will not allow a batch to proceed to the next processing step until the status indicator in the DMF has been updated. The filename is [CENSUS.DATA]DMFMAS.DAT.

The Locality File (LOC) contains a list of all known localities and is used for validation purposes. The filename is LOCMAS.DAT.

(ix) Census Reporting Software

2.25 SAS/DASE software was used to generate the various reports requested by the census office.

2.26 The data in the file census_master.dat was run through a sas program 'CECREAT.SAS' to create a permanent data set 'CETESTI.SSD'. This data set resides on disk7 in the directory user_disk7:[census]. All census table procedures access this data set to produce the different reports required.

Procedures for generating reports

- 2.27 All SAS procedures reside on disk5 in the directory USER_DISK5:[CENSUS]. Along with these are five permanently formatted files which are also input to the tables run.
- 2.28 The files are indfmt.sfn, provfmt.sfn, locfmt.sfn, majfmt.sfn and minfmt.sfn. A description of these files and tables for which they are required is supplied under the heading PERMANENT FORMAT FILES.
- 2.29 The input data set for all the tables is CETEST1.SSD. It contains a total of 843,252 observations.
- 2.30 The table programs are external files. They can be identified by the code 'CETAD' followed by a number representing the report that will be produced. The '.SAS' extension indicates that the program has SAS statements in it.

- 2.31 The SAS procedures can be run online or through batch processing. If several tables are to be produced then batch processing is recommended. Here the SAS programs are submitted in a command file and then executed. An external file "CETAB.COM" has been set up to allow tables to be queued for batch processing. The file also resides on USER DISK5: (CENSUS)
- 2.32 When the table program is run, two types of files are produced. The file names with extension "LIS" and "LOG" match the name of the program.
(EXAMPLE : If a sas program "cetabl.sas" was executed the output files would be "CETAB.LIS" and "CETABL.LOG").
- 2.33 The "LIS" file is an external file containing the output from the run. This is in the form of a report that can be despatched to the users. The reports for tables 5, 5a and 40a vary a little. These tables have the extension "DAT" instead of "LIS". For tabel 5 the reports are REPORT1.DAT and REPORT2.DAT. The outputs from table 5a program are REPORT1A.DAT and REPORT2A.DAT. The table 40a program produce REPORT40A.DAT AND REPORT40B.DAT.
- 2.34 The "log" file is also an external file containing the sas log notes and messages produced by the sas software. The success or failure of the table run can be checked on this log. This file must be checked before the .LIS or .DAT file is printed.

(xi) Security/Backup

2.35 There are 3 securities for the 1986 POPULATION CENSUS.

- . The batches of CENSUS data as it was keyed in is secured on to a tape labelled "CENSUS DATCHES". The volume lable is 3314.
- . Data that has been processed through the structural edit and edit input is secured on to a tape labelled "CENSUS-MASTER.DAT". The volume labels are 3315, 3316 AND 3343.
- . Census directory is backed up on to volume 700103. This contains all source programs, the .LIS files and also the log of all the runs.

(xii) Restore Facility

- 2.36 In order to restore any file from the different save sets above use the backup utility.
- 2.37 To re-create the data set for the table runs, the CENSUS MASTER.DAT is to be restored and the sas program CECREAT.SAS executed. The output is a permanent data set CETESTI.SSD and it is 350024 blocks.

(iv) Conclusion

- 3.1 Development of the software to process future census should be initiated at the time the planning for the census is started. EDP personnel should be an integral part of the census team and work closer to the statisticians.
- 3.2 Despite the problem of inadequate planning of the processing of data, the outcome of the processing stages at the EDP centre do achieved remarkable success with the co-operation between staff of the two offices. All the reports of the 1986 Census of Population were released within two years of the Census Date.

8 . APPROACH TO THE 1991 CENSUS OF INDIA

APPROACH TO THE 1991 CENSUS OF INDIA

Presented at the
Twelfth Population Census Conference
Beijing, China
September 5-9, 1988

V.S. Verma
Registrar General &
Census Commissioner, India

N. Rama Rao
Assistant Registrar General
(Census & Tabulation)

Office of the Registrar General, India
2A, Mansingh Road
New Delhi-110011

APPROACH TO THE 1991 CENSUS OF INDIA

INTRODUCTION

With a population of 685 million at the last Census in 1981, India ranks as the second most populous country in the world. It is a land of many races, religions and languages who live in amity in different geographical regions even while preserving their individual traditions and cultures.

The Indian Census is generally acknowledged as an authentic and comprehensive source of information about the land and people of India. The tradition of conducting a population census once in ten years is more than a century old in the country, though the population count as such, dates as far back as the third century B.C.

The first census was conducted in the year 1872 by the British rulers in different provinces at different times. In 1881, the practice of taking a synchronous census was initiated and since then the country has witnessed a census every ten years in the first year of each decade (i.e., years ending in one).

The 1991 Census will be the thirteenth decennial count of India's population, reckoned from 1872, and the fifth since Independence.

The decennial census has been providing the trends in the population size of the country and serving as a population barometer as it were, to the Government. The year 1921 is generally described as the 'great divide' in the history of Indian population growth and since then the size of population has shown an upward trend. Table-1 shows the size and growth of population of India since 1901.

TABLE - 1
GROWTH OF POPULATION

Year	Population (thousands)	Annual Growth Rate (per cent)
1901	238,396	-
1911	252,093	0.56
1921	251,321	-0.03
1931	278,977	1.06
1941	318,661	1.34
1951	361,088	1.26
1961	439,235	1.98
1971	548,160	2.24
1981*	685,185	2.28

*Includes projected population of Assam State.

The importance of reliable population data has increased with India attaining Independence and adopting a democratic system of government. The rapid growth of population as revealed by the censuses is sought to be checked through Family Welfare programmes. The Government of India was the first amongst the few Governments in the world, to adopt in 1952 family planning as a national policy and as an integral part of its socio-economic development plans. The size of the legislatures and the demarcation of electoral constituencies depend on statistical information about the population and its geographical distribution, provided by the census. In addition, these data are very useful in the implementation of certain constitutional provisions such as amelioration of the socially and economically backward classes.

India embarked on the path of planned economic development well nigh four decades ago with the First Five Year Plan in 1951. In this process of development, the size of the population and its growth rate are recognised as the most important factors influencing economic and social progress. The implementation of the five year plans called for a variety of census data at different

levels of administration with population growth, its distribution, urbanisation, manpower, employment, literacy, migration and housing. Information on working population, fertility patterns, age and education levels interest not only the planners and administrators but also research workers in different organisations including universities. Indian Census, in recent times has, therefore, shown a greater degree of sophistication in the methods of collecting, tabulating and analysing the statistical data.

LEGAL STATUS

Recognising the need for placing a legal obligation upon the public to cooperate and to give truthful answers to census questions and for placing a legal obligation upon the enumerator to record the responses faithfully, the Census Act, 1948, was placed on the statute book on 3rd December, 1948. After Independence, the censuses were conducted under the legal authority of the Census Act. The Act empowers the Central Government to take a census of the country after duly notifying their intention to do so. While the Act makes it obligatory for the public to answer all the questions faithfully, it guarantees confidentiality of information collected. There are also provisions in the Act for penalties in case of default.

CENSUS ORGANISATION

The population census is a Union subject under the Ministry of Home Affairs. The Office of the Registrar General and ex-officio Census Commissioner of India is attached to the Home Ministry. A senior officer of the Indian Administrative Service who has previously conducted a census in one of the major states is generally chosen to be the Registrar General and Census Commissioner. Since the registration of births and deaths and compilation of vital statistics are also overseen by this office under the Registration of Births and Deaths Act of 1969, the above mentioned officer is appointed Registrar General, India, to function concurrently as the census commissioner. He is assisted by deputies who head different branches of work within the office.

In each of the states and union territories, a senior officer is appointed as Director of Census Operations under the Census Act during the census period. For the major states the directors are generally drawn from the Indian Administrative Service. For the union territories and other areas, officers of the Indian Administrative Service or State Civil Service are appointed. The directors are responsible to the Registrar General and Census Commissioner of India in organising and supervising the census operations within their jurisdiction.

The directors of census operations are assisted by a number of deputy and assistant directors. During the peak census period these cadres are strengthened by drawing officers on deputation from the state civil service.

Though a Union subject, the census is successfully carried out by the cooperative effort of both the central and state Governments. The entire field operation, both the houselisting and enumeration, is organised through the general administrative machinery of the State Governments. The census enumerators are mostly drawn from school teachers, but they are also drawn from the staff of central and state Governments and local bodies. This type of administrative set up helps in keeping constant touch with the field agency and it is a matter of pride for the Indian census organisation that preliminary results are generally announced within a few days of the census count itself, thanks to this system. The credit for the success of the Indian census, which is the largest single administrative exercise in the country, goes mainly to the people of India themselves and to the vast army of devoted enumerators, supervisors, charge officers, district and city census officers who bear the brunt of the huge task at every stage.

CENSUS DATE

Censuses are generally taken with a given periodicity to facilitate study of change over time. Repeating the Census on the same date tends to eliminate the seasonal effect. However, in fixing the Census date, care has to be taken to see that there is minimum of mobility among the population over the period selected. The climatic conditions have also to be taken into account while fixing the Census date.

As regards the reference date for the Census, March 1, was generally adopted in the Censuses of 1941, 1951 and 1961. For the 1971 Census also March 1 was fixed as the reference date but as the mid-term polls to the Lok Sabha intervened, the date had to be changed to April 1. In 1981, the reference date was March 1, 1981. In the case of Jammu & Kashmir, however it was fixed as May 6, 1981 due to climatic conditions.

The houselisting operations which constitute the first phase when houses are numbered and listed, were completed in the course of 1980. It is proposed to have the next houselisting in 1990 and enumeration in 1991. It is also proposed to have sunrise of March 1, 1991 as reference date for the census. The enumeration would be made during February 9 to 28, with the enumeration of houseless people on the night of February 28. The revisional round would be made during March 1 to 5 for updating the data collected as at the reference time.

ENUMERATION METHODS

Of the two well known methods of enumeration, viz., canvasser method and householder method, the former is followed in India. Under this method, the enumerator visits every household and fills in the questionnaire based on the answers given by the

respondent. In the latter, the questionnaire is either mailed or supplied to the household for being filled up by the household itself. Owing to the low level of literacy in our country, the householder method can hardly be considered for adoption. Therefore, as in the previous Censuses, the 1991 Census will have to go by the canvasser method only.

PLACE OF ENUMERATION

In Census, information about each person can be collected where he is found at the time of the Census or at his usual residence. The first method is called a de facto approach while the second method is referred to as a de jure approach. In India, the earlier Censuses were taken by the de facto approach when a one-night enumeration was carried out for most parts of the country. This method will appear simpler but it required avoiding duplication or omission of persons during enumeration. On the other hand, the de jure approach required careful probing about one's usual place of residence. The de jure approach has the advantage of providing a better picture for the purposes of development and planning. Considering the pros and cons of both the methods, the Indian Census has made a compromise and has of late been taking Censuses by what is known as the 'extended de facto' approach. Under this method, an enumeration period of two or three weeks is fixed and the data are brought upto date with reference to the Census date by a revisional round after the enumeration period is over. The following extract from the 'Instructions to Enumerators' issued for the 1981 Census will explain how the extended de facto method is adopted (an enumerator had to cover):-

- i) All those who normally stay and are present in that household during the entire period of enumeration i.e., from February 9 to February 28, 1981(both days inclusive);

- ii) Also those who are known to be normally residing, and had actually stayed during a part of the enumeration period (February 9-28, 1981) but are not present at the time of the enumerator's visit;
- iii) Also those who are known to be normally residing and are not present at the time of enumerator's visit but are expected to return before February 28, 1981; and
- iv) Visitors who are present in household censused by the enumerator but are away from the place(s) of their usual residence during the entire enumeration period. For the purpose of enumeration such visitors will be treated as normal residents of the household where they were actually found during the enumeration period provided they have not been enumerated elsewhere.

The houseless population is generally counted on the last night of the enumeration period it was the night of February 28 in 1981. In the 1991 Census also, the extended de facto approach will be adopted.

ENUMERATION BLOCKS

At the time of 1981 Census, the Indian Union comprised 22 States and 9 Union Territories. It was divided into 412 districts. Each district was further sub-divided into sub-divisions, namely tahsils/talúks/police stations/anchals, etc. The smallest unit of census taking was a village or an urban block in a town.

The main object of the general enumeration scheme was to "catch" every person and catch him or her only once. This is the basic principle of census taking and it is easier to achieve

if the system fits into the ordinary district administration. Further, the size of the enumeration block must be governed by the ability of the enumerator, given the physical circumstances of his block, to complete the enumeration within the given period.

The unit of enumeration was an enumeration block. This was a village or a part of a village in rural areas and an urban block in urban areas that could be covered by a single enumerator in the 20 days period of enumeration. The average size of enumeration block in rural areas was 750 persons or approximately 150 households and in urban areas, about 600 persons or approximately 125 households. Large villages were divided into more than one block, but in no case did an enumeration block cut across the boundary of a town or village. It was also ensured that an enumeration block did not cut across the boundary of a ward in a town since ward-wise population data are required by the municipal or town authorities. Altogether there were more than a million enumeration blocks in the country in the 1981 Census. In the 1991 Census, there could be about 1.25 million enumeration blocks in the country.

PRETESTS

The schedules to be canvassed are generally finalised after pretests spread over the country. These pretests in the past served two purposes. They helped in highlighting the difficulties encountered. They also provided valuable field experience for the census office personnel. Apart from the two usual pretests, in 1981, there was a pilot survey to test the questions relating to the economic activity of the population. The analysis of the results of the pilot survey resulted in certain modifications in the content of the question proposed to be included. Two pre-tests of 1991 schedules are proposed to be held - one towards the end of 1988 and the other in the middle of 1989.

TRAINING AND FIELD WORK

Training the vast army of enumerators is a challenging task. In 1981, manual of instructions to enumerators and supervisors for filling the schedules was prepared at the headquarters and arrangements were made to print them in Hindi, English and regional languages. The training was organised in three phases. In the first phase, the census directors and their deputies in each state and union territory were imparted a thorough training. For this purpose, training seminars and conferences were organised by the office of the Registrar General and Census Commissioner, India during the preparatory stages of the 1981 census. The census directors and their deputies, in turn, took up the training of the district officers and the urban and rural census charge officers. The charge officers trained the enumerators and supervisors. Thus, it was ensured that the instructions for filling the census questionnaires permeated down to the grass-root levels.

A detailed programme of training was developed by each state director of census operations. Each enumerator was expected to attend four or five training classes which included practice enumeration classes as well. A similar procedure is proposed to be followed in 1991 census also.

COMMITTEE STRUCTURES

The preparation of the census schedules involve wide consultation with data users which include both Government departments and non-Government organisations. The economic questions included in the 1981 census were discussed and cleared by the technical advisory committee (TAC) on statistics of employment and unemployment constituted by the Department of Statistics prior to census. The TAC had members from Government and non-Government agencies. On the basis of these discussions, a preliminary version of the questionnaire was developed for field testing. The

results of the field testing of the questions on worker and non-worker were placed before the data users conference. A sub-committee of the data users went into the actual wordings of the questions to be included in the 1981 census. These were approved by the technical advisory committee and were included as part of the census. Questions relating to school attendance; seeking/available for work for non-workers and marginal workers, reason for migration, more questions on fertility in the Individual slip and simple information on amenities available to households were the new features of the 1981 census. A preliminary version of tabulation plan was also discussed in the data users conference. On the basis of the recommendations of the data users, the tabulation plan was finalised.

For the 1991 census the first such conference of data users inaugurated by the Union Home Minister was held in New Delhi during April 18-20, 1988. An Advisory Committee consisting of experts in the field (both official and non-official) has been set up for giving advice on technical issues connected with the 1991 census. Post Enumeration Checks and Census Evaluation Studies were done after the 1981 Census which may be undertaken in 1991 Census also.

CHANGES IN CONCEPTS OVER TIME

Important areas where changes have occurred are concepts relating to economic questions and rural-urban classification including the concepts of Urban Agglomeration and Standard Urban area.

ECONOMIC QUESTIONS

The concepts adopted to classify economic data has varied from census to census. Upto 1951 the classification was based on what is generally termed 'dependency approach' through the enumeration of the 'gainfully occupied', as against the 'work approach' resulting in production of economic goods and services, adopted in 1961, 1971 and 1981 censuses. According to the 'gainfully occupied' approach a person is identified as economically active if he considers himself to have gainful occupation.

In the censuses of 1901, 1911 and 1921 there was a dichotomy of population into 'actual workers' and 'dependents'. The term 'actual worker' included all persons who actually did work or carried out business whether personally or by means of servants or who lived on house rent, pension, etc. Persons who were not actual workers were dependents. In 1931, 1941 and 1951 censuses there was a trichotomy of population into what is broadly known as self-supporting persons, earning dependents and non-earning dependents. A self-supporting person was defined as one who worked or carried on business whether personally or by means of servant for which a return was paid in cash or kind, or who lived on house rent, pension, etc. A working dependent was one who worked and had an occupation but did not receive any income in cash or kind or did not control the means of subsistence gained. A non-working dependent was one who did not work and had no occupation. In the 1951 census, the terms used were self-supporting, earning dependent and non-earning dependent. In the 1941 census these were defined as independent worker, partly dependent person and wholly dependent person. From the 1961 census onwards, however, a person is classified on the basis of his work rather than means of livelihood. A person is classified as worker if his activity is in the nature of productive work.

Even though the 1961, 1971 and 1981 censuses used work as a basis for classifying a person, there were certain differences between the concepts used in the censuses. In 1961 a person qualified as a worker if he had worked regularly during the last season or if he had worked at least for a day in regular (nonseasonal) work during the preceding fortnight. In the 1971 census, a person was treated as a worker only if he spent his time mainly in work or if he worked at least for a day in regular (nonseasonal) work during the preceding week. The dichotomy of worker/nonworker of the 1961 and 1971 censuses was discarded, and a trichotomy consisting of main workers, marginal workers, and nonworkers was adopted for the 1981 census. For the main worker, a time criterion

of engagement in work for the major part of the year (i.e. at least 183 days) was adopted. Those who worked some time during the last year, but not for the major part, were treated as marginal workers. Those who had not worked at all during the last year were non-workers. This trichotomy was designed to permit comparability between economic data from the 1981 census and those from 1971 and 1961: the main workers of the 1981 census were expected to correspond to the workers of 1971; the main workers plus marginal workers of 1981, to the workers of 1961. In the 1991 census, the economic concepts of 1981 census are proposed to be retained for the sake of comparability, as per our present thinking.

DEFINITION OF RURAL AND URBAN AREAS

Prior to the 1961 census, there was no uniform definition of urban areas. The urban areas included, however, (1) Municipalities, (2) All civil lines not included within municipal limits, (3) Every cantonment and (4) Every other continuous collection of houses inhabited by not less than 5,000 persons which the provincial superintendent might decide to treat as a town for census purposes. In native states, where there were no municipalities, this definition was to be exhaustively applied.

In dealing with the question arising under category (4) above, the provincial census superintendent had to keep in mind the character of the population, relative density of the dwelling, importance of the place as a centre of trade and its historic associations and had to keep in mind that it was undesirable to treat as towns, overgrown villages which had no urban characteristics.

From the 1961 census onwards, the following definition was adopted for an urban area: (1) Municipal corporation or town committee or notified area committee or cantonment board, (2) Selected places with a minimum population of 5,000 and a density of not less than 400 persons per sq. km. and three-fourth of the

working population engaged in non-agricultural occupation or any other place which, according to the superintendent of the state, possesses pronounced urban characteristics and amenities. The latter discretion was provided to accommodate new emerging industrial townships.

In 1971 and 1981 the criterion adopted was essentially the same except that it was stipulated that at least 75 per cent of the male working population was to be in non-agricultural occupation. The urban criterion adopted for the 1981 census varied slightly from that of 1961 census. In the 1961 and 1971 censuses, males working in activities such as fishing, logging, etc., were treated as engaged in non-agricultural activity. In the 1981 census, however these activities were treated on par with cultivation and agricultural labour for the purposes of this criterion. In the 1991 census, the criterion adopted in 1981 census regarding urban area will be continued.

URBAN AGGLOMERATION

In 1971 for the first time, the concept of urban agglomeration was used in the place of town groups, apart from individual towns and cities. Very often large railway colonies, university campus, port areas, military camps, etc., come up outside the statutory limits of the city or town but adjoining it. Such areas may not by themselves qualify to be treated as towns but if they form a continuous spread with the town, they are outgrowths of the town and deserve to be treated as urban. Such towns together with their outgrowths have been treated as one urban unit called 'urban agglomeration'. An urban agglomeration may constitute:

- (a) A city with continuous out-growth (the part of out growth being outside the statutory limits but falling within the boundaries of the adjoining village or villages);

- (b) One town with similar outgrowth or two or more adjoining towns with their outgrowths as in (a); or
- (c) A city and one or more adjoining towns with their outgrowths all of which form a continuous spread.

This concept was continued to be used in 1981 and is expected to continue in 1991.

STANDARD URBAN AREA

A new concept of standard urban area introduced in the 1971 census was also followed for the 1981 census. The essential requirements for the constitution of a standard urban area were:

- (i) It should have a core town of a minimum population of 50,000.
- (ii) The contiguous areas made up of other urban as well as rural administrative units should have mutual socio-economic links with the core town, and
- (iii) In all probability this entire area should get fully urbanised in a period of two or three decades.

Certain standard urban areas were determined on this basis in 1971 and some basic data were presented for 1951, 1961 and 1971 for such areas and their components. Similar data were presented for the standard urban areas in 1981 also. The idea was to present basic data for these areas for four to five decades so that the urbanisation process in those areas could be studied. However, some minimum changes in the standard urban areas of 1981 census as compared to those of 1971 were inevitable. In the 1991 census also this concept will be continued with some minimum changes in the frame wherever necessary.

USE OF SAMPLING

Sampling has been used in the Indian censuses from time to time for tabulation purposes. Sampling for tabulation was adopted as early as 1941, due to stringent financial restrictions. A two per cent systematic sample of individual slips was selected and data on age and means of livelihood were based on this sample. In the 1951 census, while many of the tabulations were made from cent per cent data, some detailed cross classifications of age and civil conditions were made from a 10 per cent sample of individual slips selected systematically. Again in the 1961 census, household economic data collected in the household schedules were tabulated on a 20 per cent sample basis.

In the 1971 census, extensive use was made of sampling for tabulation. Most of the B, C, D and F series tables relating to economic, social, cultural, migration and fertility aspects of the population were tabulated on a sample basis. The sample size was 10 per cent in rural areas and 20 per cent in urban areas. The sample was selected systematically after arranging the individual slips by age, sex and industrial categories at stratum level. The strata adopted for sample selection were city, non-city urban areas of a district and rural areas of tehsil. Apart from this, an advance tabulation at national level was undertaken on the basis of one per cent sample of individual slips. For housing tabulations, a 20 per cent systematic sample of census houses was drawn from the houselist separately for rural and urban areas. The strata adopted were city, non-city urban areas of a district and rural areas of each tehsil.

In the 1981 Census, for the first time in the history of the Indian census, sampling was adopted at the enumeration stage to broaden the scope of the census programme. The sample enumeration was simultaneously carried out with complete enumeration. Two

slips were canvassed. These were termed universal and sample slips. The universal slip was canvassed for every individual. The sample slip was canvassed for all persons living in the 20 per cent sample of enumeration blocks in the major states i.e., all the states with an estimated population of 10 million and above in 1981. In all other states and union territories, both the universal and sample slips were canvassed for all the individuals.

The universal slip collected particulars on relationship to head of household, age, sex, marital status, religion, mother tongue, other languages spoken, whether an individual belonged to scheduled caste or tribe, whether a person was worker or non-worker with details of industry and occupation of the worker. The sample individual slip contained information on migration by place of birth and place of last residence, duration of residence at place of enumeration, reasons for migration. It also contained four questions relating to fertility. These pertained to age at marriage, number of surviving children, number of children ever born and whether any child ever born during the last year. The first three questions were canvassed for all ever married women while the last one was canvassed only in respect of currently married women.

In the 1981 census sampling was also used at the tabulation stage and this is explained later while discussing tabulations.

MAPPING

Mapping has been an important adjunct of census taking in India. The Census organisation is a major producer of maps in India. To achieve proper coverage in the Census, it was necessary to have complete and upto-date maps of all administrative areas - states and union territories, divisions, districts, tehsils and villages. Maps of proper scale were obtained and the boundaries of each administrative area were identified. In order to simplify the map-making tasks and to make it possible to update all maps before the census was conducted, administrative boundaries were frozen a year before the census.

After identifying every village and town on each tehsil map, code numbers were allotted to the units that were to be accounted for at the census. Notional maps were prepared of every village and of every ward and block of towns. These notional maps showed the location of every house to be visited and proved extremely helpful at the time of enumeration. They served as the frame for the clear demarcation of each enumerator's jurisdiction and enabled him to cover every household in his block and to take notice of additions that might have taken place after the preparation of the notional map. Complete coverage of areas was thus ensured. Preliminary work on mapping for the 1991 census has already been taken up by the mapping units in the Directorates of Census Operations in States and Union Territories under the overall guidance of the Map division in the headquarters.

DATA PROCESSING

Prior to the 1971 Census, the entire census data was processed manually. Computers were introduced for data processing for the first time in the 1971 Census. At that time, though tabulations of Individual Slip data for urban areas (based on 20 per cent sample) were made on computer, the rural data, based on a 10 per cent sample of individual slips were processed manually. For the first time in the 1981 Census, most of the data were processed on computers, barring some minimal manual compilations. In the 1981 Census, direct data entry systems were used to expedite the processing.

In the 1971 Census, 49 basic tables were prepared. In the 1981 Census the number of basic tables increased to 111. While in 1971 all the tables for rural areas and a number of tables for urban areas were prepared on manual compilation basis, in the 1981 census only five basic tables were compiled manually. Remaining tables were prepared by using computers. While in 1971, only 20 per cent sample of individual slips in urban areas was processed on computers, in 1981, all slips on 20 per cent sample of enumeration blocks were processed on computer.

In 1981, schedules were manually edited and coded in different Directorates and then sent to Direct Data Entry Centres (DDE) for keying in of the data. There are 32 Direct Data Entry Systems with 16 terminals each. These centres work in two shifts. The data were keyed-in there.

The keyed-in data received from DDE centres were brought to the headquarters for main frame computer processing. Initially from DDE output tapes, master tape files were generated after ensuring that record counts on the master tape tally with the manual inventory counts.

The master tape was then edited and auto-corrected with the help of a very comprehensive edit and imputation programme. This programme was developed on the basis of a detailed list of range and consistency checks formulated by a team of data processing and subject matter specialists after elaborate discussions.

The fully edited data, together with the estimation factors, were then processed for getting estimated summary tabulation figures and control figures. The summary data were then aggregated at various levels of tabulations (district, State and all India levels) and printed with appropriate headings and titles. This was the final computer printout.

Because of the large volume of data (about 218 million records) and complex nature of the tables, 1981 Census data were processed on seven computers, namely CYBER-720, HP-1000, Tata-Burrows, IBM-370, PDP-11 and IBM-1401, all located in Delhi and DEC System at Chandigarh. This needed high level of programming and processing expertise like conversion of software for each computer system, movement of large volume of data tapes from one system to another, detailed monitoring of processing at various computer installations, scrutiny of the intermediate outputs for

ensuring accuracy and thorough scrutiny of final computer printouts before they were finally released to data users in book form. All this involved rapport between subject matter specialists and data processing experts as well as close coordination between the census office and computer installations. Lack of an in-house computer in the Census Organisation proved a major handicap in the processing of 1981 census data. Steps are being taken to acquire for the Organisation a main frame computer system well in advance of the 1990 Houselisting.

TABULATIONS

The following details are furnished in respect of the tabulation done on the basis of the 1981 Census data:-

1. Table H-1 on the use to which the Census Houses were put was processed on the computer based on 20% sample of Census Houses and released in July, 1982.
2. Table H-2 on disabled population was processed manually on 100% basis and released in March, 1983.
3. Provisional results giving figures of population and literates were announced within a few days after the completion of Census enumeration. The details of rural-urban break up and economic characteristics of population were released a few months thereafter.
4. An advance tabulation based on the sample of 5 per cent enumeration blocks for the country and the major States was done on the computer for the immediate use of the data users and planners. The tables based on 5 per cent sample were released in October, 1983.

5. The A-Series tables on general population (including the Primary Census Abstract) were manually tabulated and published for India/States/Union Territories.
6. District Census Handbooks giving basic population data at village/urban ward level and amenities available in these areas were processed manually and finalised for all the States and Union Territories by end of 1986. Majority of these volumes have been printed while a few are in the press.
7. Tabulation of religion and language data based on the religion of the head of household and the language usually spoken in the household, respectively was manually done and published.
8. Individual Slips of 20% sample of enumeration blocks in respect of bigger States and 100% blocks in respect of smaller States and Union Territories with less than 10 million population were processed on the computer for generating the B-Series - General Economic Tables, C-Series - Social and Cultural Tables, D-Series - Migration Tables, F-Series - Fertility Tables and HH-Series Household Tables. The State/Union Territory volumes containing economic, social and cultural, migration and fertility tables have been published. The volumes on All-India HH-Series tables on households are in the press. Household tables of States/Union Territories have been printed in most cases while the remaining are at an advanced stage of printing. The all India volumes containing economic, social and cultural, migration and fertility tables have been brought out.

9. The Special Tables for Scheduled Castes and Scheduled Tribes for the SC-Series and ST-Series are getting printed State by State. Some selected tables of the B and C Series are also being generated on the basis of 20 per cent systematic sample of Individual Slips ('P' Slips).

In the 1991 census it is proposed to bring out some basic tables like the Primary Census Abstract and tables on religion and language by manual tabulation of household schedules and simultaneously process a sample of individual slips for computerisation to bring out all the remaining tables. The houselist will also be processed on the computer on a sample basis.

STATISTICAL QUALITY CONTROL(SQC)

For the first time in the history of the Indian Census, SQC was adopted in 1981, with a view to achieve an adequate level of accuracy of coding the schedules. This project was somewhat exploratory in nature; it aimed at providing useful information as a guide to full scale implementation of the technique in subsequent Censuses. The application of the SQC was restricted to the coding of responses to the economic questions on the individual slips. Further, this SQC was restricted to certain areas only.

The SQC scheme was designed to achieve several objectives. In general, it was designed to ensure that data of a satisfactory quality for electronic data processing and tabulation are obtained effectively and economically. Specifically, its primary purpose was to screen out errors made in the course of coding so that only a tolerable proportion remain in the data passed for further processing. It is designed to achieve this with the minimum amount of checking.

By using sampling techniques, much of the checking that was necessary with a 100 per cent check could be avoided. The use of sampling implied, however, a risk that some proportion of the unchecked coding would be in error. Keeping this proportion at a tolerable level was the object of the quality control exercises. Even with a 100 per cent check, not all errors could be detected.

The quality control check allowed substantial precision in assessing the accuracy of the census results. It increased the depth of understanding of the causes of errors arising in coding census data and suggested measures to reduce these errors in future Censuses.

The first experiment in the 1981 census in SQC in coding of schedules had brought out that if SQC was adopted in controlling the quality of coding, routine extra checks might be dispensed with. This would result in considerable savings in terms of money and man-power requirements. This aspect will be considered in depth while organising the editing and coding operations in the 1991 census.

A REVIEW OF 1981 CENSUS AND APPROACH TO THE 1991 CENSUS

At the 1981 census three schedules, namely the houselist, the household schedule and the individual slip were canvassed. An enterprise list was canvassed alongwith the houselist in 1980 as part of the Second Economic Census for and on behalf of the Central Statistical Organisation(CSO). Degree Holders and Technical Personnel(DHTP) cards were given to eligible persons in the 1981 census for filling up. After filling, these were either handed over to the enumerators during revisional round or mailed directly to the Council of Scientific and Industrial Research(CSIR), who processed them. The individual slip of 1981 census was the key schedule for each individual.

The schedules in the 1981 Census and the instructions for filling these were translated into various regional languages, printed in adequate numbers and supplied to the field staff for carrying out the Census. As much as 6000 MT of paper was consumed in printing the schedules, the instruction booklet, etc. The requirement of paper for the 1991 Census is likely to be much more than this. While generally an enumeration block was in the charge of an enumerator, a supervisor was appointed for every 5 enumerators. On the basis of this criterion, about a million and a quarter enumerators and supervisors were appointed and trained to carry out the 1981 Census. They were paid an honorarium ranging from Rs.100/- to Rs.120/- for the entire enumeration. Though no decision has yet been taken about the quantum of honorarium to be paid to enumerators and supervisors in 1991, it will definitely have to be higher than what it was in 1981.

Planning the census in a vast country like India cannot be rushed. The scale of the operations and the importance of the results require careful planning of each stage of the census operations. Since one individual slip will be canvassed and filled in for every single person in the country, irrespective of sex or age, the census operations would, in 1991, call for the handling of about 840 million slips, 170 million household schedules, innumerable charge registers, maps, houselists, etc. It is estimated that about 1.6 million enumerators and other staff will be involved in the field operations, calling for vast training programmes, distribution systems for the forms, extensive storage and filing systems, etc.

The content of the census questionnaires is governed by certain factors. These include the need to collect basic demographic data bearing in mind the U.N. recommendations, the desirability and comparability with previous censuses and the special needs of data users, including government departments. The demands for data by the data users cover a wide range of topics based

on their specific needs and it is quite often necessary to determine the priorities of such demands and the feasibility of meeting them. In considering the variety of demands for data certain constraints operate. The census organisation has to examine (a) what topics can reasonably be included in a widespread and vast field operation like the census, taking into consideration the capacity of the enumerator and the reactions and comprehension of the respondent; (b) what topics can be canvassed in quick visits to households, given the time constraints of the census; (c) what questions would be acceptable and not evoke adverse reaction; (d) how much money can be spent since every extra question has a cost in terms of printing, paper, storage and handling for tabulation; (e) can any of these demands be met from other sources or by other agencies.

In determining the topics that are included in the questionnaires, certain standards have to be applied. If the census has to be operationally successful and its results beyond doubt, the questions included in the questionnaires have to be factual and clear. Only those questions which can be expected to be easily understood and reliably answered, can be included. Questions that are difficult to ask or are an intrusion of privacy beyond acceptable limits cannot be asked. Equally important are the workload of the enumerator and the cooperation of the public. The enumerator in the Indian Census is a public official, usually the school teacher, who is pressed into service and paid an honorarium for the work he performs as a census enumerator, and he is expected to visit each household for canvassing the census schedules, ask questions and note down the answers. This he has to do within a specified time limit. It is evident that if the enumerator has to perform his work successfully, the workload has to be carefully assessed and trimmed. Otherwise, inevitably, the enumeration will be rushed and information recorded may be unreliable or incomplete. Simplicity and clarity in the census questions are important not only to avoid overloading of the enumerator, but also to ensure public cooperation and response. A heavily loaded questionnaire is most unlikely to be received kindly by the public.

The need for financial controls in the census is obvious - the operations are so extensive that unless there is careful financial planning, the overheads can be heavy. The field costs - namely the costs relating to printing, distribution and storage of the forms, the training of the enumerators, the payment of honoraria and essential staff costs - are relatively inelastic and are almost of the nature of fixed costs. But the processing costs are more elastic since the scope of the tabulations, the depth of the analysis of data, the scale of publications and the time over which tabulation and publication can be spread are capable of some choice. However, these costs can only be said to be relatively elastic because there would have to be a certain minimum tabulation for effective utilisation of the data and these would have to be published quickly if the data are not to be quoted as out of date. The length of the questionnaires and the topics included therein become important from the point of view of costs also.

The 1991 Census of India is being planned within the broad framework indicated. In view of the gigantic scale of the census operations in India, the planning process for the Census commences years ahead. Even from the middle of 1987, the census directorates in the state/union territories had commenced the preliminary work relating to the monitoring of changes in the administrative jurisdictions since the 1981 census, the preparation of list of villages in consultation with the state governments, updating of district and tehsil maps, etc.

Suggestions relating to the contents and coverage of the 1991 census received from individuals and organisations have been systematically listed for consideration at the appropriate time.

Draft questionnaires for the 1991 census were developed on the basis of past experience, current needs and international recommendations and these were presented at the First Data Users' Conference (April, 1988). The schedules developed were the houselist, household schedule and the individual slip (Annexure I, II & III) as was done in the 1981 census.

In the 1981 Census the individual slip contained two parts, i.e., universal and sample. The sample portion of the individual slip contained questions on 'migration' and 'fertility'. This was filled in 20 per cent sample of the enumeration blocks. Preliminary thinking is that in the 1991 Census area sampling is not considered feasible due to certain practical difficulties experienced in the last census. This will be, however, reviewed again. The 1981 census questions and the concepts were mostly retained and the questions in the part 1 of the household schedule of 1981 Census were transferred to the houselist of 1990 so as to save time and quicken the publication of results relating to houses and household characteristics. Questions on school attendance, seeking or available for work and any child born in the last one year to currently married women are proposed to be dropped.

The data users discussed threadbare the questionnaires under the chairmanship of the Registrar General, India, and made many valuable suggestions. Some data users made strong cases for the retention of the above mentioned questions which were proposed to be dropped. Further, with regard to the economic questions, it was pointed out that as per the current census definitions women's participation in workforce is not properly reflected. Productive activities are not considered as work if the article produced is consumed in the household itself and not sold. Another important suggestion was that in the definition of cultivation of growing of crops like coconut, betelnut, etc., and vegetable and fruit crops are not taken under cultivation. There was a suggestion that efforts should be made to see that workers growing these crops are also treated as cultivators. There was also a demand that the Primary Census Abstract should be brought out in the following nine categories as was done in the 1961 and 1971 censuses; 1. cultivators, 2. agricultural labourers, 3. livestock, forestry, fishing, hunting and plantations, orchards and allied activities, 4. mining and quarrying, 5. manufacturing, processing, servicing and repairs, 6. constructions, 7. trade and commerce, 8. transport

storage and communications and 9. other services. In the 1981 Census the village-wise and town-wise Primary Census Abstract contained only the following four categories, namely 1. cultivators, 2. agricultural labourers, 3. household industrial workers and 4. other workers.

The present position regarding the 1991 census schedules is that taking into account the suggestions made by the data users and also the views of the Advisory Committee which has been set up these schedules will be finalised. There is also a strong suggestion that instead of individual slip and household schedule being two separate schedules they may be brought into one place in the form of a 'register'. This will also be considered by the Advisory Committee and may also be pre-tested in the field.

DISSEMINATION

The Census data are disseminated through a number of publications brought out by the Office of the Registrar General, India. The headquarters office at Delhi, publishes volumes relating to India as a whole. These are termed Series 1. Data relating to each state and union territory are published in separate series. These are numbered Series - 2 to Series - 32. In each of the Series (for all India and each state/union territory) different series of tables are brought out in separate volumes. These are termed: A Series - General Population Tables, B - Series - General Economic Tables, C - Series - Social and Cultural Tables, D - Series - Migration Tables, F - Series - Fertility Tables, H-Series - Housing Tables, HH -Series - Household Tables and SC/ST-Series - Special Tables for Scheduled Castes and Scheduled Tribes. Apart from these publications containing the census tables the following publications are brought out; Reports on selected towns and villages based on surveys, craft studies, ethnographic notes on Scheduled Castes and Scheduled Tribes, Census Atlas both at all India and State levels. Further, census-papers, occasional papers and monographs each devoted to an important topic are also brought out.

It is proposed to release copies of tapes containing the census data after blanking off certain individual details in view of the confidentiality of the schedules. In the 1991 Census, the same procedure for dissemination may be adopted with increased emphasis on release of data tapes.

PUBLICITY

The success of census enumeration depends, to a great extent, on wide publicity explaining the importance and objectives of the census to the public well in advance. A vast publicity was mounted before the 1981 census with the help of the All India Radio, Doordarshan(TV), the films Division of the Ministry of Information and Broadcasting, the directorate of advertising and visual publicity and the press itself. Publicity regarding the census operations included imaginative posters, slides and films. Interesting informative skits were also broadcast over the All India Radio and Doordarshan. The media carried the census messages to all parts of the country.

In the 1991 Census also, major publicity has to be mounted. This will have to start even from the houselisting stage in 1990.

ANCILLARY STUDIES

The Indian census has a tradition of going beyond the analysis of data collected during the census and also to collect, analyse and publish information lying in various segments of life and disciplines and bring out special studies based thereon. Among other things, this endeavour takes the form of a study of selected towns, villages and crafts during the inter-censal period. A large number of such studies taken as part of ancillary studies in 1981 census are at different stages of completion in the State directorates and in the headquarters. The Indian Census has also the tradition

of undertaking ethnographic studies. But the studies conducted before Independence had their focus more on the exotic customs and manners followed by different castes and communities rather than on the social and economic aspects. After Independence, studies on Scheduled Castes and Scheduled Tribes were continued so as to collect data relating to different socio-economic aspects for enabling the Government to discharge its constitutional obligations towards them. In the post-1991 census period, it is tentatively proposed to concentrate more on ethnographic studies on lesser known tribes, Compendium on Scheduled Tribes and Glossary notes on Scheduled Castes.

The District Census Handbooks are by far one of the most valuable products of Census operations. With their micro-level data down to the lowest administrative unit, i.e., village and ward these publications are constantly referred to by planners and administrators for several purposes like delineation of electoral constituencies, adjustment of boundaries, formulation of local and regional plans, etc. The publication of District Census Handbooks began in the 1951 Census and has been continued with some improvement from census to census. In the 1981 Census, the District Census Handbooks contained two parts, Part (a) comprising village and town directories and Part (b) containing the village and town primary census abstract. It also contained the primary census abstract for Scheduled Castes and Scheduled Tribes at tehsil and town levels. In the 1981 census the state volume of the town directory which was earlier brought out in 1971 and found useful to administrators, planners and research workers is continued to be published. An all India town directory is also brought out on the basis of 1981 census data. Several studies under the plan schemes were also taken up during the inter-censal period after 1981 census. Such studies will be taken in the inter-censal period of 1991 census also with such modifications as may be deemed fit.

A N N E X U R E S

Pad No.....

Slip No.....

Location Code.....() Sl. No. of Household []

1 Name			
2 Relationship to head		8 Religion	
3 Male (1)/Female (2)		9 Whether S.C. (1) or S.T. (2)	
4 Age		10 Name of S.C./S.T.	
5 Marital status		11 Literate (1)/Illiterate (2)	
6 Mother tongue		12 Educational attainment	
7 Two other languages known			

13A Worked any time at all last year ? Yes
No (H/ST/D/R/B/I/O)

13B If yes in 13A, did you work for major part of last year ? Yes (1)/No (2)

14A Main activity last year ? Yes in 13B (C/AL/HHI/OW)
No in 13B (H/ST/D/R/B/I/O)

If HHI/OW in 14A

- (i) Name of establishment
- (ii) Description of work
- (iii) Nature of industry, trade or service
- (iv) Class of worker

13B Yes—Any other work any time last year ? Yes (C/AL/HHI/OW)/No
14B 13 B No—Work done any time last year ? (C/AL/HHI/OW)

If HHI/OW in 14B

- (i) Name of establishment
- (ii) Description of work
- (iii) Nature of industry, trade or service
- (iv) Class of worker

15A Whether you are an Ex-serviceman? Yes (1)/No (2)

15B If yes in 15A, Pensioner (1)/Non-Pensioner (2)

16 Birth Place

(a) Place of birth

(b) Rural (1)/Urban (2)

(c) District

(d) State/Country

17 Last residence

(a) Place of last residence

(b) Rural (1)/Urban (2)

(c) District

(d) State/Country

18. Reasons for migration from place of last residence (Code)*

19. Duration of residence at the village or town of enumeration

20 For all ever-married women only

(a) Age at marriage

(b) Number of children surviving at present

Male Female Total

(c) Number of children ever born alive

Male Female Total

*Employment (1) Education (2) Family moved (3) Marriage (4) Others (5)

9 . A NOTE ON THE METHODOLOGY OF THE
1990 POPULATION CENSUS OF INDONESIA

A note on the methodology of the
1990 Population Census of Indonesia

Azwar Rasjid

1. Introduction

Preparatory work for the Indonesian Population Census of 1990 had already begun in 1987. The forthcoming census is the fourth carried out since Indonesia proclaimed its independence in 1945. The first three censuses were taken in 1961, 1971 and 1980. All of these censuses covered all residents of the Republic of Indonesia, regardless of permanent residence. Enumeration of ships, crews, isolated population groups, and homeless persons was conducted simultaneously on the census date, and limited to information on age and sex. Foreigners were also recorded in the census, except members of the diplomatic corp and their families.

Basically, in all censuses persons are recorded on a de jure basis, i.e. in their place of usual residence. However, persons who do not have a permanent residence and persons who have more than one residence are recorded in the place where they are found by the census enumerator [de facto]. The forthcoming census-taking will commence in mid September 1988 and last through 1 November 1988. October 31, 1988 is established as the census date.

Over the years, census-taking in Indonesia has undergone a lot of enhancement. This includes improvements in sampling design, field organization, the use of more refined concepts and definitions, more efficient data processing techniques, publication and data dissemination. Despite the application of varying procedures in census taking, comparability of results has been maintained to the maximum.

Presented at the Twelfth Population Census Conference.
Beijing, China. 5-9 September 1988.

2. Overall Plan

Census activities are planned to be carried out in 5 fiscal years. The annual plan in broad detail is as follows:

1987/1988 fiscal year: Planning and preparation of the census, including several pilot tests in statistical area mapping.

1988/1989 fiscal years:

- a. Preparation of mapping of all regencies/municipalities (kabupaten/kotamadya), sub-districts (kecamatan), and villages (desa) as well as all enumeration areas in the capitals of provinces and municipalities.
- b. General rehearsal of mapping, and pilot tests for the enumeration (complete and sample census).
- c. Study of border line cases of villages to determine the urban/rural classification.
- d. Construction of master sampling frame.
- e. Implementation of mapping.

1989/1990 fiscal year:

- a. Planning and preparation of mapping in 20 percent of areas outside the capitals of provinces and municipalities.
- b. Construction of master sampling frame.
- c. Planning and preparation of complete and sample census.

1990/1991 fiscal year:

- a. Checking of village and enumeration area sketches.
- b. Implementation of complete and sample enumeration. At the same time, information on village potentials will be collected. Release of population counts.
- c. Post Enumeration Survey.

1991/1992 fiscal year: Data processing and evaluation of census taking in general, and data quality in particular.

1992/1993 fiscal year: Publication and analysis.

3. Mapping and Formation of Enumeration Areas

Formation of enumeration areas and mapping of these areas are necessary tools to achieve successful count of all persons in the census. The primary objective of statistical area mapping is to identify these areas, particularly their boundaries. This is aimed at facilitating field work in an attempt to minimize omissions and double counting. Another use of the maps is in connection with the master sampling frame, which is used as a basis in drawing samples for future household surveys.

Prior to the 1980 Population Census, two types of statistical units were formed. The area of each of the villages in the country was classified into enumeration areas (EA), and each EA was broken down further into census blocks. An EA was delineated by permanent/natural or clear boundaries, and consisted of 200 to 300 households. To meet the requirements of a census block, an area had to have a clear boundary and comprised of a maximum of 100 households. On average, a census block consisted of 70 households.

Although theoretically vacant areas should have been included in a census block, in many cases they were left out, and were not part of any census block. When areas such as these were settled, they presented a problem because they did not belong to any of the existing census blocks. Since 1980 there has been a lot of changes due to in and out migration of the people, development of buildings or housing complexes, construction of public utilities, and changes in administrative status which generally involved shifting of boundaries. Some census block gained, others lost their

population. Also, since one census block covers a very small area, clear and permanent boundaries were not always present, and it was not possible to maintain these census blocks during the intercensal period.

Basically, the statistical maps are based on administrative/geopolitical areas. As such, the actual boundaries are often unclear, and may not coincide with easily recognized landmarks such as roads or rivers. In consideration of this problem, the Central Bureau of Statistics decided to modify the mapping procedure by adopting the following rules:

- a. Administrative boundaries are used for regencies/municipalities and sub districts.
- b. For villages and EAs both administrative and natural/permanent boundaries are drawn on the map.

EA maps drawn for the census are expected to be used in the following activities:

- a. As a guide for the complete census enumerators in identifying the EAs assigned to them. This is an important aspect in census taking, as the maps should enable enumerators to avoid miscoverage of households in their EA.
- b. As a basis for drawing samples for sample census, where EAs are the first stage sampling unit and the households or cluster of households are the second stage sampling units.

4. Urban/rural classification

The criteria for classifying villages into urban/rural areas is the same as that used in the 1980 census. This allows for a comparative study across time. The classification is based on scores computed from population density, percentage of households engaged in agriculture, and availability of urban facilities such as schools, roads,

electricity and hospital.

Prior to the enumeration, villages which may have undergone a change since 1980 are evaluated. The change may be brought about by a shift in boundaries, change in administrative status, or by urbanization. Included in this evaluation are villages which scored between 17 and 22 in 1986. (Based on the criteria currently used in the urban/rural classification, a village is automatically regarded as urban if it scores 22 or above, and rural if the score is 17 or less). Altogether, 9,200 villages will be evaluated in 1988.

5. Complete enumeration

The complete census will be carried out as a door-to-door canvass except for homeless people, isolated population groups, and ships' crews who do not have residence on land. To ascertain successful enumeration, EA borders determined during the mapping stage have to be maintained. To help identify buildings, a household listing will be carried out prior to complete census enumeration in which each building will be numbered sequentially with a sticker.

With the aid of EA maps, the stickers and list of households, coverage error in the complete census is expected to be minimal. During the enumeration, the field workers have to check the completeness and accuracy of the household list. New buildings should be added on the list and given stickers, while buildings/households which do not exist any more are crossed out.

The content of the complete census will be substantially simplified. It is limited to basic demographic characteristics, namely age, sex, relationship to head of household, and marital status. Traditionally, the complete census forms were processed manually in stages by preparing summaries at the EA, village, sub district, district and

provincial levels. However, since micro computers are available in the provincial statistics offices, plans are being made to process these summaries mechanically.

6. *Sample census*

A long form is going to be used in the sample census will content similar to that used in the 1980 census. However, several questions are taken out in consideration of the fact that they are more appropriate in sample surveys where more detailed information can be asked.

The overall sampling fraction for the sample enumeration is 5 percent, covering about 1.5 million households. However, the sampling rate varies between 2.5 and 33 percent across regions, mainly to produce district level fertility and mortality estimates with reasonable reliability in all areas. The following table displays the sampling fraction according to the size of the regency/municipality.

<i>Size of reg/mun (no. of hhs) x 1000</i>	<i>Overall sampling fraction (%)</i>
- 6	30
6 - 10	30
10 - 15	33.3
15 - 25	20
25 - 50	10
50 - 100	5
100 +	2.5
<i>Total</i>	5

Note. In some provinces, urban areas will be oversampled in comparison to the above to achieve a certain minimum sample size.

In consequence of the application of different sampling rates, a minimum sample size has to be established for each district (regency/municipality). This has some bearing on the allocation of funds. The adoption of this procedure has to take into consideration efficiency of the field operation, especially in areas which have difficult terrain.

A more crucial problem is the determination of a minimum sample size to yield a reliable estimate of mortality measure, e.g. CDR. Assuming a CDR of 10 per 1,000 population, with 95 % confidence interval (+/- 1.5 per 1,000), the required sample size is 18,000 persons or 3,600 households.

Data will be presented for each of the 27 provinces, classified by urban and rural areas. At district level, the census results will be available for the total population without urban/rural breakdown, because many urban areas in the regencies are too small to be presented separately with acceptable precision. In this case it is necessary to ensure that for each urban area the sample does not fall below a certain minimum.

In the sample census each regency/municipality is considered a domain of study. Two stage sampling design is employed. In the first stage, EAs are selected systematically after the sub districts, villages and EAs are arranged geographically within each regency/municipality. Two alternatives are being considered for the second stage sampling which involves selection of households. The first option is to draw households systematically. This method may be adopted in the urban areas where the buildings are located close to each other such that canvassing the whole EA will not create transportation problems. Another option is to select a group of households which forms a cluster. This method is more appropriate in the rural areas where households are more spread out within the EA limits.

A self-weighting design is adopted for regency/municipality, and thus allows for the application of simple ratio estimation procedure. This way the data generated from sample census can be kept consistent with the complete census result at a particular administrative level.

Based on 1980 Population Census data, the distribution of regencies/municipalities by the number of households is as presented in the following table.

*Percentage distribution of regency/municipality
by number of households*

<i>Classification of number of hhs per reg/mun</i>	<i>Number of regency/ municipality</i>	<i>Number of households</i>	<i>Percentage distribution (%)</i>
- 9,999	16	102,385	0.34
10,000 - 14,999	20	245,292	0.81
15,000 - 24,999	24	470,963	1.55
25,000 - 49,999	62	2,281,190	7.51
50,000 - 99,999	63	4,381,243	14.42
100,000 +	115	22,891,390	75.37
<i>T o t a l</i>	300	30,372,463	100.00

Data in the above table is used as a guide in determining the sampling rate in each regency/municipality. Those which have less than 50,000 households will account for 10 percent of all districts in Indonesia, most are located in "difficult" areas in East Timor, Irian Jaya and some islands in Sumatra. In consideration of efficiency in cost while maintaining the minimum sample size, the maximum sample size in these areas is set at 20 percent. Also, too large a sample may create problems in recruiting qualified field workers.

7. *Master sampling frame*

To obtain a complete sampling frame to be used in selecting samples in the census and future household surveys, several supplementary activities are going to be carried out.

a. Construction of village master file

This is a list of sub districts and villages by their geographical location, the area codes, number of EAs, and estimated number of households. The list is constructed separately for urban and rural areas.

b. Construction of master sampling frame

The EAs seem to be an ideal sampling unit in the master sampling frame because of reasons as follows:

- a. The size is optimal, not too large and not too small. Complete census field workers should be able to finish enumeration in one EA in the allotted time. In general, each EA includes 200 to 300 buildings/households.
- b. The EAs are delineated by permanent/natural boundaries.
- c. This allows for the formation of segments or clusters of households for the purpose of sample selection in future household surveys.
- d. Or also facilitate revisits by another field workers, even if sample households are drawn systematically throughout the selected EA.

Since monitoring and updating the EAs is costly and time consuming, the master sampling frame will only cover 20 percent of all EAs in Indonesia. Updating will be carried out every 3 years, as part of the decennial census taking activities. This means the population census in years ending with 0, the agricultural census in years ending with 3, and the economic census in years ending with 6. The total number of EAs in the 1990 Population Census is expected to run between 150,000 to 200,000, of which 30,000 to 40,000

are included in the master sampling frame.

The above list will be used to draw samples of EAs for the sample census and household surveys conducted in the future. The master sampling frame will include all EAs in province capitals and municipalities, and 20 percent of EAs outside the capitals and municipalities.

To fulfill the above objectives, three types of EA will be identified.

1]. Ordinary EA

An EA is an area which have a clear boundary, and comprises of 200 to 300 households, or 200 to 300 non-residential buildings, or a combination of households and non-residential buildings of the same number. Unlike in 1980, the boundary criterion have to be met although in so doing the number of households/non-residence buildings may become less than 200 or more than 300.

2]. Special EAs

A special EA is an area which has a capacity to accomodate 150 persons but is closed to the public, such as a military base, military housing compound which needs an entry permit, seminary, hospital complex and prison. Sample selection for special EAs will be done separately, and if the number is small they will be selected in one stage sampling.

3]. Preparatory EA

This is a vacant area which is planned to be developed as a residential or business area. Preparatory EAs can generally be found in urban or suburban areas.

The following items will be collected in the sample census.

a. Housing

- Type of building
- Number of households in the dwelling unit
- Ownership status
- Construction materials of roof, outer wall and floor
- Floor area
- Fuel for lighting and cooking
- Source of drinking water and water for bathing/washing
- Toilet facility
- Bathing facility
- Ownership of durable goods

b. Population

For all persons

- Age and month/year of birth
- Sex
- Relationship to household head
- Marital status
- Religion
- Citizenship
- Place of birth
- Place of previous residence
- Duration of residence

For persons 5 years of age and older

- School attendance
- Literacy
- Educational attainment
- Highest grade attended
- Type of secondary school (general or vocational)
- Place of residence 5 years ago
- Language used at home
- Ability to speak the Indonesian language

For persons 10 years of age and over

- Economic activity
- Industry, occupation and employment status in primary and secondary economic activity
- Number of days and hours worked in primary and secondary activity
- Industry in employment one year prior to the census

For ever-married women

- Age, month and year of first marriage
- Number of marriages
- Number of children ever born, surviving and dead
- Month and year of birth of last birth and their survival status.

8. Census organization

A census working group has been established in the Central Bureau of Statistics since 1987. The group consists of several teams, each responsible for a specific aspect in the census operation. A team in this group is responsible for census mapping. In cooperation with other teams, it carried out several pilot tests before finalizing the mapping instruments. The most recent activity in this area is the training of census mapping personnel between 24 August and 9 September 1988 throughout the country. Mapping will be conducted during October 1988 for areas in the province capitals and municipalities, and in October 1989 for 20 percent of the areas in the regencies, and for potential areas.

An interdepartemental team is established in the central and regional statistics offices. The team's membership includes representatives from government agencies responsible for developing population policies and programs, and those associated with the census operation. Experts and scholars in such fields as population, health, family

planning and manpower are also invited as team members. The team is expected to provide inputs in improving the quality of the census, particularly with regard to data needed by the agencies represented in formulating future plans. Since topics covered in the census are basically fixed, the team members are asked to suggest ways to make the best use of the available data.

9. Data processing

A new computer was installed in the central office to process the census data. It is an ACOS 1500 system NEC with a capacity of 64 mega Bytes. At the same time, an ICL 3920 is installed in Surabaya, capital of the East Java province, to replace the ICL machine installed earlier. Being the largest province in the country in terms of population and number of regency/municipality, the East Java Statistics office has functioned as a data processing center for some national surveys carried out by CBS. In 5 provinces (North Sumatra, Jakarta, West Java, Central Java and South Sulawesi) NEC computers of ASTRA 370 system are acquired, while the remaining provincial statistics offices are equipped with mini computers.

10. PROPOSED MEASURES FOR IMPROVING COVERAGE
OF THE 1990 POPULATION AND
HOUSING CENSUS OF MALAYSIA

**Proposed Measures for Improving Coverage of the 1990 Population
and Housing Census of Malaysia**

by

Khoo Teik Huat
Kwok Kwan Kit
Ng Man San

Paper presented at the Twelfth Population Conference, Beijing,
China, September 5-9, 1988.

August 1988.

Introduction

Since one of the main aims of a population census is to obtain an accurate count of the population, net undercoverage of the population normally receives much attention from census takers. Coverage errors mainly arise because persons are missed in the census either because the living quarters they stay in or their households have been missed in the census and/or they are left out in enumerated living quarters or households.

While it is possible that coverage error rates may be quite constant across subgroups of the population so that distributions of their characteristics are not greatly distorted, in practice, these rates often differ substantially and not only are population counts affected, distributions of population characteristics are also affected.

Therefore, it is important that measures be taken to improve census coverage. However, census coverage errors, like other types of biases, are difficult to detect and generally require external checks to detect them. The four methods that are commonly used are (a) the Post Enumeration Survey (PES); (b) the record check; (c) analytical studies; and (d) comparisons with aggregates from administrative records.

For the evaluation of the 1980 Census in Malaysia, basically methods (a) and (c) have been used. A PES conducted to evaluate the quality of the data collected in terms of coverage and content revealed that there was an underenumeration of some 609 thousand persons giving an undercoverage rate of 4.4 per

cent. The undercoverage rate was found to be 4.2 per cent for Peninsular Malaysia and 5.5 per cent for Sabah and Sarawak; the rate for Sabah was actually assumed to be similar to that of Sarawak. While a PES was carried out in Sabah, technical and operational difficulties made it impossible for an independent estimate for Sabah to be made. In terms of the three main ethnic groups in Peninsular Malaysia, the 1980 PES showed that undercoverage was lowest for the Malays (2.9 per cent) while they were higher for the Chinese and Indians (5.5 and 6.7 per cent respectively). It was also shown that males had a higher undercoverage rate than females and high undercoverage rates were found in the age groups, 0-4 years, 15-29 years and 60 years and above. The PES report also disclosed specific reasons for this underenumeration.

In addition to the PES, the census counts for the age group 0-4 years were also checked against population estimates for this group derived from birth and death statistics by forward surviving children born (that is births less deaths) in the five years before the census. Assuming that the effect of net migration was negligible for this age group, the study shows that the PES adjusted figures for Peninsular Malaysia were lower by 2 per cent.

This information derived from the PES together with other census information can be instructive in developing measures for better coverage in the 1990 Population Census. The emphasis in the current planning of the 1990 Census is on the prevention and control of the coverage errors that are likely to be made in

implementing the census plans. This involves firstly the identification of the possible sources of such errors and secondly to take the necessary steps in preventing them or minimizing their effects. Since some of these measures are expensive, cost considerations must necessarily be taken into account. Some of the measures that have been identified and are being tested to evaluate their feasibility and operational effectiveness are discussed in this paper.

Proposed Measures

First, we plan to reduce the number of topics and hence the workload of the enumerator at the enumeration stage with a view to improving coverage. In the light of this approach the use of sampling during the enumeration stage will be introduced. It is envisaged that a short questionnaire containing a few basic questions to be canvassed for all persons would be used during the first stage which includes the listing of living quarters as well. The need to collect other detailed information would be met by using a long questionnaire to be canvassed on a sample of the population as part of the second stage of the enumeration. It is expected that the approach of making use of both a short and long form would not only improve census coverage but also would result in a reduction in the cost of holding the next census.

The Malaysian experience indicates that coverage errors appear to be positively related to the number of topics covered and hence the workload of the enumerator. Table I shows the number of persons enumerated in the first and second stage of the 1970 and 1980 Census and the final results after adjustment for

underenumeration. It will be noted that in both the 1970 and 1980 Census, the first stage of enumeration, carried out in a period of seven to ten days, involved a much lighter workload since only listing of living quarters and a short housing census schedule was used. However, the second stage of the census, involved a much longer schedule and due to the heavier workload was conducted over a period of 2 to 3 weeks. Table 1 shows that the first stage figures of both the 1970 and 1980 Census were consistently higher than those of the second stage and these are closer to the PES figures. However, it should be pointed out that there is also a difference in the period of recall for the first and second stage of the two censuses. This may have also contributed to the difference in coverage.

Table 1: Population counts from first and second stage of the 1970 and 1980 Census and Post Enumeration Survey adjusted estimates

	1970			1980		
	First stage	Second stage	PES adjusted	First stage	Second stage	PES adjusted
Peninsular Malaysia	8928884	8809557	9181674	11222773	10944844	11426613
Sabah	N.A.	653604	N.A.	1007436	955712	1011046
Sarawak	N.A.	976269	N.A.	1294537	1235553	1307582
Malaysia	N.A.	10439430	N.A.	13524746	13136109	13745241

We also plan to continue the use of the de facto approach for the official census count, but are experimenting with a moving reference time point using the night before the census interview as the reference point instead of a fixed census night as has been used in the 1970 and 1980 Census. The use of a fixed census night in the 1980 Census resulted in a time lag of two to three

weeks before some households were enumerated. This has eventually led to some net undercounts because some respondents were uncertain which of their household members should be included in the census. For example, the 1980 PES showed that some 29 per cent of the omissions in the 1980 Census was subsequently confirmed to have stayed in the enumerated living quarters on census night and therefore should have been included.

It may also be recalled that part of the reason for the higher first stage figures of the 1970 and 1980 Census is that these counts are based on persons staying in the living quarters the night prior to the census interview in contrast to the second stage figures which are all based on a fixed census night.

The results of the 1980 PES also indicated that the under enumeration of persons was greatest for the very young (0-4 years), the young adults (15-29 years) and the elderly (60 and over years). Recognising the selective nature of this undercoverage, an initial census pre-test using a questionnaire with special questions focussing on these groups was tested but the results proved inconclusive. Currently, efforts are being made to redesign this questionnaire for a further test later in the year.

The extent of control of census coverage and content errors is very much dependent on the extent of supervision that can be imposed during the field enumeration stage. Very little can be done to rectify errors made in this stage once the census forms are received at headquarters. In the 1980 Census a supervisor-

enumerator ratio of 1 to 10 was used and it is now thought that this ratio should be reduced to 1 to 5 for the 1990 Census. This improved supervisor-enumerator ratio would probably allow the supervisors to carry out specific built-in checks on coverage more satisfactorily. This approach would inevitably increase census costs but the increase is not expected to be substantial since the number of supervisors required will not be large.

For the 1990 Census it is envisaged that the basic organisation used during the 1980 Census would be maintained. The Commissioner of Census (who will be concurrently the Chief Statistician) will be assisted by two teams of census officers. The first team at headquarters in Kuala Lumpur will be responsible for the administration, planning, coordination of census activities. The second team comprising of government officers from each of the 13 states and the Federal Territory in Malaysia will be responsible for the task of census enumeration at the ground level. The State Deputy Commissioners who will be assisted at the administrative district level by the respective Assistant Commissioners and District Superintendents will be solely responsible for the census operations in their state.

The regional offices of the Department of Statistics played only a limited role in the 1980 Census, but it is envisaged that their role will be enhanced so that they can provide the necessary technical support to the census teams in the states as well as to implement quality control checks on their work. It is hoped that if substantial errors are found at this early stage, the affected questionnaires can be returned to the field for

correction.

The role of census mapping in achieving a complete and accurate census count cannot be over emphasized. Suffice to say inaccurate or incorrect census maps will lead to omissions and duplications which will result in inaccurate census data. The preparation of the 1990 Census maps is currently in progress and is scheduled for completion by the end of 1988 when they will be reviewed by the state governments. The updating of the 1980 Census maps is based on maps and information from the listing stage of the current Labour Force Surveys as well as information from land development authorities and also housing developers site plans. For maps which cannot be updated in the office, a listing operation will be undertaken to update these maps. It is hoped that with the vigorous effort made in preparing more accurate maps for the 1990 Census, the quality of the census counts would be improved correspondingly.

The role of a national publicity census programme in ensuring a more complete coverage is well recognised. A good and well organised publicity programme can go a long way in making the people aware of the forthcoming census as well as to enlist their co-operation in providing complete and accurate information. For the 1990 Census, a full-scale publicity programme will be launched. The help of various organisations, associations and guilds in publicising the census enumeration would also be pursued more vigorously. A publicity programme incorporating a "missed persons" campaign whereby persons who have been omitted in the census enumeration are urged to notify

the local census office will be used as well.

Summary

In this paper we have reviewed some of the probable causes of coverage errors of the 1980 Census. These have been detected by means of analytical studies based on vital statistics and the 1980 PES. Over the past 5 years we have tried to use this information to develop several measures, which are currently undergoing field tests, that will be used in the 1990 Census to control and to reduce the extent of coverage errors. These are:

- (a) a reduction in the workload of the census enumerators;
- (b) the use of a moving reference point instead of a fixed census night;
- (c) the use of check questions on those subgroups with high underenumeration rates;
- (d) a lower supervisor-enumerator ratio;
- (e) a greater participation of the Department of Statistics regional offices;
- (f) an improvement of the census maps; and
- (g) a more intensive national census publicity programme.

11. THE USE OF COMPUTERISED TECHNOLOGY IN THE
1991 NEW ZEALAND CENSUS OF
POPULATION AND DWELLINGS

EAST-WEST POPULATION INSTITUTE
and
STATE STATISTICAL BUREAU OF CHINA

TWELFTH POPULATION CENSUS CONFERENCE
BEIJING, CHINA

September 5-9, 1988

The use of computerised technology in
the 1991 New Zealand Census
of Population and Dwellings

Michael Moore,
Director,
Population Census Division.

Department of Statistics,
NEW ZEALAND.
August 1988

The Use of Computerised Technology in the 1991 New Zealand
Census of Population and Dwellings

1. Introduction

Advances in technology have played an important role in the methods of processing recent New Zealand Censuses of Population and Dwellings. These changes have resulted in greater efficiency in operational procedures and improvements in data quality.

The 1986 Census saw a move away from the use of manual coding, and data capture using Optical Character Recognition, to 'key-to-disc' entry directly from the questionnaire, and the use of Computer-Assisted Coding (CAC) for the traditional 'hard to code' fields.

Whilst considerable savings in resources were achieved, the processing system however operated as a series of separate processes. It is our intention for the 1991 Census to integrate and automate these processes where beneficial.

2. System Overview

Following the receipt of completed questionnaires from the field, the control system will issue batches of work to the pre-capture processing operation where questionnaires are checked and prepared for the data capture operation. A key-to-disc data capture operation utilising PCs will follow, the quality of which will be controlled by range checking, field and inter-record edits and batch controls. The records will then be formatted for Computer-Assisted Coding.

The data will be transferred to the mainframe computer via floppy disk or a network link for preliminary processing where the imputation of sex, age and work status will be undertaken, together with the allocation of a family code.

The processing of the 'hard to code' fields will follow, utilising CAC software, and will include further inter-field and inter-record editing.

On completion of CAC and editing the data file will automatically pass to the final process involving automatic editing, the conversion of non-standard codes to standard codes, the derivation of new data fields, the creation of a family record and the formatting of each record for processing by the TPL based output system.

Throughout the processing operation a flow control system will ensure that each batch of questionnaires passes smoothly through each phase. It will enable pinpoint monitoring, access and control

over each batch at any stage during the operation and provide management with up-to-date information on progress and performance.

3. Census Questionnaires

The capture of responses from the census questionnaires is a two stage operation. Initially, tick-box and numeric responses will be captured directly from the questionnaire in the data capture operation, and subsequently the 'hard to code' fields will be captured at the Computer-Assisted Coding phase.

This requires the design of questions which are suitable both for data capture and respondent coding. Separate dwelling and personal questionnaires will be designed, as was the case at the 1986 Census, with the emphasis on designing questions which permit the use of ticking the appropriate answer box.

4. The Processing System

(a) Pre-Capture Processing

This phase is carried out in two parts, the first of which concerns the checking and balancing of questionnaires with the field collection records, while the second involves checking and editing the question responses prior to data capture.

Part 1 comprises separate clerical and computer activities. The clerical process includes checking that each occupied dwelling has a dwelling questionnaire and the appropriate number of completed personal questionnaires (late questionnaires and dummy questionnaires may be inserted, duplicate questionnaires removed and dummy questionnaires replaced). The computer process involves the entering of page totals (of population and dwellings counts) from the fieldbook via a VDU and balancing these counts with the number of questionnaires.

Part 2 involves the sorting and numbering of questionnaires, plus the coding of several questions including dwelling type, relationship, parent number, and absentees. Finally, the questionnaires are groomed by checking for legibility and ambiguity, and eliminating obvious errors and inconsistencies from the question responses. This grooming process is undertaken to ensure the data capture operation is fast, accurate and effective.

The major changes since the 1986 Census to the pre-capture processing operation include the computerisation of the questionnaire and fieldbook balancing operation, and the elimination of family coding which is replaced by an

identifying parent (record) number, and which will be used in a subsequent phase to derive family codes automatically.

(b) Data Capture

The capture of respondent-coded answers will be undertaken on Personal Computers utilising software such as Entrypoint 90, Rhode PC or Centry.

A pool of 50 PC's will capture the data over a 6-7 month period, after which the equipment will be distributed throughout the department.

Several evaluations of software have been undertaken, with no single product being clearly superior. All meet the basic requirements for this process.

The software will be programmed to perform range checks for each field, plus a series of validation checks (e.g. that a dwelling record is followed by a personal record).

After performing these checks, the following process will be undertaken

- each record will have a 'record type' determined
- additional fields for CAC processing will be created
- a holding or not-specified code will be allocated to CAC fields

Given the software capabilities, suitable verification strategies will be adopted to ensure a high level of accuracy is achieved.

The major changes in the data capture operation since the 1986 Census include the replacement of the Inforex equipment with Personal Computers and the use of software which permits a wide range of processes to be undertaken. Inforex had very limited capabilities in terms of range checks and verification options.

(c) Preliminary Processing

On the completion of data capture, the records will be transferred via floppy disk or a network link to the mainframe, where the following tasks are performed:

- the number of records (i.e. questionnaires punched) are compared with the Counts file (created at pre-capture processing)
- missing data for the fields sex, age and work status are automatically imputed
- family code is derived

At this point the file is copied and a limited number of inter-field edits performed. The data is then formatted for

TPL processing and a range of provisional statistics are produced culminating in a national summary being processed within 6-7 months of census date.

The original data file now passes to the Computer-Assisted Coding phase.

The main changes to this process at 1991 Census, will be the dropping of several processes which can now be performed at data capture. Also the automatic derivation of family code has replaced the former manual coding.

(d) Computer-Assisted Coding and Editing

Fields that are not captured at the data capture stage receive a holding value from the computer. This includes written responses to questions on Occupation, Type of Work, Residence 5 years ago, as well as written responses to "other" categories in certain questions. These fields will be displayed on screen requiring the operator to type in a description.

Each word entered is used as a key to access a dictionary file and any exact matches are stored. Where an exact match is not possible, the word is truncated one letter at a time from the right until a match is obtained. After all entered words have been processed, the matched entries are scanned to find the most frequently occurring references.

The descriptions and codes containing the largest number of matched words are then displayed on the screen from which the operator can select the most appropriate description. The codes of selected descriptions are automatically transferred to the data record.

If suitable descriptions are not located, the response can be re-worded and entered again or by-passed for subsequent expert attention.

The success of Computer-Assisted Coding is dependent on the quality of the dictionary files of descriptions. Considerable effort will be made over the coming years to upgrade these files especially Occupation, Religion, and Tertiary Qualifications. The availability of Standard Classifications such as Country Code ensure the adequacy of code lists for such topics. However, fields such as Tertiary Qualifications and Religion will require significant upgrading.

Information on streets is obtained from the NZ Geostatistical System which is updated on an on-going basis, while the Business Directory supplies Employers Names and Addresses permitting the coding of Type of Work and Workplace Address.

Whereas at 1986 Census operators were responsible for the coding and editing of a complete batch, the approach for 1991 Census will be to divide CAC into 3 separate operations, including:

- Labour force coding
- Non-labour force coding
- Editing and coding of by-passed fields

This approach will enable greater specialisation in aspects of the labour force (e.g. Occupation coding) or other fields (such as Location coding). The creation of a smaller edit team will also permit greater control over and observation of data discrepancies, and enable Managers to respond with appropriate coding policies.

A number of additional features will be built into the 1991 CAC System. These include a Direct File Matching (DFM) facility, where the computer finds an exact match for a respondent's description. The code will automatically be written away to the record without further manual intervention.

Variations of Direct File Matching are currently being undertaken and research to date indicates that the following fields will be suitable for DFM: Relationship, Country of Birth, Ethnic Origin and Religion.

A further enhancement is the use of a recall facility, whereby the operator is able to recall the codes/descriptions from previous records. Variables which are common within a household such as usual residence, address 5 years ago, ethnic origin, religion and country of birth can be recalled and the codes allocated to the current record.

The availability of an on-line news facility which will contain details of coding policy (in particular, difficulties or problems which occur during the processing operation) will be of assistance to operators and help improve data consistency.

Whereas inter-record editing was carried out in batch mode at 1986 Census, this process will follow the inter-field editing during the CAC phase. The introduction of automatic family coding will however, significantly reduce the number of inter-record edits required.

The CAC/Editing operation is planned to be undertaken on a 1 shift basis, using a staff of 72 persons for 1 year. Unlike the 1986 Census, shift work will not be implemented and each operator will have his/her own VDU.

(e) Automatic Editing

When the operator clears the batch of data after the CAC/Edit phase, it passes directly to the Automatic Editing phase, where a number of inter-field edits will be performed. (These were formerly part of the inter-active edit system at CAC phase in 1986.)

The program will check each record for the pre-determined edit conditions, and records containing such inconsistencies will have one or more fields altered automatically. A preliminary assessment of the edits suitable for automation has been completed. Generally, they involve format-type inconsistencies or inconsistencies where correction of only one field need be made on most occasions.

(f) Rationalisation

This final process leads to the creation of the Census masterfile. Data will pass directly from the Automatic Editing process to the Rationalisation process which involves the following:

- the input data is combined with 'areas' data to form an hierarchically structured output file comprising a Geographic, Dwelling, Family and Personal record suitable for TPL processing
- non-standard codes are converted to standard codes
- new data fields are derived (such as Total Household Income, and Household Type).

5. Quality Control

A quality control scheme for CAC processing will be developed along similar lines to that adopted at 1986 Census. This will provide information on the standard of each operator's performance, and consistency rates for each subject-matter field, and will highlight any need for retraining of operators.

The system is largely dependent on the CAC system as the initial data is obtained from duplicate coding of a batch of work by two CAC operators. The two files are matched against each other and a listing of all inconsistencies produced. This list will be checked out by a member of the Quality Control team and errors recorded by subject-matter field. The batch of work is then returned to the appropriate operator for correction.

It is our intention to check approximately 10% of all work (the equivalent figure at 1986 Census - 5% - was considered to be too low), with several levels of checking, this being dependent on the error rates achieved by each operator.

6. Flow Control/Data Management System

A flow control/data management system is currently being developed which will provide an effective means of control on the progress of questionnaires and data through the various phases of the processing operation. The system will provide for -

- the issue and receipt of questionnaires to each processing phase
- monitoring and updating progress on each phase of the operation
- maintaining a history for each batch in the system
- maintaining counts of population and dwelling records.

The system available at 1986 Census was a series of independent systems which had to be independently updated using manual methods, and compared to ensure they were consistent.

The system being developed for 1991 Census will comprise one integrated system with on-line updating facilities. The system will provide an immediate response to queries enabling management to have more control over the flow of questionnaires.

It is planned to use bar code labels on questionnaire wallets and readers to scan these codes, automatically entering batch numbers into the control system.

An enquiry screen will provide information on progress or determine the whereabouts of any batch of work.

The Counts data is also being incorporated into the Control System. The counts are created and input into the system at phase 1 of the pre-capture operation. If necessary these counts may be updated at a later date.

The Control System also includes:

- a history file - each time the control file is updated, the data for a batch is recorded on a history file.
- a geography file - containing standard area codes for a variety of geographic regions (administrative/non-administrative) enabling management reports to be produced for these various regions.

7. Output System

(a) Output Strategy

The strategy which will be adopted for 1991 Census requires provision for the production of a broad range of summary statistics which meet the department's "community good" obligations whilst developing a range of "commercial products and services" in order to conform to current government policy that requires the department to generate specific levels of revenue each year.

(b) Community Good Statistics

This range of statistics includes information releases and publications and their availability will commence within 4-6 weeks of Census date. During the first 6-8 months, only provisional statistics will be available including:

- Counts - returns of population and dwelling counts from District Field Supervisors are processed and published
- Key Statistics - as data becomes available a summary range of statistics are generated for each region and then a national summary is produced.

Final statistics will be available as soon as sufficient batches comprising a region pass through the rationalisation process. These are published in the form of regional reports and culminate with a national summary. A limited range of subject-matter reports are then produced.

(c) Commercial Products and Services

In addition to the above statistics a range of products and services will be developed in order to meet market needs and ensure that revenue requirements are met.

These products and services include a range of unpublished cross-classified tables, data from a small-area database, time series statistics on INFOS, the provision of output on magnetic tape or floppy, Supermap and the development of customised tabulations to meet client need.

(d) Output System

As each batch of data passes through the rationalisation process, an output file containing derived fields suitable for TPL processing is created. These batches aggregate to the 13 regions in New Zealand and on completion of the

processing operation will comprise the Census Unit Record Masterfile.

From this masterfile, a number of sub-files are created for the processing of the various census reports and will be of use in the official servicing of clients' data requests. These sub-files will be available for Dwellings, Families, Adults and Electoral Areas, while the range of sample files is still being assessed.

Data from the Masterfile will also be downloaded (as each region is available) to the Small Area Database from which the requirements for SUPERMAP will be generated. SUPERMAP is a mapping package developed by Space Time Research Ltd, Australia. Data from the census and digitised boundaries are stored on a compact disc, and tables and maps can be generated using a PC.

The overall design and maintenance of the output system will continue to be the responsibility of the EDP Division, while the specification, program development, testing, production, validation and the preparation/editing of tables for publication/clients use will again be the responsibility of the survey section (i.e. Population Census Division).

(e) Tabulation Software

TPL (Version 5) was the major table-generating software used at 1986 Census and, whilst it has a number of limitations, it is still considered to be the best tabulation software available at present.

The principal problem with TPL is its inability to generate table formats which meet departmental standards, resulting in considerable screen editing to bring the output up to publication quality.

Currently, it is proposed to install Version 6 of TPL, enabling the processing of part files, the cross classification of observation variables, and with additional PCL features, the enhancement of the quality of the output produced.

The opportunities provided by desktop publishing software such as VENTURA, are currently being assessed. This software is extremely powerful enabling the production of high quality products.

(f) Publication Timetable

The following is the proposed publication timetable for the 1991 Census of Population and Dwellings (being held on 5 March).

<u>Release</u>	<u>Time-frame</u>
. Provisional Counts	April-June 1991
. Provisional Key Statistics	July-December 1991
. Final Counts	November 1991
. Regional Reports	December 1991 - May 1992
. Subject-Matter Reports	June 1992 - December 1992

NOTES:

1. The number of records to be processed at the 1991 Census of Population and Dwellings is expected to be:

1,176,000 Dwelling records
3,449,000 Personal records

2. The computer hardware available to process this Census is as follows:

CPU:	IBM 4381/P13
MIPS:	3.7
Main Store:	16 mb
Disk Store:	9 x 3370, 712.8 mb
Tape Units:	2 x 3420
Printer:	1 x 4245 producing 2,000 lines per minute
CAC VDU's:	72 x IBM 3178s or equivalent
Operating System:	IBM VM/SP Release 5 CMS

12. SOME POLICY CONSIDERATIONS FOR PLANNING
THE 1991 CENSUS OF HOUSING AND
POPULATION OF PAKISTAN

EAST - WEST POPULATION INSTITUTE
TWELFTH POPULATION CENSUS CONFERENCE
September 5-9, 1988 Beijing, China
CO-SPONSORED & HOSTED BY STATE STATISTICAL BUREAU OF CHINA

SOME POLICY CONSIDERATIONS FOR PLANNING THE 1991 CENSUS
OF HOUSING AND POPULATION

by

Muhammad Aslam Chaudhry
Joint Census Commissioner
Population Census Organization
Statistics Division
THE GOVERNMENT OF PAKISTAN

C O N T E N T S

	Page No.
Some Policy Considerations for Planning of 1991	
Census of Housing & Population of Pakistan	287
1. Topics of the information	288
1.1 User's needs	288
1.2 Comparability	289
1.3 Quality of previous data	289
1.4 Type of enumerators	290
1.5 Combined or separate census	292
2. Timing	294
3. Coverage	296
4. Use of sampling in census	297
5. Tabulation	299
6. Census geography	301
7. Hiring of field staff	302
8. Training of field staff	304
9. Publicity	304
10. Census publications	306
11. Budgetary Requirement	307
Annexure - 1	309

SOME POLICY CONSIDERATIONS FOR PLANNING OF 1991
CENSUS OF HOUSING & POPULATION OF PAKISTAN

Census as collection of information about all persons living in the country and housing units occupied by them and their characteristics at one point of time, their compilation and dissemination is a gigantic task. Its planning generally starts three to four years before the actual canvassing. This planning includes framing/revision of census legislation, consultation with users, formation of advisory committees, preparation of questionnaires and manuals/training guides and development of field operation plan, tabulation plan, data processing procedure, publication plan, publicity, post evaluation survey procedure and proposal for analysis and research on the collected data.

2. The adequacy of the planned procedure is generally tested through a pretest survey. Execution of field operation is itself a huge task which involves training of trainers and subsequent training of field staff, supply of documents to the field staff and their retrieval, complete geographic coverage, proper supervision, quality control measures to be adopted during the operation, data processing and finally dissemination of field obtained information in the form of official documents. While publicity before and during the actual field operation could play an effective role in minimising various type of errors provided it is well planned. The post evaluation survey provides information about the degree of reliance upon both coverage and contents of the census returns. Analysis and research on the data facilitate in identifying shortfalls left in the questionnaires, training procedure, field operation, supervision, tabulation and data processing which can serve as feedback for improvement in quality of the next census.

3. Therefore, policy planning needs some careful considerations based on the past experience of the census personnel who are involved at various stages of census. Available literature and experience of the other countries can also be benefited. A modest attempt has been made in this paper to highlight some important policy considerations for planning the 1991 Census of housing and population of Pakistan.

1. TOPICS OF THE INFORMATION

4. The choice of topics mainly depends upon the user's needs. However, comparability with other sources at national and international levels, ESCAP recommendations on the subject, quality of previously collected data on the relevant topics, field problems, decision regarding combined or separate phases of census, availability of enumerators and their calibre, the topics of information, decision about the time to release the data and the budget granted for the census operation also play vital role in the selection of topics of information of census.

1.1 User's needs

5. During the last one decade or so awareness among the data users have tremendously increased. Besides the main users like public servants involve in planning and evaluation of different development projects in different sectors of economy, political conscious has increased to use census data for obtaining rights. Even the industrialists are now establishing their industries after knowing the availability and nature of labour force in the area of their interest. Census data is also being used for the provision of basic necessities, recruitment of public servants, control of law and order, budgetary purposes and for other miscellaneous purposes. Several pressive demands were also poured in from town planners for supply of population and housing data. The increasing awareness of these users have to be kept in mind while taking decision about the topics of the census information.

1.2 Comparability

6. Without comparing with other parallel sources at national and sub-national level it is difficult to assign any degree of reliance in the census returns. The comparison of previous censuses establishes trends of various characteristics of the population and housing units and their likely trends in future. While comparison at international level facilitates in knowing the country developmental level compared to other countries of the world. While deciding the topics of information, therefore, comparability aspects can not be ignored. UN ESCAP Secretariat have recommended some topics of information for census. Some of these are core questions and must be included in census enquiry while other are useful topics. These recommendations are based on the experience of census experts of different countries. Due consideration will also be given to these recommendations for the selection of topics of information.

1.3 Quality of previous data

7. Quality of data could be a useful guide for the selection of topics, wording of the questions and probing for seeking reliable information. Methodology research on 1981 Census have identified some weakness which could be easily ascribed to the sensitivity of the topics, wording of the question and/or probing problems. Two type of errors are found in the census returns, viz response errors and processing errors. The response errors were the outcome of weaknesses in planning of census while process errors were the result of editing, coding and keying of the filled-in entries on the census questionnaires.

8. The rate of response errors and processing errors were more in urban areas where multi-nature characteristics of population were found compared to rural areas where characteristics were less variant. Response errors were much high in literacy, school/college attendance and disability while processing errors were more common

for place of previous residence of the migrants and disability characteristics of the individuals. Some districts have shown high rate of response errors for some characteristics. The above analysis guided us to give more emphasis either on probing of such questions or urged to drop sensitive topics from the census coverage and may be asked through special surveys. Besides, analysis and research on the data have helped in identifying the shortfalls in the questionnaires. For example, through analysis it was found that information on education which filtered through question on literacy had counted only formal literate for educational characteristics. Thus informal literates were not included. This resulted under coverage of educated persons owing to the defect in the questionnaires.

9. The analysis of data also guided us that students generally state the level of education in which they are currently enrolled. This tendency inflated the level of education. Such type of problems require special attention while deciding upon the topics to be investigated through census.

1.4 Type of enumerators

10. The type of enumerators and informants of the households play a significant role in bringing out the quality of work. In the 1981 Census low paid officials of the Revenue Department, Primary school teachers (male) were engaged as enumerators. Through a test intelligent enumerators were filtered who were given special training for sample count. This sample count comprised topics either sensitive or complicated in nature with respect to their conceptual problems which demanded tactful probing. However, experience shows that quality of the sample count was no better than the complete count. That provokes the calibre of enumerators to be engaged for sample count in the forthcoming census provided such topics has to be canvassed through sample enquiry. Informants of the household were mostly the head of household and in his absence the responsible

person. In most of the cases these informants were males. It is believed that in Pakistani society males generally do not remember well the particulars of the children as compared to their female members. They also bother little in seeking information from their concerned family members before answering to the enumerator's queries. Consequently some inadvertant errors are crept into the census filled-in questionnaires. Thus, while formulating census planning the type of enumerators to be engaged and their interacting effect with the informants may also be given due consideration.

11. The planning for various sectors of economy preferably depends upon availability of the latest data. This increasing awareness is putting more and more pressure on the data suppliers for release of data as early as possible after the field operation is over. Too much topics of information not only hamper the quality of census returns but also require a lot of time for coding of open ended questions, editing of inconsistent entires, keying time, clearance of tabulation and publication of data. In the 1981 Census question on age, sex, marital status, relationship to the head of household, type of population, i.e., dejure and defacto, literacy, ability to read the Holy Quran, religion, enrolment and education were asked through the complete count. While other recommended questions such as level and field of education, migrant population, economically active population, children ever born and still surviving, disability and persons gone abroad during the last 10 years were asked through the sample count. The response on fertility question was very poor, firstly because as stated above informants and enumerators were mostly males. Many enumerators hesitated to probe this sensitive question and respondents did not bother in seeking information from their concerned female members before replying to the enumerators. Enumerators also bothered little to have probing about the nature of disability because of it touchy nature. He also failed to record properly the description of the occupations and industries of the working labour force, especially of females. Filteration of the information through male informants introduced unconscious bias which gave inexplicable results.

12. Engagement of female enumerators can have access to the housing units and can easily probe the sensitive questions particularly pertaining to female reproductive behaviour from the female informants. Availability of female enumerators perhaps impose no problem in towns and cities where primary school teachers can be engaged. However, in rural areas availability of females is doubtful where there are few schools in number which are serving scattered population spread over a large area and hardly there is one teacher per school. Moreover, engagement of such teachers would be a risky job as it is difficult for them to visit scattered population located in open fields away from the main settlements.

1.5 Combined or separate census

13. The 1972 Census was conducted in two phases, viz Population Census Complete Count and Sample Count. Initially it was planned to conduct a census in 1971 but due to war with India followed by heavy flood in most part of the country, census was delayed by one year. The complete count was canvassed in September, 1972 followed by sample count called Housing, Economic and Demographic Survey (HED-1973) a year later. The 1981 Census was also conducted in two phases but in the first phase housing information proceeded by house listing and the second phase complete count with simultaneous canvassing of sample count information was carried out. Although initially it was planned to conduct Population Census preceded by Housing Census with one year interval. The aim of Housing Census one year earlier was to obtain sample frame subsequently to be used for selection of sample for the sample count but the Housing Census was delayed by nine months. It was conducted three months earlier the Population Census. The same field staff was engaged in both the operations. The experience shows that the field staff lost enthusiasm and zeal during the intervening period of two phases. This might have affected the quality of later phase of census. Also the engagement of field staff twice was very expensive.

14. A census can be conducted on hundred per cent basis or it can be conducted with some core questions in complete count and other recommended questions on sample count basis. Housing Census can be combined with either of these two counts. Both the combined and separate censuses have their own merits and demerits. Some of the countries of the world had combined operation while most had separate operation for their 1980 round of censuses. Those who had combined operation generally had very high literacy level. They either delivered their questionnaires from door to door and then collected after a reasonable period or mailed them. These countries generally use Optical Mark Reader device for data entry. Some of them used standardised abbreviations for various occupational and industrial classifications for shortening the coding time of open ended questions and also standardised the names of the migration units for computer coding.

15. This type of experience is perhaps successful where above mentioned facilities are available and literacy rate is very high. Where literacy rate is very low like Pakistan, door to door canvassing is unavoidable. From the experience of those countries which used Optical Mark Reader (OMR) or other speedy devices for data entry and computer coding can be benefited for census planning.

16. Separate census has its own merits and demerits and perhaps sampling frame is the pre-requisite for drawing representative sample on scientific basis for which house listing has to be completed well ahead of the sample count for selection of sample. In combined operation the size of the Census Enumeration Area (CEA) has either to be reduced for its coverage by an enumerator within the specified period or the enumeration period has to be extended.

17. The former proposal requires more number of blocks, enumerators and their supervisors. While the latter likely to create problems due to population change during the enumeration period

which may occur due to natural growth and migration. Memory laps may introduce errors in recording the information. All these issues have to be addressed before taking decision to conduct census either on combined or separate census.

18. The 1981 Census took 1800 man-months for coding of open ended questions and editing of inconsistent information or in term of years it consumed two calendar years. If the same type of information has to be carried out through complete count in the 1991 Census then 7 to 8 times additional time will be required for editing and coding. In view of the increasing awareness amongst data users Population Census Organization of Pakistan is thinking to finish the processing work in one year. In other words either 15 times Key Punch Operators have to be recruited or very speedy device like OMR has to be used for direct transcription of data from filled-in questionnaires to computer tapes without keying it. Plus an efficient system of either computer coding or use of self coded questionnaires has to be considered. The Census Organization is actively thinking the merits and demerits for adoption of either combined complete count operation or complete count followed by sample count operations of census.

2. TIMING

19. Census is a time oriented activity. From planning upto dissemination of data all activities are interdependent. If one activity is disturb it is likely to disturb the whole census operation. Before the actual canvassing entire planning including publication of data has to be pendown before its pretest. The pretesting has to be carried out as census itself a year before the actual operation, so that planning could be revised in the light of shortfalls detected through pretest.

20. The decision of the census date mainly depends upon the availability of enumerators, accessibility of informants, nationwide activities, cropping pattern in rural areas, seasonal movement of

the people, favourable weather conditions and ease in analysis of the census data and its comparison with past censuses. In Pakistan generally low paid teaching staff and Patwaris of revenue department are engaged as enumerators. The availability of teachers depend upon the annual examination of the students which normally held in the first or second week of March every year. In the last two or three months preceding the examination teachers generally have tuition for their additional earnings. To them this coaching is a prestigious job for their additional salary. The availability of patwaris depends upon the sowing and harvesting timings of different crops. Again there is a time lag of almost one month in these timings of same crop from northern to southern part of the country. The revenue staff has to visit the fields for assessing quality and quantity of various crops for deciding the amount of revenue to be levied upon the formers. Therefore, from their visit to the fields upto the harvesting time revenue staff normally remains busy. Apart from revenue staff the informants who are generally the formers also remain busy at the time of sowing and harvesting. Thus in rural areas the availability of revenue staff and male informants of the households to a greater extent depend upon cropping pattern.

21. The most part of NWFP, Northern Area, Azad State of Jammu and Kashmir, Tribal Area adjoining to NWFP and Northern part of Baluchistan normally remain snow-bounded from December to February. During this cold winter season people living on the top of the mountains come down alongwith their animals. Some of them have alternate residence at the base of mountains and spend their cold season there. While other move here and there where they found suitable land for grazing their animals. Normally moonsoon sets in during the months of June and July with heavy rainfall and sometime followed by flood. The months of June and July are also very hot when the temperature goes up 40 degree centigrade in Punjab, Sind and Baluchistan and south eastern part of NWFP.

22. With the exception of the 1972 Census the remaining censuses of Pakistan were conducted in first quarter of the year ending at 1, while 1972 Census was conducted in September, 1972. The most of analysis demand 5 or 10 years interval between two statistical inquiries. Thus, the cohort analysis became very difficult without doing adjustment of the 1972 Census for making 10 years interval. Also it became difficult to compare data between the two consecutive censuses when tabulation was produced in 5 year age groups.

23. Holding of any nationwide activity during or just before the census enumeration period such as national and provincial assembly elections, students examination, visit of revenue staff to fields for assessing the revenue, etc., and natural calamity like heavy flood, political disturbance, etc., do affect the quality of the census returns if not the time schedule.

24. Due weight, therefore, has to be given to these considerations while deciding about the census date and census enumeration period.

3. COVERAGE

25. The coverage can be bifercated into area coverage and population coverage. The entire area of Pakistan was covered in the 1981 Census which includes provinces of Punjab, Sind, NWFP and Baluchistan. Besides, Tribal Areas adjoining NWFP either controlled by Federal Government or Provincial Government, Azad State of Jammu & Kashmir and Northern Areas were also covered.

26. The entire population of the above mentioned areas were enumerated both on dejure and defacto count basis yet population was published on dejure basis as the country normal practice. The analysis of the 1981 Census data and field supervision have

identified some pockets which was either not covered or population was under enumerated in the covered areas. Also duplication in counting was observed in some pockets which have also been identified and listed for policy consideration for the forthcoming census.

27. Instructions were also given for enumerating special population such as living in hospitals, hotels, jails, sanatoria, military restricted areas, etc. Nomads and baggers who otherwise did not have any residence were covered under the special instructions. In some parts of Baluchistan tribes moved from place to place in different seasons. Their area of movement and tribes were identified before the actual enumeration and they were also covered under special instructions. Instructions were also given not to enumerate Afghan Refugees. Those living in camps were not enumerated and separate record was obtained from Afghan Refugee Commission. In many cases those who were living in regular settlements could not be identified and hence enumerated as Pakistani. Had the question on nationality included in the 1981 Census questionnaire these Afghan Refugees would have been identified easily. All foreign diplomats working in different embassies were not enumerated except foreigners working with other agencies.

4. USE OF SAMPLING IN CENSUS

28. Sampling method can be used for collection of information on sensitive and complex characteristics of population and housing units, for quality control and for post evaluation. Pakistan used sampling in 1972 and 1981 Censuses. In both these censuses stratified sample design was used with selection of sampling units with probability proportionate to the population size. In the 1972 Census restriction was imposed to select every fifth household within the selected block in systematic random start. While in 1981 Census the whole census block was enumerated. This disallowed the spreading of sample and hence became less-efficient compare to the

former design. For the former design sample size was around 3.2 per cent while for the later it was around 10 per cent.

29. There are two more proposals for the adoption of sample design for the 1991 Census if canvassing has to be carried out partly on sample basis. These are selection of self-weighted design with systematic selection of census block with random start and the sample size is recommended to be 5 per cent. Another proposal is the selection of housing units in systematic order with random start and sampling fraction of 20 per cent. The advantages and disadvantages of these designs are given in annexure-I. While deciding upon the choice of the sample design one more thing has to be kept in mind that is tabulation of data at administrative district level. The 1972 Census sample count had given some inadequate cell information in those tables where variables are cross classified for the districts where population was very scattered. These cell information were particularly related to age and occupational and industrial classifications of the labour force. As census planners these issues have to be given due consideration for policy decision.

30. Use of sampling for quality control measure can not be ruled out. Sampling can also be used in pretest, field operation, coding and editing processing and keying of the data. Use of sampling in pretest can permit in identifying the shortfalls in the plan and provide opportunity to rectify them before actual operation. In the census history of Pakistan every time supervision is carried out without the use of sampling. Thus, the supervisors are allowed to subjective selection of housing units for checking. Though in the 1981 Census restriction was imposed to identify some important buildings before field visit for checking purpose but much was left desirable. Such a subjective approach could not guide the decision makers to take any decision based on the magnitude of errors intolerable. The selection of the housing units on planned sample basis would help in measuring non-sampling errors generally

committed by the enumerators and respondents and shortfalls in the plan. The planned sample enquiry needs development of proformas to be filled-in by supervisors and bound them to visit the housing units. Such filled-in proformas would facilitate in identifying and measuring the magnitude of above mentioned errors by matching their information with the information collected by enumerators on the census schedule. Certainly it will be a pains-taking job and invoke the competency of the census planners but at the same time can not be ignored.

31. Checking of editing and coding job on sample basis can facilitate to adopt statistical control measures. Establishment of tolerance limits would permit the acceptance of the editing and coding work done and rechecking of discarded cases with less cost and time while ensuring quality. Similarly sample can also be introduced for quality control measures in keying of data by the individual Key Punch Operators. However, no such measure is require if OMR device has to be used for transcription of information directly from census filled-in questionnaires to the data computer input device.

5. TABULATION

32. There are two ways of developing tabulation plan. First one is the development of the plan directly from user's needs. Another method is development of the plan from the topics of information in the questionnaires. But the objectives of development of the plan through either of the above two methods is to meet user's needs on priority basis, timely release and adequacy of the data they needed, resources at hand, unambiguous and meaningful statistics for full analytical insight without omission or duplication of un-necessary information. Also tabulation plan should permit comparability with national and international data and change in definition and in area. The data should be consistent within the source of collection with regard to number and reasonability of the

information and should provide adequate information for each cell especially when cross classified information are tabulated. It should also be consistent with other parallel sources.

33. Tabulation plan for 1980-81 housing and population census was developed from information contained in the questionnaires. It included 10 tables of housing and 24 tables for population characteristics. The housing tables covered information on household size, construction period, tenure, construction material used in house building and housing facilities like drinking water, light, cooking fuel, bathroom and kitchen with cross classification of usually two to four variables. The tabulation plan of population also included two more tables at district level giving some basic population information in respect of individual urban and rural localities. The plan covered population size, its growth, distribution and composition, marital status, religion, literacy, enrolment, education, migration, labour force, fertility and disability characteristics. Language spoken in the household and persons gone abroad during the last 10 years were also included in the plan. These tables were approved by the committee of data users set-up for the purpose. However, because of poor quality of data on children ever born and still surviving, the data was not published in regular publications, though it was made available in the office for interested users. For earlier release of data we decided offset printing which limited the cross classification of information upto 132 horizontal digit place in computer tabulation. Initially it was planned to separate spanner and stub headings in title headings through the word 'by'. But in view of the large matrix in some tables exceeding 132 horizontal place slight changes were made. The data published generally met the data users requirements. However, there is an increasing demand to provide data on religion at the lowest identifiable administrative unit level. Also the tabulated data was found inadequate for formulating developmental project with respect to individual large urban localities and research work. The

table of migration published in national, provincial and district reports could not facilitate in arriving at figures of in- and out-migrants at district level which was the basic unit of migration. The tabulation plan also could not provide adequate cross classification of labour force population and homeless population.

34. In the 1981 Census tabulation some tables were deleted compared to the 1972 Census while some new tables were added and some tables were modified. The tables which were dropped from tabulation plan of the 1981 Census either had poor quality of the data, could not provide meaningful information or a question was not asked in the census. Those which were added were either based on new questions asked through the census or had useful information. However, some tables which could not provide adequate information in the earlier census were modified for producing more useful data.

6. CENSUS GEOGRAPHY

35. Census geography can be divided into two subjects, i.e., delimitation of areas and development of maps. The workload of census geography mainly depends upon the size of the population. In 1981 the entire area of Pakistan was divided into 78,279 blocks, of which 16,092 were in urban areas. Though it was initially planned to prepare blocks of 200 to 250 households on the average or 1000 to 1500 persons. However, in rural areas revenue estates smaller than this size were considered as independent blocks. In some cases blocks exceeded the predetermined number of the households. The average size of block came down to 160 households.

36. Complete description was prepared for identification of the block boundary. For rural areas, maps of revenue estates were not prepared because of their well known boundaries to the revenue staff. As the population is increasing with higher annual growth rate of 3 per cent per year it is expected to reach 125 million by the 1991 Census. Following the same scheme for delimitation, the

entire country including Azad State of Jammu & Kashmir and Northern Areas will be divided into 120,000 blocks.

37. In view of the financial constraints and paucity of other modern facilities in the past, the 1981 Census maps were developed for urban areas only. In some cases essential ground features were missing on these maps. This imposed some problems in coverage of area which resulted under coverage of population in some areas and in some cases their double counting.

38. Two types of maps were prepared for large urban areas; one showing census charges, census circles and electoral wards, other maps were prepared for individual wards showing census circles and block boundaries. The purpose of the first maps was to provide guideline to the supervisors for locating their supervisory areas. The purpose of other maps was to trace out the area of the census blocks by the enumerators, and enumeration of housing units and population characteristics of the census blocks. For each small urban areas only one map was prepared showing different census tiers. This procurement, updating, tracing and reproduction of maps took two years.

39. The above issues and increasing population demands reactivation of efforts rigorously and strengthening of geographical staff for improvement of census geographic work for the 1991 Census.

7. HIRING OF FIELD STAFF

40. To arrange and organize a large field force for census operation in a country like Pakistan is a big challenge. Population Census Organization can not maintain such a large field force on regular basis. This field force is hired from the Provincial Governments for census period. In this regard the role of Provincial Government is very vital and needs to be properly mobilized to

provide good quality staff. Generally un-willing and uncommitted staff is provided for census work. Though significant efforts are put in at training stage for generating motivation to do census work but still this staff remains passive and gives low priority to census work with few exceptions. Efforts are required to obtain good quality staff through greater persuasion of the provincial governments.

41. The phases of hiring staff from the Provincial governments from the hiring of Directors to head the Provincial Census Offices. Every time it is aspired that the Directors should be the officers who are competent, have sufficient administrative experience, sufficient knowledge of the province as a whole, command good reputation and can create good communication link with the Provincial Governments. But the officers provided are generally having qualities contrary to the above. As such it is being seriously thought that status of Directors is raised to Secretary level of the Provincial Governments. Another alternate being considered is that Provincial Census Offices are headed by the officers of the Population Census Organization.

42. Similarly it is being planned to enhance incentives for the field staff which may be monetary and also recurring incentives. Last time the honorarium for the field staff was though increased manifold but still dissatisfaction was expressed by most of the staff. Therefore, this issue needs sympathetic consideration. In the last census problems in obtaining required number of field staff in metropolitan areas was also faced. This problem needs to be visualised before hand. This time various available options be considered. The recruitment of special field staff in these areas for a month or so on regular basis needs to be examined. The services of retired governments officials, students, etc., could be obtained through advertisement in the press. The recruitment of female staff in urban areas can improve the situation in case such staff is also available.

8. TRAINING OF FIELD STAFF

43. Training is an important aspect in the execution of census plan particularly with a large field force. If the field staff is well trained and motivated they can produce good quality data. Therefore, efforts have to be made at appropriate time after finalising the census questionnaires, instructions, etc., to prepare an effective training plan. In the last census a direct approach was adopted to train the field staff which proved to be effective and is expected to be continued this time. The direct approach here refers that the field staff is trained by a census trainer at all steps of field hierarchy. The census trainers were taken from the college Lecturers, Assistant Education Officers and Headmasters of High Schools. They were trained by the senior officers (called Master Trainers) of PCO and FBS. Before the 1981 Census, indirect method was used in which the census district officers were trained in the first phase, who in turn, trained charge superintendents in the second phase, charge superintendents trained circle supervisors in the third phase who finally trained the enumerators. This method of training was found problematic as the basic content of training in terms of concepts and definitions changed from one phase to another. This time the excessive use of audio-visual equipment like TVs, VCRs, etc. is planned for training as this is a popular mean to bring changes in the attitude of the people and has a greater receptivity.

9. PUBLICITY

44. In modern age the publicity plays an important role in projecting the needs of census taking which calls for extensive and effective publicity. The publicity creates awareness in public that a national census is taking place on a particular date and a government representative, called enumerator, is visiting them for collection of information. It highlights importance of census as a basic exercise in the national statistical endeavour for socio-economic development and streamlining of administrative machinery. It seeks cooperation of informants because without their

cooperation census returns are likely to be spurious. It also creates confidence among informants about the confidentiality of the information about the individual characteristics and interest of the government in publishing the aggregate data of established administrative and geographical units.

45. For this purpose a publicity plan was prepared for the 1980-81 Housing and Population Census. This plan covered the Secretary's interview on TV, Census Commissioner's daily message on TV for a week before the census operation, Jingles, use of radio for delivering message especially to poor people who could not afford the expensive media like TV. Newspaper advertisement for three weeks before census operation; display of fifteen minutes documentary film in all the cinemas in Pakistan for one and a half month, message of President and Governors which were published in all the daily newspapers on the first day of the census operation, holding of seminars and publication of broucher in english and urdu.

46. Though this publicity plan gave good coverage in achieving publicity objectives but still there is a room for improvement in the plan for the next census. For example, though the people were successfully informed and they gave cooperation but still they remained unaware about the type of questions to be asked them and the correctness of the reply to be given. They bothered little about age reporting, resulting severe distortion in age pattern. Similarly response about occupational and industrial characteristics of the employed persons were also very poor. Therefore, people needs education for their greater participation in providing the more reliable information. This plan will be developed in consultation with Ministry of Information in promoting the cause of census. Under this plan a journal will be released for the education of masses about topics of information covered through census, especially those topics whose responses are generally poor. It will be a biannual journal. For creating interest of the readers this journal will also cover socio-economic and cultural aspects of the population.

47. Alongwith above media a proposal is being considered for the introduction of almanac calender and postals carrying census messages. The effectiveness of publicity largely depends upon the prioritisation of publicity media with respect to their receptivity by the target groups. In this connection the entire population of Pakistan can be divided into two major target groups. One group comprises literate people while other illiterate people. In 1981 about 26 per cent of the people were literate which is likely to reach around 35 per cent by 1991. But at that time literacy level in urban areas is expected to be 60 per cent and in rural areas around 25 per cent. It is easy to approach literate persons to elicit their cooperation for seeking reliable information through the use of any publicity media. The problem is with illiterate population, the largest target group, who can not be educated easily for census purpose except through radio and television. Again above 70 per cent of the rural people still do not have access to electricity therefore the use of television will not be so effective. However, this group have easy access to radio and it is found almost in every house in rural areas. It can not be denied that some interested people of this group generally keeps them inform about the national activities through others. These problems have to be addressed for prioritisation of various publicity media.

10. CENSUS PUBLICATIONS

48. In the past the practice was to release census data in the form of national, provincial and district reports and administrative report. However, for meeting urgent requirement of the Planning Division a short bulletins were released before the publication of the above mentioned reports. For the 1981 Census in addition to the above reports and bulletins, five handbooks were prepared; one for national and one each for provincial levels. The handbooks contain almost all published data in condensed form as rates and ratios and some derived statistics like singulate mean age at marriage, crude birth rate, gross fertility rate, total fertility rate, gross reproduction rate, age specific fertility rate, infant mortality

rate and child mortality rates. Population of administrative units and mauza lists at district level showing comparative population from 1951 to 1981 census were also published.

49. In addition to the above publications, for the first time we are also releasing census monographs on country socio-economic and demographic profile; migration and urbanization; fertility and mortality; economically active population; and literacy, enrolment and education. These are likely to be published by the mid of next year. Before the release of such condensed statistics, different data users generally manipulated census data according to their own desire and level of understanding about the subject resulting use of different estimates derived from the same census as data source. The endeavour of Census Organisation has brought a significant change in use of census data and its interpretation. It has not only made the life of data users easier but also persuaded them to use same estimates. It is also anticipated that census monographs will further facilitate them in the use of census data and its interpretation, drawing conclusions, implications and recommendations for socio-economic development plans. It is, therefore, being thought to make these innovations as regular features of the census publications. In view of the increasing pressure of town planners a serious thought is also being given to release some more publications providing detail data of each city on population and housing characteristics.

11. BUDGETARY REQUIREMENT

50. For the 1981 Census the sanctioned budget which included establishment charges, field staff honoraria, printing, publications and census kits for the financial year 1980-81 was Rupees 43.18 million, for 1981-82 Rupees 15.1 million and for 1982-83, Rupees 7.56 million. The cost per person was Rs. 0.8 (US 6 Cents). The amount of honoraria was paid to the field staff at the rate of Rupees 300 to enumerators, Rupees 450 to Circle Supervisors, Rupees 600 to Charge Superintendents, Rupees 750 to Census District

Officers and Rupees 100 to Trainers. Charges on purchase of marking ink Rupees 5000, patrol/oil Rupees 3000. Revolving sum of Rupees 1000 for meeting miscellaneous expenditure were allocated for each census district.

51. The cost of living has increased tremendously since the 1981 Census. Also population is expected to increase by more than 30 per cent. In view of the increasing cost of living and past experience the Census Organization is thinking to double the amount of honorarium to field staff including trainers. Cost of logistic requirement and other charges has to be linked to the average cost of living expected from census planning stage upto the release of last publication. Therefore, the increasing cost of living and the expected population in 1991 has to be considered while allocating budget for the next census.

THE COMPARATIVE ADVANTAGES AND DISADVANTAGES OF THE FOUR DESIGNS

Design-I Sample design on the pattern of 1972 Census	Design-II Sample design on the pattern of 1981 Census	Design-III Sample design suggested by the Consultants	Design-IV Systematic
<u>GENERAL</u>			
I. Gives less variance than designs II, III and IV	Gives less variance than design III	Gives more variation than designs I & II	Give less variance than designs II & III
II. Unbiased	Unbiased	Unbiased	Unbiased
III. Difficult to select and complex to develop estimates	Difficult to select and complex to develop estimates	Easier to select sample and to develop estimates	Selection depends upon the numbering of all households in serial order.
IV. House listing is the pre-requisite	House listing is the pre-requisite	No pre-listing is required for selection of sample	House listing is the pre-requisite
V. Possibility of under or over enumeration of households in selection of housing units in cyclical order by the enumerators. Also possibility of errors due to marginal cases.	Possibility of committing errors due to marginal cases.	Possibility of committing errors due to marginal cases.	Possibility of errors through advertant replacement of sample households of large size by non-sample households of small size.
VI. Difficult to supervise work of enumerators due to scatterness of sample.	Easier to supervise the work of enumerators due to less number of blocks in the sample compare to designs I & III.	Easier to supervise the work of enumerators due to less number of blocks in the sample compared to design I.	Difficult to supervise due to wide spread of sample households.
<u>CONDITIONED TO PHASING</u>			
VII. Listing operation has to be carried before other operations.	Sample count follows the Housing Census while Population count can be combined with either of the two.	Three operations viz; Housing Census, Population count and sample count of Population Census can be combined.	Listing operation required before other operations.

Design-I Sample design on the pattern of 1972 Census	Design-II Sample design on the pattern of 1981 Census	Design-III Sample design suggested by the Consultants	Design-IV Systematic
VI.II. Requires more number of sample blocks than design II & III and more enumerators than for design III but less than for design II.	Requires less number of sample blocks of relatively big size and less number of enumerators for each operation but more for both operations.	Requires more number of sample blocks of relatively small size and hence more enumerators compared to design II or requires longer period of enumeration.	No such condition is required.
IX. Requires separate training in two batches - one for Population Census (Complete Count) and one for Housing and sample counts.	Requires separate training in three batches-one each for Housing Census, Population Census (Complete Count) and sample count.	Requires separate training of enumerators in two batches - one for Housing & Population Complete Count, and other for sample count.	Requires separate training of enumerators for sample count.
X. Difficult to link with complete count information without introducing bias through ratio estimation.	No problem in linking the information with complete count but difficult to coincide the results of two simultaneous operations.	No such problems	Difficult to link with complete count information without introducing bias through ratio estimation.
XI. Less chances of losing enthusiasm because of change of mode from complete count to sample count and engagement of special enumerators for the later operation.	More chances of losing enthusiasm in the 2nd operation.	No such problem	Enthusiasm depends upon the decision of combining Housing Census and complete count of Population Census.
XII. Ranks at second with respect to cost.	More expensive than designs I & III.	Least expensive	More expensive than all other designs.

13. THE POPULATION CENSUS OF PAPUA NEW GUINEA
WITH SPECIAL REFERENCE TO THE 1990 CENSUS

TWELVETH POPULATION CENSUS CONFERENCE
(Beijing, China, 5-9 September 1988)

THE POPULATION CENSUSES OF PAPUA NEW GUINEA
WITH SPECIAL REFERENCE TO THE 1990 CENSUS
(A COUNTRY PAPER)

MR NICK SUVULO
NATIONAL STATISTICIAN
NATIONAL STATISTICAL OFFICE
PAPUA NEW GUINEA.

The population censuses of Papua New Guinea
with special reference to the 1990 census

1. NATIONAL BACKGROUND

Papua New Guinea, which became an independent state in 1975, comprises the eastern half of the island of New Guinea and some 600 islands in its neighborhood, the largest of which are New Britain, New Ireland and North Solomons. The terrain of the mainland and the other larger islands is rugged, with high mountain ranges, some of them exceeding 4,000 metres. Rivers flowing fast from the mountain ranges descend to coastal plains and join some of the world's largest swamps. The coastal population is thinly clustered in villages, except in a few areas of high density.

According to the population census of 1980, Papua New Guinea had a population of a little over 3 million. The estimated population for 1990 would be 3.7 million. In relation to a land area of over 460 thousand square kilometres, the population density is estimated at 7 persons per sq. km. Over a third of the population lives in the densely populated highland valleys of the mainland.

Papua New Guinea is noted for its linguistic diversity. Over 700 dialects are spoken. However, Hiri Motu and Melanesian Pidgin are now generally understood and widely spoken. English is the language of education and administration. The main means of transport is still by air and sea. Roads are slowly stretching out into rural areas but the capital city is still not linked by road to the highlands or to the other provinces.

For administrative purposes, the country is divided into 19 provinces, each with its own government, with the capital district, centrally administered. There are in all 20 primary administrative units.

2. HISTORY OF CENSUS TAKING

The first census of Population in Papua New Guinea was held in 1966. Prior to 1966, there were two distinct types of census. The less important one was that which covered the non-indigenous population at the same time as the Australian census and the other more important one was the administrative census of the indigenous population. These censuses collected some information for administrative purposes and a summary of population figures for each patrol area or census division. They were of limited value because of the few items which were only available in summary form, a lack of simultaneity and incomplete coverage. The 1966 census was the first attempt to provide estimates of the total indigenous and non-indigenous population at one point in time. A complete enumeration was conducted in all urban areas and rural non-villages, and a ten percent sample was taken of the traditional villages. The first census was generally accepted to have been a success.

The 1971 population census was a virtual repeat of the 1966 operation. But the social and political climate had changed considerably and the time was not very suitable for conducting a census. Only a short preparation time was allowed and direct assistance from Australia was considerably less than in 1966. Preliminary results showed a major under-count. By using limited post enumeration survey data and demographic rates calculated from 1966 and 1971 census data, a set of adjustment factors were produced and only the adjusted results were published.

Proposals to continue a five year census cycle by holding a census in 1976 was deferred each year until a firm date was finally fixed for 1980 to be the first post-independence census.

3. 1980 POPULATION CENSUS

3.1 Preliminary preparations

Unlike the 1971 population census, the 1980 census, being the first post-independence census, had sufficient time for preparations and support from the government. The two previous censuses were largely designed by outsiders and processed in Australia. The Papua New Guinea statistical office had only a small permanent population section and a recently established field survey section on which to build the new census project. It was not until the UNFPA appointed a census advisor in September 1977 and the government appointed a Census Director in January 1978 that serious planning commenced for the 1980 census.

3.2 Complete enumeration - It became obvious from the beginning that the 1980 census would have little support unless it was a complete enumeration. Planners at the national and provincial governments required small area data. Determination of electoral boundaries was another major requirement needing small area data. But uniform application of coverage rule and enumeration method throughout the country was not feasible because of scattered and widely varied physical, social and cultural features of the land and the people.

3.3 Enumeration Method - The universally practiced house-to-house enumeration method was applied in the urban and rural non-village sectors in all the three censuses since 1966. But for the large majority of the population in the rural village sector, house-to-house enumeration was not possible even in the 1980 census owing to field problems. In the Highlands Region, settlements are usually not nucleated type but very dispersed; houses for one clan are often miles apart and scattered across rugged, virtually inaccessible terrain making house-to-house enumeration almost impossible. The alternative was to use the established procedure of administrative census in which all the people for the village or clan gathered at a central point. This was called the 'line-up' system of enumeration or group enumeration. The percentage turnout in this system of enumeration had been decreasing and in 1980 census it was often only 50 percent and sometimes as low as 20 percent. This situation obviously leads to the dangers of coverage errors in the census.

3.4 Coverage Rules - Unlike most other countries, a uniform coverage rule was not applied in the Papua New Guinea census. Rural enumeration was to be spread over at least three months and urban enumeration over two weeks. No fixed census night was possible in either sector. The de-jure concept of a usual resident was reasonably applicable in rural villages. However, in urban areas, previous surveys had shown no clear interpretation of the term, usual resident, except for long term urban dwellers. It was also essential that the rules be as simple as possible as the educational standard of many of the field staff and respondents would be quite low. Finally a mixture of the de-facto/de-jure rules were applied. In urban areas and rural non-villages persons who spent the night previous to the day of enumeration at that dwelling were counted. In rural villages, all persons who usually lived in that village, but who did not spend the night previous to the day of enumeration in an urban area or rural non-village were counted.

3.5 Questionnaire - The 1980 census being the first census after independence, the census data users were pressing strongly for a wide range of questions. It was, however, realized that the quality of field staff was likely to be low, particularly in remote rural areas and a complex questionnaire could not be canvassed there. Thus a compromise strategy was adopted which directed that:

- All urban areas, rural non-villages and a sample of rural villages be enumerated with a 'long' questionnaire.
- The remaining rural villages be enumerated with a 'short' questionnaire.

After consultation with the Census Users' advisory Committee, the 'Long Form' finally consisted of 26 questions and the 'Short Form' had 12 questions. All 'Short Form' questions appeared on the 'Long Form'.

- 6 Operational Strategy - The initial plan of collecting all the information on both forms through a single visit was changed as the field staff had considerable difficulties in managing two questionnaires in a single visit. Thus a two-phase operation was conducted; the first phase being held in 1980 and the second in 1981.

In 1980, the 'Long Form' was used to collect information for all persons in the urban and rural non-village sectors and for all persons in a few selected parts of the rural village sector. The 'Short Form' was used in 1980 to collect information for all persons in the remaining parts of the rural village sector.

In 1981, the 'Long Form' was used to collect information from a sample (7%) of the villages in the areas where 'Short Form' had been used in 1980. The sample data collected were used to arrive at estimates for each province relating to the additional topics on the 'Long Form', in particular, migration, fertility and mortality.

- 7 Census operation - Because of the difficult terrain in most part of the rural villages, census enumeration was conducted through a patrol system. A patrol consists of a supervisor and a group of 4 or 5 interviewers. Patrol duration averaged about three weeks to enumerate all the Short Form census units in a particular census division. In urban areas interviewers were allocated one census unit of about 50 dwellings to be enumerated over a two week period. Interviewers were grouped into teams of about five under a supervisor. In rural areas, the census enumeration under the patrol system took place during the period from mid-July to September 1980. Rural non-villages were enumerated in late August and September 1980. All urban areas and some easily accessible rural areas were enumerated in the main enumeration period between 22nd September and 3rd October 1980. The second phase of rural sample enumeration came into operation 12 months after the first phase and was staggered over a period of 9 months in different provinces.

- 8 Data processing - All data processing and tabulation work were carried out in PNG for the first time. All coverage checks, quality control work, manual editing and coding were performed by the census staff before the data were entered into the computer. A Honeywell level 6, Model 57, one Megabyte, central processor and 20 visual display units were used to perform all the data processing. Beginning from January 1981 all short-form and long-form data from 1980 operation were processed by Christmas 1981. Only part of the 1981 sample data remained at the end of the year which was completely entered by early April 1982. Tabulations were produced using the 'COCENTS' tabulation package.

9 Publication and dissemination of Census Results - The first Census results were the preliminary Census Unit totals based on field counts which were published on 30 December 1980. Final tables of the 1980 data began to appear in mid 1982. The tables were made available to users as they were produced. Final tables including the 1981 data were all available by the end of 1982. The first major publication was the Pre-Release: Summary of Final Figures, giving major summary tabulations at the national level. A provincial summary bulletin for each of 19 provinces was published. A national bulletin corresponding to the provincial bulletins was also published. Research monographs containing detailed analysis of census results on a specific topic such as urban migration, fertility and mortality were also prepared and published. Seven such Research Monographs based on the 1980 census results were released.

4. PLANS FOR THE 1990 POPULATION CENSUS

- 4.1 Previous census recommendations - The technical and administrative reports prepared by the census director and other officers of the 1980 population census offer a wealth of material to consider before embarking upon planning the next decennial census. These reports contain detailed description of the various alternative plans, proposals, activities, etc and the relative advantages and disadvantages in carrying them out. Based on their experiences and difficulties, they have also put forward suggestions for the future census planners to consider. These technical reports thus provide important inputs for planning the next census.
- 4.2 Census legislation - The 1966 and 1971 censuses were conducted under the authority of the census (Papua New Guinea) Act 1966. This Act was considerably out of date after the independence in 1975. So work commenced on a revised Statistical Services Act in 1976. It was intended that this Act covers both the population census and all other statistical collections. But this Act was not fully approved until February 1981. Thus the 1980 census had to be conducted under the revalidated census (Papua New Guinea) Act 1966. The Statistical Services Act 1980 which was approved in 1981 and now in operation, empowers the National Statistician to conduct the census of population. It is under this act that the 1990 census will be conducted.

- 4.3 Preparatory work - Although the population and Social Statistics Division of the National Statistical Office initiated some correspondence with the Provincial Authorities in preparation for the 1990 census in late 1987, active planning for the next census really started from the beginning of 1988. Census Users' Committee was formed and two meetings of the committee were held in March and April 1988. These meetings considered mostly the users' requirement of census data and an outline plan for the 1990 population census. Preliminary views and suggestions of the Regional Advisor for censuses and surveys, ESCAP were also obtained in this meeting and in subsequent discussion with him. A high powered National Steering Committee consisting of secretaries and their representatives from various concerned departments for obtaining and coordinating their active support for conducting the census is in the process of formation and its first meeting is scheduled to be held on 18th October 1988. One provincial steering committee for each of the 19 provinces will also be formed from early 1989. The National Executive Council, the cabinet of the government of Papua New Guinea, has not only approved, in April 1988 the taking of census in 1990 but also has given a directive to all Ministries and Departments to assist the National Statistical Office to conduct this census.
- 4.4 Enumeration Method - With a view to reducing the coverage error in the census of rural village sector, where the 'line-up' system of enumeration was applied in the 1980 census, the house-to-house enumeration method is planned to be extended to the greatest possible extent. Only the inaccessible and most remote rural villages may have to be enumerated through the 'line-up' system. A list of such villages is in the process of preparation. Special provision of transportation for such areas at the time of enumeration may reduce their number significantly from being enumerated through 'Line-up' system. House-to-house enumeration method will be applied in all other urban, rural no-villages and accessible villages. As the extent of house-to-house enumeration will be much larger in 1990 census compared to the 1980 census, the extent of preparatory work for the 1990 census, such as rural mapping, listing of households, delineation of census units, etc. will be much larger compared to the 1980 census.

- 4.5 Coverage Rules - As comparability of census figures between the censuses is one of the criteria to be followed in the census, the coverage rules followed in the 1980 census will be repeated in the 1990 census. An attempt will be made to eliminate the third sector of population in the country i.e. rural non-village sector. This sector consists of district offices, mission stations, aid posts, plantations, mining camps, settlement schemes, schools, etc. which are mostly service centres adjacent to the villages. These centres will be considered urban if they fulfil the urban criterion. Otherwise, they will be treated as rural. Thus it is expected that the 1990 census population will have only two classifications namely, urban and rural. This will reduce many problems of processing and tabulation of the 1990 census data.
- 4.6 Questionnaires - The views expressed in the census users' committee meetings showed that the wide variety of information collected in the 1980 census should almost be repeated in the 1990 census. But all these information totalling about 30 items cannot be canvassed in every household throughout the country owing to low literacy rate among both the interviewers and the respondents who will find it difficult to understand some of the complex items of information in the questionnaire. Thus a two-form strategy as applied in the 1980 census is planned to be followed in the 1990 census. About 14 items of information which are basic and easy to understand will be canvassed universally in all the households as a 'Short Form' questionnaire. The other items of information including these fourteen questions will comprise the 'Long Form' questionnaire, which will be canvassed to a sample (about 10%) of households throughout the country.
- 4.7 Operational Strategy - Although a two-form strategy is being adopted for the 1990 census, a single phase enumeration operation is being planned as apposed to the two-phase enumeration operations in the 1980 census. Both the Short Form and the Long Form will be canvassed simultaneously during the main enumeration period 2nd to 11th July 1990 by two different interviewers. While the Short Form will be canvassed in all the census units throughout the country, the Long Form will only be canvassed in about 10% of the census units. But all the households in the selected census units will be interviewed by a specially selected and trained interviewer during the main census enumeration period.

- 4.8 Estimate of coverage error in the census - The 1980 population census was the first full enumeration census in PNG. Complaints of under-enumeration were received from a particular province which were subsequently nullified by the figures of out-migrants resident in other provinces. However, some areas of under-enumeration and of over-enumeration were identified. But no post enumeration survey could be undertaken to arrive at an estimate of the overall coverage error in the 1980 census owing to lack of necessary resources.

Post enumeration Surveys for both the census enumeration and the sample enumeration are being planned for the 1990 census so as to arrive at the coverage errors in the census as well as in the sample survey.

- 4.9 Data processing plan - Considering recent developments in computer hardware and software, it is clear that the existing computer hardware in the National Statistical Office (NSO) is barely comparable in performance with those available in the market. An evaluation of the existing computer facilities in the NSO and recommendation for the requirement of 1990 census data processing were sought from a mission consisting of two Regional Data processing advisors one from ESCAP and the other from SPC. This mission paid a visit in April 1988 and submitted its recommendation. Considering the volume of data to be processed for the 1990 census, the mission did not recommend any big frame but recommended several micro-computers. The system recommended by the mission is a Compaq Desk pro 386, Model 300 type which is likely to be used in several countries in the Pacific region for the 1990 round of censuses. UNFPA assistance in securing these micro-computer facilities are being sought.

4.10 Tabulation and publication of Census Results - The tabulation of the 1980 census data was not straight forward for some important items due to two-phase operation strategy and many classifications of the population viz. urban, rural non-village, rural sample village, rural village, etc. in the 1980 census. For arriving at the national total for any item of information, the sub-totals for the different sectors of the population were to be summed up. The national total for some characteristics which were collected on sample basis in the rural sector, had to be obtained by summing up the rural sample estimate with urban, rural non-village and rural village full enumeration sub-totals. All these classifications and operational divisions made the tabulation difficult and clumsy. The 1990 census operations are being designed to minimize these tabulation difficulties and to make them straight forward. The 'Short Form', which will be universally canvassed, alone will provide full enumeration totals for each item contained in it for the country and for other administrative units. And the 'Long Form', which will be canvassed on a sample basis, will provide estimated total for each item contained in it for the country and for each of the 19 provinces by district. Results obtained separately from 'Short Form' and 'Long Form' operations need not be summed up for any characteristic.

The publications of the 1990 census results will generally follow the pattern of publications issued for the 1980 census. These publications will begin with a preliminary release of census field counts within three months from the census enumeration date. Census monographs, which were a special feature of the 1980 census, may also be repeated in the 1990 census for providing comparable results and analysis on special topics such as urban migration, fertility and mortality.

5. CONCLUSION

In planning the 1990 population census, the difficulties and problems encountered in the past censuses especially the 1980 census and the recommendations made by the past census directors and officers are being carefully considered. Comparability with the 1980 census will be maintained as far as possible so that the results remain comparable for the two censuses.

As the house-to-house enumeration will be extended to the large majority of the rural population, an army of 18,000 enumerators and supervisors will be casually recruited throughout the country so that the census enumeration could be completed within the 10 day enumeration period. Elaborate preparation for delineating the census units throughout the country and for casual recruitment and training of the census enumerators and supervisors are in the process of development. Data processing operations will be fully tested before they are introduced into the system with the assistance of the ESCAP Regional Advisor.

Finally, it may be said that Papua New Guinea is one of the most difficult countries in the world for taking a census. This statement testifies to the fact that the 1980 census was the first census taken to cover the entire population. Its significance is brought forward by the selection of a photograph depicting an enumeration scene in PNG for the cover of the book "Census of Asia and the Pacific, 1980 Round", an East-West Population Institute publication of which all of us are well aware.

14. URBANIZATION INDEX BASED ON THE CENSUS OF
POPULATION AND HOUSING RESULTS OF PHILIPPINES

URBANIZATION INDEX BASED ON THE CENSUS
OF POPULATION AND HOUSING RESULTS

CONTENTS

	Page
I. BACKGROUND	331
II. URBAN DEFINITION: THE PHILIPPINE EXPERIENCE	
A. The 1939 and 1948 Definitions	332
B. The 1956 Urban Definition	333
C. The 1963 Definition	333
D. The 1970 Definition	333
III. METHODOLOGICAL CONSIDERATIONS	
<u>Factor Analysis</u>	336
<u>Discriminant Analysis</u>	339
IV. IDENTIFICATION OF URBAN AREAS	
A. Use of Discriminant Scores	342
B. Cut-off Limits Based on Raw Scores of Key Variables	342
C. Identification of Urban Areas Below City/Municipality Level	344
V. URBAN AREAS: A REDEFINITION	344
VI. CONCLUSION	346
BIBLIOGRAPHY	347

URBANIZATION INDEX BASED ON THE CENSUS OF POPULATION
AND HOUSING RESULTS*

By: NELIA R. MARQUEZ**

I. BACKGROUND

Urbanization is a phenomenon of important concern to planners and policy makers. Trends and patterns of urbanization have implications for socio-economic development and vice-versa. The United Nations in its report on "Urbanization: Development Policies and Planning" acknowledged the difficulty of aggregating international statistics on urban population because of non-comparability of definitions.

The U.N. had recommended two criteria to generally distinguish the urban element in a country's population. These are (1) quantitative, in terms of population concentration, and (2) qualitative such as characteristics of the economy and mode of living. According to the U.N., the correlation between these two criteria varies with time, location, and the "urban" features of a particular locality. However, because of the divergence in the urban definition used by many countries, it adopted a simplified definition of urban place which is "that of localities with 20,000 or more inhabitants." Localities with 20,000 to 99,999 inhabitants are referred to as "towns" and larger localities as "cities." Although the U.N. came up with this recommendation, it has taken the position that given the different conditions obtaining in various countries, each country is the best judge of what criteria will best define urban in its particular circumstances.

It is recognized that population size and density are the most common indicators used by many countries to identify urban areas. As early as 1938, Wirth had identified three key features of city life in his article "Urbanism as a Way of Life." These key features are size, density, and heterogeneity. However, other scholars claim that urbanization cannot be explained in terms of these three features alone, or that one or more of these key factors are superfluous in the explanation of urban phenomenon. Wirth defended his stand by saying that "one should not confuse urban way of life with some historical conditions under which it is now observed." For this reason he cautioned his critics not to assume in advance that (1) all cities have their economy on the factory and large firms; (2) all city centers are surrounded by slums; (3) mechanical means of transport are dominant in all cities and (4) all cities generated their residents a minimum of sanitary services.

*

Paper to be presented in the Twelfth Population Conference in Beijing, China on September 5-9, 1988.

**

Deputy Administrator, National Statistics Office, Philippines.

Wirth did not assert, however, that size, density and heterogeneity were the "prime movers" and that the other features of the city were merely their consequence.

The Philippines has adopted various definitions of urban areas and the latest is that of the 1970 version. The 1970 definition of "urban" remains inadequate and problematic, often leading to a misclassification of areas or to an over- or underestimation of the level of urbanization. Based on the present definition, sometimes an area may be classified as urban but the population per se might be in a more rural environmental condition than those classified as rural areas. Because of the shortcomings of the present urban definition, the National Economic and Development Authority (NEDA) through its Inter-Agency Committee on Population and Housing had created a Technical Working Group (TWG) on Urbanization Index whose function was to study an alternative definition of an urban area in response to the requirements of planners and policy makers.

II. URBAN DEFINITION: THE PHILIPPINE EXPERIENCE

The level or index of urbanization for statistical purposes is generally defined as the proportion of the population residing in urban areas to the total population. The definition of an urban area in the Philippines has undergone several modifications. This is understandable considering that urbanization is a dynamic process such that classification of areas into various levels of urbanization is necessary over time. Consequently, indices of urbanization and/or their levels are subject to change. The fact that many countries have altered their definition of urban through time and/or developed concepts to supplement urban definitions is clear indication that governments are sensitive to settlement patterns and their implications on overall national development.

A. The 1939 and 1948 Definitions

The then Bureau of the Census and Statistics (BCS), now the National Statistics Office (NSO) had set criteria in 1939 to distinguish the urban population from the rural population. Classified as urban areas during the period were administrative centers or poblaciones, irrespective of population size, of chartered cities and municipalities and the entirety of some large cities existing at the time. This definition was not quite satisfactory because it included in its entirety the population living in big cities although there were certain portions in these cities which were more rural in character than those earlier classified as rural.

In 1948, a modification in the definition was made to include as urban areas only the poblaciones or central districts of chartered cities and provincial capitals plus the poblaciones of all municipalities and municipal districts. Big cities were no longer considered "all urban" in their entirety.

B. The 1956 Urban Definition

Because of many developmental changes that took place in chartered cities and in the Metropolitan Manila area, the definition of an urban area for purposes of the Labor Force Survey, which was to be conducted for the first time, was expanded to cover the entire areas of chartered cities and municipalities designated as provincial capitals and all other poblaciones. Although majority of the cities and provincial capitals appeared to be largely urban, there were still others which covered large areas whose population was still living in relatively rural conditions.

C. The 1963 Definition

Recognizing the limitations of the previous definitions adopted by the BCS, new sets of criteria were formulated where population size, density and population agglomeration were considered.

Urban places included the following:

1. In their entirety, all cities and municipalities with a population density of at least 1,000 persons per sq. km.;
2. Poblaciones of other cities and municipalities with a population density of at least 500 persons per sq. km. and barrios with at least 2,500 population, and any barrio contiguous to a poblacion with at least 1,000 population;
3. Poblaciones of other cities and municipalities with a population of at least 20,000 persons and barrios with at least 2,500 population contiguous to the poblacion; and
4. All other poblaciones having a population of at least 2,500 persons.

The above definition did not include some qualitative aspect of the environment but it was the first attempt to incorporate in a census operation the identification of barrios which are contiguous to the town center or poblacion.

D. The 1970 Definition

Further refinements on the urban area definition were made in 1970 and a special barrio schedule was administered during the census operation in 1970 to gather some qualitative characteristics of the population at the barrio or village level. The revised criteria to identify an urban place, which is being adopted up to the present time are the following:

1. In their entirety, all cities and municipalities

B. The 1956 Urban Definition

Because of many developmental changes that took place in chartered cities and in the Metropolitan Manila area, the definition of an urban area for purposes of the Labor Force Survey, which was to be conducted for the first time, was expanded to cover the entire areas of chartered cities and municipalities designated as provincial capitals and all other poblaciones. Although majority of the cities and provincial capitals appeared to be largely urban, there were still others which covered large areas whose population was still living in relatively rural conditions.

C. The 1963 Definition

Recognizing the limitations of the previous definitions adopted by the BCS, new sets of criteria were formulated where population size, density and population agglomeration were considered.

Urban places included the following:

1. In their entirety, all cities and municipalities with a population density of at least 1,000 persons per sq. km.;
2. Poblaciones of other cities and municipalities with a population density of at least 500 persons per sq. km. and barrios with at least 2,500 population, and any barrio contiguous to a poblacion with at least 1,000 population;
3. Poblaciones of other cities and municipalities with a population of at least 20,000 persons and barrios with at least 2,500 population contiguous to the poblacion; and
4. All other poblaciones having a population of at least 2,500 persons.

The above definition did not include some qualitative aspect of the environment but it was the first attempt to incorporate in a census operation the identification of barrios which are contiguous to the town center or poblacion.

D. The 1970 Definition

Further refinements on the urban area definition were made in 1970 and a special barrio schedule was administered during the census operation in 1970 to gather some qualitative characteristics of the population at the barrio or village level. The revised criteria to identify an urban place, which is being adopted up to the present time are the following:

1. In their entirety, all cities and municipalities

a municipality and having a relatively small population, although the density may be high, may not have been classified as urban.

4. The classification of the economic activity of the population of a barrio/barangay was left to the judgement of the enumerator and was not based on the census results. The perception of the enumerator on the pre-dominance of agricultural or non-agricultural activities of employed persons may be inaccurate thus resulting to erroneous classification.

While the Census of Population and Housing may be able to provide the basic data requirement in the study of urbanization, the geographic level of observation is limited to the municipality/city level considering that censuses are taken on a sample basis and the data may be considered reliable at this level. At the village or barrio level only population size could be utilized due to the absence of data on area thus density can not be used to identify urban places. This limitation is also true to the study which will be presented.

III. METHODOLOGICAL CONSIDERATIONS

Several approaches or methods were tried to arrive at a definition of an urban area taking into consideration the function, the socio-economic conditions and population distribution and concentration of a particular area.

Initially, with the constraints on data availability, only demographic variables were used to facilitate the measurement of the level or index of urbanization.

The first attempt was to set up population size cut-off at the barrio level since density can not be generated due to lack of data on the areas of the different barrios. The result of the experiment is shown below:

Cut-off Population at Barrio Level	Percent Urban
2,000 & over	43.38
2,500 & over	35.44
3,000 & over	29.74
3,500 & over	25.64
4,000 & over	22.85
5,000 & over	20.40

Upon examination of the percentage of urbanity of areas known to be highly urbanized, like those in Metro Manila which yielded a low percentage of 19.19 for Pasay City and 45.95 per cent for Caloocan City at a population cut-off of 3,000 and over, the results were not acceptable. Because of the unsatisfactory result of this experiment this approach was abandoned.

The second attempt done to measure urbanization was to alter the first criterion in the 1970 definition by increasing

the density per sq. km. and the population size to 5,000 and 100,000, respectively. This was on the assumption that if an area has met these two conditions, it is almost certain that it is an urban area. To confirm this assumption, some urban indicators on social and physical characteristics were gathered at the city/municipality level, namely:

1. % employed in non-agricultural activities, 1975 Census
2. % built-up areas, 1978-1980
3. % of households with electric service, 1980
4. % of households with piped water, 1980
5. % of households with toilet facilities, 1980
6. % of households with urban housing (strong & mixed type of construction materials)

There were 15 areas which satisfied the density and population size criteria but upon examination of their characteristics, there was diversity in the characteristics of these areas and therefore the assumption which was made earlier was not confirmed. On the basis of this finding, it is not therefore advisable to use only population size and density to classify an area as entirely urban.

Some attempts were made to give scores to the different municipalities and cities on the basis of the six variables, giving an equal weight for each variable but setting the limits on the basis of the average scores to identify an urban area was difficult.

Regression analysis was also used to identify the factors which influence urbanization but there was a problem again of measuring the level of urbanization.

A factor analysis was also done to enable us see the underlying pattern of relationship that exists such that data may be rearranged or reduced to a smaller set of factors or components that may be taken as source variables accounting for the observed interrelationships in the data. The same problem of measuring the level of urbanization was also encountered but it gave us an idea of the key indicators of urbanization.

Since the problem is to discriminate the areas in accordance with urban characteristics, a discriminant analysis was performed. The application of discriminant analysis on the urbanization data served two purposes. First, is to statistically distinguish between urban and non-urban areas in the country; and second, to determine which of the variables under consideration can best measure the degree of urbanization of a particular area.

Factor Analysis

Factor analysis is a helpful tool in multivariate statistics which allows for the reduction of the dimension of the problem before proceeding to other analyses such as discriminant or cluster analysis.

The principal factoring with iteration, which can handle most of the factoring needs and the most widely accepted procedure, was applied to determine what would be the best linear combination of variables. That is, the particular combination of variables accounting for more of the variance in the data as a whole than any other linear combination of variables will be determined.

There were two factor analyses done, the first involved seven variables and the second one deleted the data on non-agricultural employment because the reference period was different from the rest. The seven variables which are at city/municipality level are: (1) percentage of households with water-sealed toilets, (toilet for short) for the exclusive use of the reporting household; (2) percentage of households using electricity for lighting, (electricity for short); (3) percentage of households residing in houses made of strong materials or makeshift materials since only slum areas in big cities have this type of housing, (construction materials for short); (4) percentage of households with piped water from a community water system (Water system for short); (5) percentage of non-agricultural workers to total employed persons; (6) population count of an area; and (7) population per square kilometer (population density for short). The results are as follows:

Table 1. FACTOR MATRIX USING PRINCIPAL FACTOR WITH ITERATIONS

VARIABLES	FACTOR LOADING		
	Factor 1	Factor 2	Communality
Analysis I:(7 Variables)			
V1 - Toilet	0.67352	-0.09404	0.46247
V2 - Electricity	0.82895	-0.21116	0.73175
V3 - Construction Materials	0.77541	-0.22037	0.64982
V4 - Water System	0.64874	-0.16456	0.44795
V5 - Non-agric employment	0.82289	-0.09379	0.68595
V6 - Population size	0.52286	0.62346	0.66209
V7 - Population Density	0.49884	0.57566	0.53577
EIGENVALUE	3.36227	0.81351	
PERCENTAGE OF VARIANCE	80.5	19.5	
CUMULATIVE % OF VARIANCE	80.5	100.0	
Analysis II (6 variables)			
V1 - Toilet	0.68217	-0.13281	0.48300
V2 - Electricity	0.77463	-0.20803	0.64332
V3 - Construction materials	0.78347	-0.26267	0.68282
V4 - Water system	0.65918	-0.20778	0.47769
V5 - Population size	0.53173	0.55948	0.59575
V6 - Population Density	0.53539	0.55476	0.59440
EIGENVALUE	2.6832	0.79385	
PERCENT OF VARIANCE	77.2	22.8	
CUMULATIVE % OF VARIANCE	77.2	100.0	

Based on the above results, there are two factors extracted by the principal factoring method from the original correlation matrices, which are shown in below:

Table 2. CORRELATION MATRICES

A. Original Variable Set

	V1	V2	V3	V4	V5	V6	V7
V1	1.00000	0.50064	0.53459	0.58203	0.54517	0.29242	0.28959
V2	0.50064	1.00000	0.73381	0.51889	0.76565	0.29454	0.30786
V3	0.53459	0.73381	1.00000	0.52201	0.63041	0.26064	0.28277
V4	0.58203	0.51889	0.52201	1.00000	0.52027	0.24178	0.21899
V5	0.54517	0.76565	0.63041	0.52027	1.00000	0.38323	0.34791
V6	0.29242	0.29454	0.26084	0.24178	0.38323	1.00000	0.59521
V7	0.28959	0.30786	0.28277	0.21899	0.34791	0.59521	1.00000

B. Reduced Variable Set

V1	1.00000	0.50054	0.53459	0.58203	0.29252	0.28945
V2	0.50064	1.00000	0.73381	0.51889	0.29475	0.30796
V3	0.53459	0.73381	1.00000	0.52201	0.26091	0.28285
V4	0.58203	0.51889	0.52201	1.00000	0.24178	0.21887
V5	0.29252	0.29475	0.26091	0.24178	1.00000	0.59517
V6	0.28945	0.30796	0.28285	0.21887	0.59517	1.00000

It can be seen that Factor 1 alone involving seven variables can explain 80.5 percent of the total variance and the inclusion of Factor 2 into the model will explain the remaining 19.5 percent of the total variance. Furthermore the variables V1 to V5 have high correlation with Factor 1 with V2 (electricity) having the highest correlation followed by V5, non-agricultural employment and V3, construction materials. The correlation of V1(toilet) with the first factor is a little more than that of V4 (water). Thus, Factor 1 could be interpreted as the physical indicator of urbanization since the variables having high correlations with this factor refer to the physical aspects of the area. Factor 2, on the other hand, may be taken as the demographic determinants of the degree of urbanization of an area because of its high correlation with V6 (population size) and V7 (population density). In terms of cause and effect relationship, the two factors may be interpreted as follows: Factor 2 seems to indicate the degree of urbanization of a particular area in relation to other areas and Factor 1 follows as a result of the urbanization of the said area.

The second analysis, where V5 (non-agricultural employment) was removed, showed some changes in the relative importance of the variables in Factor 1 and the explained variance had also reduced to 77.2 per cent. It can be noted that construction materials had the highest contribution to Factor 1, followed by electricity and toilet.

The interpretations given to the two resulting factors in the first analysis can also be attached to the resultant Factor 1 and Factor 2 in the second factor analysis. In addition, Factor 1 in the second analysis may be interpreted further as the visual aspects of urbanization. This result confirms the general impression that the tangible aspects of urbanization are the physical characteristics of the area.

While the factor analysis has successfully classified the six variables into two factors, it failed to provide parameters by which cut-off limits can be set to identify a city or municipality as entirely urban or not. Further analysis therefore is necessary to statistically distinguish the urban and non-urban cities and municipalities based on the six variables under consideration.

Discriminant Analysis

The discriminant analysis technique was employed as a means of classifying the cities and municipalities into specific categories. This technique computes the probability of membership of a city or municipality to a particular group based on the key variables earlier identified and assigns it to a group where its probability of membership is optimum.

An initial requirement in discriminant analysis is to group the different cities and municipalities according to some classification. Guided by the result of the Factor Analysis, three groups were set up based on arbitrary cut-offs on population size and density. It was assumed in this grouping that those areas falling under Group I are more urbanized than those in Group II and Group III.

The three groups classified are the following:

Group I - Those areas with population size of 200,000 and over and a density of at least 2,500 persons per sq. km.

Group II - Those area with population size of 100,000 to 199,999 and a density of 2,500 persons per sq. km.

Group III- All other cities and municipalities not falling in either Group I or II above.

To distinguish between the groups, four discriminating variables which were known to be urban characteristics on which the groups may differ were used. These variables are toilet, electricity, construction materials, and water system as earlier defined.

The first discriminant analysis performed using four variables (V1 TO V4) showed that 90.73 per cent of the cities and municipalities were correctly classified based on the above.

groupings. Only 9.3 per cent of the areas possess characteristics which are quite different from the group they were initially classified. The following table reveals the extent of misclassification.

Table 3. PREDICTION RESULTS OF GROUPING

Actual Group	No. of cities/ Municipalities	Prediction Group Membership		
		Group I	Group II	Group III
Group I	11	9	2	0
Group II	15	4	10	1
Group III	1,539	19	119	1,401
Total	1,565	32	131	1,402

Except for Group II where one area does not belong to the more urbanized group (Groups I and II), most of the misclassified cases belong to Group III totalling 138. This misclassification shows that 19 cities/municipalities in Group III have characteristics similar to those in Group I and that some 119 cities/municipalities have more or less the same attributes as those in Group II.

For purposes of this study, however, there is no need to reclassify these areas because it is already certain that the initial classification used gave very satisfactory results. Table 4 will show the results of the two discriminant analysis done.

The first discriminant analysis yielded a 99.24 per cent explained variance attributed to the four discriminating variables. This variance is measured in terms of the relative percentage of the Eigenvalue in Function 1. Since Function 1 has already explained most of the variation, the second function may be ignored.

Among the variables, V1 (toilet) has contributed the highest with 0.625 which represents the absolute value of the standardized coefficient. The sign attached to the coefficient merely denotes whether the variable is making a positive or negative contribution to the discriminant function. The other two predictors which can be considered also as key indicators of urbanization are electricity and construction materials.

In the factor analysis which was done earlier, construction materials gave the highest correlation followed by electricity and toilet. Both statistical methods of analysis (factor and discriminant) yielded consistent results in the identification of the three major indicators of urbanization.

Since one of the objectives in the use of discriminant analysis is to reduce the number of variables to as few as possible, water was removed in the second analysis. The

exclusion of water is dictated by its low contribution in the first function and therefore is the least predictor of the precise group to which a city/municipality belongs. Besides the correlation matrix as shown in Table 5 shows a high correlation between toilet (V1) and water (V4). In other words, it follows that if there is a flush toilet, there is the presence of a water system.

Table 4. RESULTS OF DISCRIMINANT ANALYSIS

VARIABLES	STANDARDIZED COEFFICIENT		Percent Correctly Classified
	Function 1	Function 2	
Discriminant Analysis I:			
V1 - Toilet	-0.62497	0.56672	90.73
V2 - Electricity	-0.37318	-0.13026	
V3 - Const. Materials	-0.15192	0.01151	
V4 - Water System	-0.03590	0.59341	
Canonical Correlation	0.403	0.038	
Eigenvalue	0.19313	0.001147	
Relative Percentage	99.24	0.76	
Wilks Lambda	0.8369	0.9985	
Chi-square	277.84	2.293	
Degrees of Freedom	8	3	
significance	0.0	0.514	
Centroid:			
Group I	-3.50790	0.30916	
Group II	-2.75245	-0.28788	
Group III	-0.05190	0.00060	

Discriminant Analysis II:			
V1 - Toilet	-0.62924	0.90695	90.48
V2 - Electricity	-0.38068	-1.12621	
V3 - Const. Materials	-0.15784	0.12136	
Canonical Correlation	0.402	0.034	
Eigenvalue	0.19295	0.00117	
Relative Percentage	99.40	0.60	
Wilks Lambda	0.8373	0.9988	
Chi-square	277.223	1.824	
Degrees of Freedom	6	2	
Significance	0.0	0.402	
Centroid:			
Group I	-3.50274	0.27609	
Group II	-2.75501	-0.25644	
Group III	-0.05188	0.00053	

Table 5. CORRELATION MATRIX OF DISCRIMINANT ANALYSIS I

VARIABLES	V1	V2	V3	V4
V1	1.0000			
V2	0.4342	1.0000		
V3	0.4762	0.7038	1.0000	
V4	0.5398	0.4741	0.4792	1.000

On the other hand, construction materials (V3) was retained in the list of variables although it is highly correlated with electricity (V2) because of its importance in the physical configuration of an area. Among the three key variables established, construction materials could easily be recognized by a mere visual inspection and therefore accurate data can be gathered for future classification of areas.

The outcome of the second discriminant analysis, which is also shown in Table 4 is nearly similar to the first set of analysis. There is a slight increase in the relative importance of Function 1 to 99.40 per cent. More importantly, however, is that the porportion of correctly classified cases decreased slightly to 90.48 per cent. In effect, the exclusion of V4 (water) did not affect much the group membership.

IV. IDENTIFICATION OF URBAN AREAS

A. Use of Discriminant Scores

To distinguish or identify cities and municipalities which are more urbanized, the discriminant scores were computed based on the three variables. The scores were from the lowest to the highest. Take note that the contribution of the discriminant variables to the first function (Function 1) as shown in Table 4 is negative. It is therefore expected that the lower the discriminant score of an area, the more urbanized it should be.

Upon examination of the cities and municipalities classified as more urbanized based on the arrayed discriminant scores, it was observed that two cities with very low percentage of strong and makeshift construction materials (31 and 39 per cent) were classified as highly urbanized. There were also areas where electricity usage was low and yet were classified as highly urbanized. Because of this problem, a more strict cut-off limits were imposed thus the use of the discriminant scores to identify highly urbanized areas was abandoned.

B. Cut-off Limits Based on Raw Scores of Key Variables

The alternative basis for setting cut-off limits are the raw scores reported for the three key variables isolated as predictors of urban areas. In order to ensure that only areas which are really urbanized will be classified as such, the mean scores of the three variables in Group I were used as the

acceptance level for areas to be considered as urban in their entirety. It was found, however, that these limits were too rigid because only few areas could pass the evaluation. After several cut-off marks were tested based on the mean score (\bar{X}) minus the standard deviation to lower the acceptance limit, it was finally decided that minus three standard deviations from the mean of Group I will be adopted as the minimum acceptable score for each of the three variables. Table 6 below shows the mean scores of the three variables and the corresponding standard deviations for each group.

Table 6. GROUP MEANS AND STANDARD DEVIATIONS

VARIABLES	GROUP NUMBER		
	I	II	III
Arithmetic Mean			
V1 - Toilet	69.66	55.23	15.93
V2 - Electricity	89.89	82.41	21.10
V3 - Const. Materials	95.40	84.94	34.39
Standard Deviation			
V1 - Toilet	12.25	17.51	15.05
V2 - Electricity	10.49	19.26	23.98
V3 - Const. Materials	12.58	22.65	21.61

In addition to the minimum acceptance level set for each variable, there was still a need to impose additional restrictions to avoid the inclusion of areas which may be more rural in character. Since the degree of urbanization can be measured in terms of population size and/or density of an area, it was also decided to incorporate any of the two in addition to the three key variables.

Based on the preceding discussion, the final criteria formulated to identify cities/municipalities as entirely urban are the following:

A city or municipality with at least a population of 50,000 or a density of 2,500 persons per square kilometer with the following characteristics:

1. at least 33 per cent of the total households use water-sealed toilets;
2. at least 58 per cent of the total households use electricity for lighting; and,
3. at least 58 per cent of the total households reside in houses made of strong and makeshift materials.

All the above criteria must be satisfied before a city or municipality may be classified as entirely urban. Cities and municipalities which do not meet this first level of evaluation are then considered in the second level evaluation, which involves the identification of urban barrios/villages.

C. Identification of Urban Areas below City/Municipality Level

For cities and municipalities which are not considered all urban in their entirety, the function of the area and population size were used independently to identify urban villages/barangays. All areas or barrios/barangays which are seats of political, economic, social, religious and cultural activities are considered as urban places, regardless of population size. These areas are usually called poblaciones or central districts of cities or city proper. It is assumed that the basic facilities and services in these areas are urban in character.

For other barrios/barangays which are located outside the town/city centers, a population size cut-off of 7,000 was used. The decision to adopt this cut-off figure was on the assumption that with 7,000 population which is approximately 1,200 households, basic urban facilities and services are present. The assumption concerning the presence of urban characteristics in the barrios/barangays with this population size was confirmed through a verification of the 1980 Barrio/Barangay Schedule collected during the 1980 Census of Population and Housing. This cut-off has to be further studied for possibility of lowering it as there may still be areas below this cut-off which are really urban in character.

V. URBAN AREAS: A REDEFINITION

On the basis of the studies made by the Technical Working Group on Urbanization Index, the proposed criteria in identifying urban places in the country require two levels of evaluation, the first level involves the evaluation of cities and municipalities which are to be considered urban in their entirety and the second level, evaluates each barrio/barangay in cities and municipalities which did not pass the first level evaluation.

The proposed definition requires that the following conditions be met:

1. To be considered as urban in its entirety, a city or municipality must have a population of at least 50,000 or a density of 2,500 persons per sq. km. and at the same time must meet the following characteristics:
 - a) at least 33 per cent of the total households use water-sealed toilets;
 - b) at least 58 per cent of the total households use electricity for lighting; and,

- c) at least 58 per cent of the total households reside in houses made of strong/makeshift materials.
2. All barrios/barangays in cities or municipalities not classified as entirely urban in criterion 1 will be considered as urban located:
- a) within the city proper or town center (commonly regarded as poblacion); and
- b) outside city proper or town center but having at least a population of 7,000 persons.

The following table shows the level of urbanization based on the present definition and the proposed definition.

Table 7. Level of Urbanization by Urban category, Philippines:1980
(Present and Proposed Definition)

Urban Category	Number of Cities/Mun.		% of total Cities/mun. Population		% of total Population	
	OLD	NEW	OLD	NEW	OLD	NEW
All Categories	1565	1565			<u>38.62</u>	<u>37.69</u>
Cities	60	60	75.10	76.34	15.90	16.16
Municipalities	1505	1505	28.83	27.31	22.76	21.53
I. Entirely Urban					<u>20.88</u>	<u>19.54</u>
Cities	20	15	57.66	52.89	12.21	11.20
Municipalities	49	36	11.00	10.58	8.67	8.34
II. Poblaciones /City Proper					<u>13.31</u>	<u>14.26</u>
Cities	*	45	9.56	12.20	2.02	2.58
Municipalities	*	1469	14.33	14.82	11.29	11.68
III. Non-Poblacion Barangays					<u>4.43</u>	<u>3.89</u>
Cities	*	22	7.89	11.25	1.67	2.38
Municipalities	*	51	3.50	1.91	2.76	1.51

* Information not available.

OLD - refers to present urban definition.

NEW - refers to the proposed urban definition

Source of Basic Data: 1980 Census of Population & Housing.

The comparative summary of both the present and the proposed definition of urban areas shows that the proposed definition is more strict in the sense that only 15 out of 60 cities were considered as all urban in their entirety as compared with 20 in the present definition. However, percentage-wise, the proposed definition had included more areas from cities as urban considering that there was a slight increase in the coverage from 75.1 per cent to 76.34 per cent of the total city population. Most of this increase came from city proper and city districts as reflected under Urban Category II above.

VI. CONCLUSION

This study is the first attempt to apply statistical methods in the determination of key indicators of urbanization. The findings of this study is significant towards the improvement of the definition of an urban area in the Philippines.

It should be noted that while the level of urbanization derived from the proposed definition is 0.93 percentage point lower than that of the present urban definition of the NSO, the two definitions used different criteria and the locations of identified urban places are not exactly in the same cities or municipalities.

The advantage of the proposed definition is its simplicity in administration in the future and with the assurance that some statistical methods were used to isolate the variables which reflect the best indicator of urbanization under Philippine conditions. Those areas which were identified in this study as urban places need not be evaluated in the future. Only those which are considered non-urban based on the 1980 Census of Population and Housing results may be evaluated.

A similar study using the same statistical methods in isolating urban indicators may be done after every census year to find out the change in the degree of urbanization and its effects in the environment. While three key variables may be the best indicators of urbanization at this time, there may be other variables that may come in depending upon the development of the country. As the degree of urbanization increases, the cut-off limits established may be reviewed for possible alteration in order to ensure that those cities and municipalities which are to be considered as urban in their entirety are predominantly urban in character.

BIBLIOGRAPHY

- Mijares, Tito A. et al. The Growth of Urban Population in the Philippines and Its Perspective, Technical Paper Number 5, Bureau of the Census and Statistics, Manila, Phil., (October 1973).
- Nie, Norman H., et al. Statistical Package for the Social Sciences (SPSS), 2nd ed. MacGraw-Hill, Inc, 1975, USA.
- United Nations, Urbanization: Development Policies and Planning, International Social Development Review, No. 1, ST/SDA/Sec. XII, United Nations Publications, N.Y., 1968, pp 9-10.
- Wirth, Louis. Urbanism as a Way of Life, American Journal of Sociology, XLIV, No. 1 (July 1938).

**15. MEASUREMENT OF ECONOMIC ACTIVITY
IN SRI LANKA**

MEASUREMENT OF ECONOMIC ACTIVITY
IN SRI LANKA

by

R.B.M. Korale

PAPER PRESENTED AT THE 12TH POPULATION CENSUS CONFERENCE

4 - 9 SEPTEMBER 1988, BEIJING, CHINA.

MEASUREMENT OF ECONOMIC ACTIVITY
IN SRI LANKA

by

R.B.M. Korale

PAPER PRESENTED AT THE 12TH POPULATION CENSUS CONFERENCE

4 - 9 SEPTEMBER 1988, BEIJING, CHINA.

MEASUREMENT OF ECONOMIC ACTIVITY
IN SRI LANKA

The principal source of statistical data on the economic activity of the population is the decennial census of population and housing. A regular household survey has not been conducted in Sri Lanka to measure the labour force characteristics of the population, ad-hoc surveys were undertaken to meet special needs, and therefore the role of the census in providing data on the work force has become all the more important. Further, data is also not available from an establishment survey or from an employment registration system. The employment exchanges which registered and processed the unemployed for placement mainly covered the urban sector and had become ineffective as a means of processing job placements due to high unemployment in the country and were discontinued in 1977. Thus it became necessary to fall back on the census of population to meet the data which would otherwise be gathered through these different sources.

The census data on employment and unemployment and occupational and industrial distribution of the population can be disaggregated to any local area level required for study. It has this advantage over labour force and other survey data which cannot provide estimates for small areas, because of sampling errors. Thus detailed statistics for small areas on a number of parameters such as on occupational and industrial distribution, employment status, age, sex, educational status, marital status, religion, and ethnic breakdown of the workforce is available from the census. In addition the census also provides data on the current and potential labour supply for small areas, which makes it possible to use it for projecting the supply of labour. These uses and possibilities have enhanced the value of the census as a source of labour force data, at a time when political devolution and administration decentralization is being implemented, through the establishment of Provincial Councils, and Pradeshiya Sabhas.

Differentiation by Activity and Occupation

Unemployment was not a serious issue of concern to the administration until the 1930s. Upto this time there was not much differentiation of labour by industry and occupation. Except for a relatively small urban labour force which found employment mainly in commercial enterprises ancillary to the plantations, the majority were engaged in rural peasant farming activities. The plantation industry depended very much on immigrant labour from India. Here the labour supply contracted through returned migration to India when the demand for labour decreased. It is said that the labour force in the plantation industry at the time expanded and contracted with the vagaries of the market, the weather and harvests in India. The acceptance of a policy to allow Indian immigrant labour to the estates and for some unskilled and skilled tasks in the urban sector did not require the generation of these vocational skills, these were significant features of the labour market at the time.

The early censuses of population collected information on occupation and employment status. But this information was not collected based on the activity status or economic activity of the population. The concept of activity, including economic activity is relatively new in the local context. The word industry is associated largely with factory or manufacturing industries. This weak attachment to industry and the lack of adequate terminology to exactly convey, the concept of activity status has been an issue which made it difficult to accurately measure the characteristics of the work force. At these censuses the concept of an attachment to industry was not followed and they attempted to gather information on occupation. The concept of occupation which was used in the past for differentiation by caste when used as a measure would have a bearing on the measurement of employment characteristics and dimensions.

The problems encountered by the census authorities in the measurement of employment is evident from the following paragraph. "Thus the Indian Census Report of 1931 began its Chapter on 'Occupation' by calling it "the most troublesome and complicated return called for on the census schedule." The Ceylon Census Report of 1921 averred that "the difficulty in obtaining accurate data is very considerable. The great variety of the different occupations, the difficulty of framing an unexceptionable classification of them, and of obtaining sufficiently precise information, either from the enumerator or the householder to enable specific occupations to be located in the classification, make the tabulation of the occupations a work of great complication, and the results are, in most countries, probably subject to considerable error". The Australian Census of 1911 described the field of investigation comprised under the head of 'occupations' as one of the most complex of those that came within the ambit of a census". Census of Ceylon: General Report 1946, p. 242.

Labour Market Situation and Characteristics

At the end of the 1940s by which time the country had gained independence the economy was characterised as a dual economy, with a plantation sector producing primary commodities for export, and domestic agriculture largely peasant based producing for home consumption. The manufacturing and tertiary sectors were small, and they too had developed as ancillary to the plantation industry. The labour market was segmented at this time with one serving urban sector employment, another for the plantation sector functioning in the enclaves and with a partially monetized labour market serving rural agriculture. The economic development policies and programmes undertaken during the past 3 - 4 decades have raised the share of the manufacturing, construction, wholesale and retail trade and

services sectors both in terms of output and in employment. Although output in agriculture in relative terms had declined from about 62 % in 1951 to around 24% at present (1986) the relative share of agriculture in total employment in agriculture had declined only from about 53.5% in 1953 to around 45.5% in 1985. Thus the productivity of this sector in relative terms is substantially lower than in respect of other sectors, and continued to hold the increased additions to the labour force in low productivity tasks, until they secure work in more remunerative sectors. This slow rate of growth of the economy except for brief periods during this time resulted in an inadequate diversification of the economy and created labour surpluses, and extended the period of waiting for employment for jobs in the modern sector. The rapid expansion of educational opportunities provided under a free education programme created increasing outputs from secondary and tertiary educational levels. This also contributed to an increase in the aspirations of the job seekers, and created a mismatch between available opportunities, and expectations. As a consequence the prevalence of discouraged workers has been evident from the early 1960s, and continues to this day. With this brief background it is possible to summarise the main features of the Sri Lanka labour market.

(1) Labour Market Segmentation

The labour market is segmented into the modern sector (mainly in urban industrial and services sectors), plantation agriculture, and a partially monetized rural labour market.

(a) Modern Sector

In the modern sector there is regularity of work, fixed hours and regularity in wages. Formal recruitment procedures based on educational skill requirement criteria are adopted. Inclusive of the public service this sector had expanded and is estimated to be around 20-25% of the total employed. The urban informal sector still remains small.

(b) Plantation Sector

The contraction of the extent under crops in plantation agriculture, has reduced the labour absorption possibilities of this subsector. There is a supply demand imbalance, and the adjustment practices adopted include the spreading of available work among the workforce. The inclusion of young adults as eligible workers raises the level of participation, but in adverse work situations contributes to spreading under employment.

(c) Rural Labour Market

The rural labour markets are not fully monetized and seasonal work is common. The extent of share cropping in paddy cultivation is estimated to be as high as 20% of the holdings cultivated. Payment in kind in addition to wages specially as a component of wages is still widely prevalent in peasant agriculture.

(2) The Mismatch of Available Opportunities and Expectations

The increased size of cohorts a result of continued population growth and expansion of educational opportunities under a free education scheme from primary through university education has resulted in significantly higher numbers entering the workforce with educational attainments at all levels when compared with the attrition due to retirements, deaths and other causes. The labour demand for many of the categories has not matched with the supply, and this mismatch is specially acute in respect of females. This mismatch has caused the creation of a pool of labour, specially female labour which could enter the labour force at short notice depending on the labour market demand situation.

(3) Seasonal Employment

The single largest category of labour in the workforce comprise agricultural labour in self employment and paid employment. Work intensity in agriculture is highly dependent on seasonality. Apart from agriculture even construction sector employment is to some extent determined by seasonal variation, specially construction and rehabilitation of irrigation works.

(4) High Labour Mobility

The vast majority of business enterprises and activities are small, and they employ a few workers, and much of this employment is of a casual nature and varies with the intensity of the work and production. Many undertake casual jobs for the supply of goods and services and geographical mobility is also a feature in the case of such activities as construction work undertaken by small contractors. All these lead to raising the employment mobility of persons between enterprises and even across occupations. These lead to short term changes in the activity status of persons which should be noted in deciding on methodology for labour force measurement.

(5) Absence of Employment Service Institutions

At present there is no institutional machinery for the registration of employment and job placement. Individual institutions select their own employees either by advertising them in the media or through informal means. Thus there is no mechanism for persons seeking employment to apply and register themselves for occupations of their choice. This lack of formal institutions for the unemployed to declare their availability for work, makes it difficult to ascertain the activity statuses of the individuals specially of females.

(6) Discouraged Workers

The continued expansion of educational opportunities and consequent rising aspirations, and the prevailing job profile which created the mismatch and the supply and demand imbalance in overall terms has increased the period of search for employment, and created the issue of discouraged workers. Therefore the need to include persons who are not actively seeking employment but who are available for work had been recognized as early as the Census 1971. The high level of unemployment recorded in the country for a long period of time, which is the result of a supply and demand imbalance in absolute terms as well, makes it necessary to recognize this issue of discouraged workers in the measurement of employment and unemployment.

These labour market characteristics are important in deciding upon the conceptual framework and the methodology that could be used in the measurement of the economically active population.

The Concept of Economically Active Population

Sri Lanka has generally adopted the ILO recommendations on the measurement of employment and unemployment and the UN recommendations concerning the measurement of economic activity. The International Standard Classification of Occupations (ISCO) with suitable modifications is in use in the classification of occupations and the International Standard Industrial Classification and the SNA have been adopted in measuring economic activity. The recent censuses and surveys undertaken by the Department of Census and Statistics (DCS) have adopted the concept of the "economically active population" for measuring labour force characteristics of the population. According to the ILO recommendations arrived at by the 13th International Conference of Labour Statistics (ICLS) "the economically active population" comprises all persons of either sex who furnish the

supply of labour for the production of economic goods and services as defined by the United Nations Systems of National Accounts and Balances, during a specified time reference period". The economically active population as defined by the resolution, allows for two measures, the "currently active population" or the labour force measured in relation to a short reference period, such as one week or one day, and the "usually active population" measured in relation to a long reference period, such as a year.

According to the definition of the currently active population or labour force, the population is divided into 3 mutually exclusive groups employed, unemployed and not in labour force, and the labour force comprise the employed and unemployed. The 13th ICLS has defined the employed and unemployed categories, and according to this definition the employed "comprise all persons above a specified age who during a specified brief period, either one week or 1 day were in (a) paid employment; or

(b) self employment.

Paid employment and self employment have also been defined in the resolution. Further the unemployed have also been defined, and they comprise all persons above a specified age group, who during the reference period were "without work", i.e. were not in paid employment or self employment and were "currently available for work", i.e. available for paid employment or self employment during the reference period and were "seeking work", i.e. had taken specific steps in a specified recent period to seek paid employment or self employment. The relaxation of the requirement "seeking work" in the standard definition has also been recognised where employment service institutions are weak and the labour market is unorganized and paid employment opportunities are not dominant.

The current status approach is appropriate where the secondary and tertiary sectors are large, and the labour market is characterised by regular full time paid employment. However, when there are significant seasonal variations and substantial labour force mobility in and out of economic activities, the cross sectional picture provided by a survey using this measure is not truly representative of the general pattern of labour utilization.

According to the definition of the usually active population in paragraph 7.1 of the ICLS Resolution 1982, the usually active population comprises all persons above a specified age whose main activity status as determined in terms of number of weeks or days during a long specified period (such as the preceeding 12 months or the preceeding calendar year) was employed or unemployed as defined in paras 9 and 10 and in paragraph 7.2 where this concept is considered useful and feasible the usually active population may be sub-divided as employed, and unemployed in accordance with the main activity. The population not usually active has also been defined and they comprise the functional categories: (a) Students; (b) Home makers; (c) Income recipients; and (d) Others (recipients of public aid or private support, children not attending school, etc.). The recommendation also allows the introduction of separate functional sub-categories to identify: (a) persons engaged in unpaid community and volunteer services; (b) other persons engaged in marginal activities which fall outside the boundary of economic activities.

The adoption of the concept of the usually active population is preferable in situations where there are seasonal variations in employment, where full time regular paid employment is not predominant, and where movements in and out of the activities and from the labour force is significant. In these situations this approach could provide a better

picture of the prevailing status, and is specially suitable at a census, where only a single one time measurement can be taken. The advantages of this method has been cited as allowing the measurement of the intensity in terms of the extent of employment and extent of unemployment over the entire reference period of one year, and also allowing a wider coverage and comprehensiveness through a listing of employed and unemployed, including persons who were economically active some time during the year, which would be missed under a current status approach. The data on the usually active population is also preferred for analyses using national accounting data and household income and expenditure data which are also prepared for one year reference periods.

According to this method the respondents are first classified as "usually active" or "not usually active" and then "usually active" are again classified as employed and unemployed. The reference period for the usually active population at a census can be fixed as the 12 months preceeding the census, and at a survey it allows the use of a moving reference period using the 12 month period immediately preceeding the survey month.

The classification as usually active or usually inactive is based on the time duration of there statuses. In practice the duration employed and duration unemployed are taken together and if this time duration in weeks or days exceeds the time duration recorded for the usually inactive then the respondent is classified as usually active. One important feature here is that a period such as off season or adverse weather conditions during which the respondent was not available for work should be excluded in determining the time duration. Thus with regard to seasonal workers they could become eligible to be classified as usually active and then as employed if they had worked for more than 13 weeks during the year. In

the sub-division of the usually active population into employed and unemployed under this concept, the longer time duration of these two statuses in the reference period is used as the determinant. Thus in this approach the persons classified as employed would have had some experience of a shorter time duration as unemployed, or even as not been in the labour force and thus their status need not be unique during the entire reference period. In view of the longer time duration involved the problem of recall certainly surfaces and how best to elicit this information on different statuses is an issue which is discussed later.

Measurement of Employment and Unemployment

From the inception of the measurement of the workforce or the economically active population in Sri Lanka, it appears that the total magnitude and the various labour force components ascertained through censuses and surveys have been subject to different margins of error and most of these parameters have been under estimated. The fluctuations observed in the labour force participation rates, and other labour force magnitudes were more due to these errors in measurement, which have arisen due to a number of factors. When one considers the long time frame of about 40 years methodological issues are very relevant. The conceptual framework used for the measurement of the economically active population has been evolving during this period.

Following the recommendations made at the time, the concept of gainfully occupied population was used in the census conducted in the 1940s, 50s and 60s. The gainfully occupied population "comprised persons who reported themselves as having an occupation or a profession or a trade from which they had earned an income, in cash or in kind, or in which they had assisted in the production of goods and services, regardless of whether they had actually worked or sought work during the current period".

Hussman, S.R. and Mehran, F. Currently Active Population, p.6. ILO. The measurement of the gainfully occupied population was based on the reporting of an occupation by respondents who could be retired persons, but excluded new entrants to the labour force, who did not have an occupation, and these were limitations when measuring the economically active population. This approach also did not provide for a proper integration of the employed and unemployed as parts of the economically active population.

The enumeration of the employed in the Censuses of Population 1946 and 1953 were based on this concept. The respondents were directly asked to state their attachment to an occupation. In fact in the Census of 1946 this question on occupation was listed fairly high in the schedule, before "district of residence" and "relationship to head of household", and the question on literacy. The attachment to industry was elicited after these questions and only from those who had responded to the question on occupation. The problems of measuring the attachment to economic activity by this distance between the questions on occupation and industry was understood, and in the Census 1953 the questions on occupation, industry and employment status were grouped together.

As regards the unemployed the Census 1946 enumerated only those who were previously employed but without work at the time of enumeration which number amounted to 25,082 and excluded all other categories of unemployed. The number unemployed at the time of this census had been estimated at 284,000. Snodgrass, R. (1968).

Even in the Census 1953 no effort was made to measure the unemployed component of the work force, and only those persons who were previously employed and currently unemployed were enumerated. Thus the concept of economically active population comprising persons who are contributing to the creation of economic goods and services and that component

which seeks to participate in these processes were not part of the measurement provided in censuses and conducted in Sri Lanka by the 1950s.

At the time the ILO Survey of Employment, Unemployment and Underemployment 1959 was conceived and executed, industry and occupation were not carefully differentiated and a combined classification had been used. In this survey the concept of gainfully employed population with modifications drawn from the labour force concept had been attempted. Even in the Census of Population 1963 although classification by activity status was attempted the employed were enumerated using the gainfully occupied population method. In measuring unemployment the labour force concept which was then available had influenced the definition of unemployed in this census, which included the element "seeking work". By this time the problem of unemployment had become significant and of interest to planners and policy makers with the Ten Year Plan of 1959 also focussing attention on employment. The concept used in the 1963 was too restrictive and could not capture the total unemployment and several researchers attempted to prepare adjusted estimates of the unemployed population. One adjusted estimate of unemployed was double the enumerated unemployed population. ESCAP (1974). Population of Sri Lanka. p. 227.

The unemployment problem continued to worsen in the country and the Census 1971 used a broader definition than that used until then. It covered both those actively seeking work and not actively seeking work. When one evaluates the data it is evident that the Census of Population 1971 has elicited the most comprehensive information on employment and unemployment available through any census or survey.

Evaluation of Employment Data from Census of 1981

As regards comprehensiveness and quality of data ascertained through the Census of Population and Housing 1981 it is now evident that the data on the economically active population has been subject to a number of limitations and these could be characterised as the weakest elicited through the census.

Considering the development and the recommendations made by the International Labour Organisations (ILO) and the United Nations (UN) on the measurement of the economically active population it was believed that it would be desirable to move away from a concept of gainfully occupied population adjusted with the usual status approach used in the Census 1971, and use the labour force concept for the Census 1981. The concept of the currently active population had not evolved fully at the time the 1981 Census was planned and executed in 1981, and a mixture of the currently active population, as now defined and the usual status approach had been attempted. In conformity with the labour force concept the population aged above 10 years were divided into employed, unemployed and the economically inactive, however the recommended short reference period of 1 week or 1 day and a priority rule was not accepted and a reference period of 1 month, the month being the 30 days preceeding the date of census enumeration was stipulated.

Although the employed and the economically inactive group, viz.. students, retired and unable to work, engaged on own housework were identified as economically inactive and separately listed, the unemployed were included in the residual and an attempt was made to enumerate them on the basis of a question, as to whether the respondent had been seeking work during the preceeding 30 days. This procedure of treating the unemployed as a residual, appears to have over enumerated the unemployed by including those who were marginally employed, such as casual workers on whom a strict definition on

employment had been used. The definition on employed did not specify a time duration of work during the reference period, to qualify to be included as employed. But in the case of casual workers (the majority of employed fall into this category) the definition of having to "work the major part" of the reference period of 30 days had been stipulated. This naturally under enumerated the employed population significantly, and this reduced the employed in respect of the agriculture, construction and manufacturing sectors.

In order to circumvent the problems of exclusion of seasonal workers due to the adoption of the currently active population measure, seasonal workers who had a usual occupation but did not do any work during the reference period "because of seasonal factors such as non-cultivation season, unfavourable weather, etc." were regarded as employed. The reference period of 1 month used in this census did not conform to the measurement of current activity based on a short reference period of 1 week or 1 day, and it is believed that this would have also contributed to raising the number enumerated as employed. A number of surveys particularly those conducted by the Central Bank of Sri Lanka has used a reference period of one month and in some instances 2 months. The work of 1 day in the reference month being used for classification as employed, thereby raising the employment estimates and under estimating unemployment. "When one studies the concepts and definitions used in these censuses and surveys they lead to the impression that the conceptual framework adopted would have invariably resulted in the under estimation of the unemployed population." Korale (1988). The under estimation of the employed population has also occurred, but the degree of under estimation would be to a lesser extent than in the case of the unemployed, this comment does not apply to the Census 1981 where significant under enumeration of employed population has occurred.

The use of a reference period considered longer than that required for the currently active method and the concept of seeking work being considered as a necessary factor for enumeration of unemployed contributed to a reduction in the number enumerated as unemployed through the exclusion of discouraged workers and persons who were not actively seeking work, who had been included in the Census 1971. These changes in the concepts and definitions have resulted in the loss of comparability of data between 1971 and 1981, apart from the problem of quality of data arising from the issues discussed earlier.

Choice of a Method.

The primary consideration as regards measurement of economic activity is to achieve comprehensiveness of coverage of the different components of the economically active population.

The use of the concept currently active population would obtain a snap shot picture of the form and pattern of labour utilization immediately preceeding the census enumeration day. From preceeding experience it is evident that the use of such an approach could not completely enumerate the population as employed, unemployed and economically inactive groups. The existence of significantly large proportions of seasonal workers, discouraged workers, educated females whose labour participation depends on the status of the labour market itself and other marginal groups would not enable a census utilising a pure current status approach methodology to provide a complete picture.

As regards the total magnitude of labour force variables it is possible to obtain what is derived through a census using a current status approach through a large survey. For these reasons there is a stronger case, for the use of the usually active population approach in the measurement of the economically active population at the next census. The use of this

approach has the weakness that it would not make the data comparable with the data from the Census 1981, but as stated earlier the employment data from this census were subject to limitations. The methodology used in 1971 has approximated to that of the usual status approach, and the definitional framework used was wide enough to capture the different activity groups with a degree of comprehensiveness that has not been achieved hitherto. Therefore it is more prudent to formulate a methodology to ensure comparability as far as possible between Census 1971 and the one to be undertaken. The adoption of the usually active approach is justified on the basis of the current size of the economically active population in manufacturing and tertiary activities which still remain small. Further, as already discussed, absolute numbers and even proportions of the categories of seasonal workers casual workers, discouraged workers and other marginalised groups are large and they could be best enumerated only using long reference periods under the usual status approach.

Specially in the case of employment what is most important is to obtain a comprehensive and complete account as is feasible so that the necessary industry-occupation matrices, occupational-educational attainment, professional skills level matrices can be obtained disaggregated to the sub-national level required. Census data on occupation industry, education and age, sex distribution of the work force are used as baseline data in educational and skills development planning, manpower estimation, and for economic analyses. Thus the census should be able to obtain a complete or nearly complete list of all persons who practice a particular occupation with specific skills etc. to be meaningful for use as baseline data. This data is not normally available through other means. The value of the census data on the economically active population ultimately depends on the extent and degree to which data of good quality on these conditions could be provided.

We have already acquired some experience in using these methodologies at a fairly large survey, namely the Labour Force and Socio-Economic Survey conducted in 1985-86. Preliminary data from this has already been released and the final tabulations are being processed. The data for the two approaches are however still not available for comparison. The preliminary report used the currently active population approach which uses a reference period of one week. The survey of 25,000 households was conducted in 12 monthly sub rounds from March 1985 to April 1986.

Enumerating the Economically Active Population in the Census 1991

The Census 1981 schedule contained 7 questions to measure the economically active population. This was the highest number of questions used at any preceding census and the need therefore to restrict the number of questions to 7 or to try and reduce the number further is evident. The reduction in the number of questions would naturally enable the collection of other vitally important information.

Specially because of the issues encountered and the deficiencies which reduced the quality of data at earlier censuses this aspect is being examined. The careful framing of questions and the provision of clear and lucid notes and instructions would certainly help to reduce the ambiguities and misclassifications of different statuses. The desirability of reducing the length of the instructions while at the same time providing clear descriptions in summarized form is understood.

The Census 1991 will have to engage approximately 90,000 enumerators. This would certainly not provide for the recruitment of the type of interviewers with the experience and skills as those used in a sample survey. This factor is also important in the development of the methodology and instructions for enumerating the economically active population. On the positive side the persons available for recruitment as enumerators would

have a somewhat better educational background than those engaged in 1981, but from the point of view of training them to admit the schedule the costs involved would not make it possible to utilise significantly different procedures from those adopted in 1981.

As far as the coverage of topics are concerned the census would canvas the following basic items:

- (i) Economic activity of the population
- (ii) Occupation
- (iii) Industry
- (iv) Employment status
- (v) Unemployment

The employed would comprise:

- (1) Paid employees
- (2) Self employed
- (3) Seasonal workers
- (4) Unpaid family workers
- (5) Apprentices and trainees in receipt of an allowance

The ILO recommendation relating to persons "with a job but not at work" will be used to classify them as employed.

Improving the quality of the data ascertained through the questions on occupation can be achieved by eliciting information on tasks performed in the job or profession, combined with the skills required to perform the occupation. It is felt that the quality of information on industry status, can also be improved by asking information on the name and address and the type of work carried out by the employing organisation, and main activities or products of the enterprise or establishment. The format of the questions and the flow of questions are still under study and are in the early stages of formulation.

A tentative approach for this is set out here with the object of canvassing comments and observations:

Economically Active Population

Q.1 Type of Activity

During the preceeding 12 Calendar months as a whole, state whether (you were) mainly economically active (employed or unemployed) or economically inactive

1. Employed	Go	To
2. Unemployed	Go	To
3. Economically Inactive	Go	To

Q.2 Occupation

State clearly, principal occupation - Job, Profession, Trade or type of work.

1. Description of work or tasks performed
2. Skill required

Q.3 Nature of Industry/Business or Service

Type of work carried out by employing organization or in the place of work

- (1) Name & address of employing organisation/
place of work
- (2) The main activity/
products of the enterprise/
establishment

Q.4 Employment Status

Indicate employment status in his principal occupation stated in cage (Q.2) by marking appropriate number

1. Government employee
2. Public Corporation/Local Government/Semi-Government employee
3. Other employees
4. Employer
5. Own account worker
6. Unpaid family worker

<u>Q.5 Unemployment Status and Duration</u>		<u>Duration of Unemployment</u>
For unemployed persons state	(1)	0 - 10 weeks
whether the person has been	Available and	10 - 20 weeks
available for work and seeking	seeking work	20 - 30 weeks
work or whether available for	(2)	30 - 40 weeks
work and not actively seeking	Available but	40 - 50 weeks
work and duration of	not actively	
unemployment	seeking work	

In conclusion data from Censuses of Population on the economically active population have been subjected to several weaknesses. The lack of comprehensiveness, incompleteness in coverage, misclassification of different statuses, etc. These weaknesses in data quality do not often get detected in pilot surveys and is learned of too late after the enumeration. These lessons of experience from preceeding censuses are being taken count of in planning the Census 1991.

QUESTIONS ON ECONOMIC ACTIVITY - CENSUS 1946 + 1953

TYPE OF ACTIVITY

1946
NOT INCLUDED

1953
NOT INCLUDED

PRINCIPAL OCCUPATION

Principal occupation or means of subsistence of earner

Occupation - State principal occupation of self (if earner) or of person on whom dependent. If dependent add the following letters: (S) for student (H) for housewife (D) for other dependents. If unemployed or retired add (U) or (R). State specific occupations such as Typists, Accounts Clerk, Advocate, Tea Plucker, etc.

INDUSTRY

Industry or business in which engaged or employed

Industry - State the character of the industry or place where work is performed, or source from which income or wages are derived. For example: Motor Works, Bank, Tea Estate, Government Service

EMPLOYMENT STATUS

Whether employer, employee or working on own account

Status - State whether Employer (E), Paid Employee (P), Worker on Own Account (OA), Unpaid Family Worker (UFW). (Follow instructions carefully).

QUESTIONS ON ECONOMIC ACTIVITY - CENSUS 1963

TYPE OF ACTIVITY

The question of type of activity sought to classify persons aged 5 years and over into one of the following categories:

1. Employed persons: i.e. persons who are engaged in any kind of work for pay or for profit.

This category includes:

- (a) Persons who work for salary or wages
 - (b) Persons who practice a profession or do cultivation or business on their own
 - (c) Apprentices, if they are in receipt of an allowance
 - (d) Seasonal workers, e.g. cultivators who work during the cultivation season only; and
 - (e) Unpaid family workers
2. Persons engaged in household work in their own homes
 3. Full time students
 4. Persons who are not working, but are in receipt of an income
 5. Unemployed persons previously employed, but unemployed at the time of the census
 6. Unemployed persons not previously employed but seeking work for the first time
 7. All other persons

PRINCIPAL OCCUPATION OF KIND OF WORK

State specific occupation, e.g. Ayurvedic physician, Accounts clerk, Paddy cultivator, Carpenter, Tea plucker, etc.

KIND OF INDUSTRY, TRADE OR SERVICE

State the type of establishment or place where the person works or source of income e.g. Tea Estate, Marketing Department Store, Factory, Paddy, etc.

EMPLOYMENT STATUS

State whether Employer, Paid Employee, Self Employed, Unpaid Family Worker

QUESTIONS ON ECONOMIC ACTIVITY - CENSUS 1981

PRINCIPAL ACTIVITY DURING THE LAST 30 DAYS

- | | | |
|---------------------------|------------------|--------------------------|
| 1. Employed | 4. Own Housework | 1. Emp.-go to P19 |
| 2. Student | 5. Other | 2. Stud.- go to P22 |
| 3. Retired/Unable to Work | | 3. Ret/Un.Wk - go to P24 |
| | | 4. O.H.W. |
| | | 5. Other |

DID HE/SHE SEEK WORK DURING THE LAST 30 DAYS

1. Yes
2. No - go to P18

PERIOD OF SEARCH FOR WORK

Ignore fractions of a year. Enter, under
1 year as 0, 9 years or more as 2

Years - go to P24

REASON FOR NOT SEEKING WORK

- | | | |
|------------------------------|-----------|-------------|
| 1. Has private income | 1. Pr.In | } Go to P24 |
| 2. Does not want to work | 2. D.N.W. | |
| 3. Thinks work not available | 3. W.N.A. | |
| 4. Other | 4. Other | |

PRINCIPAL OCCUPATION OR KIND OF WORK

State clearly exact occupation or kind of
work done during the last 30 days. eg. Mechanical
Engineering, Paddy Cultivator, Tea Plucker,
Accounts Clerk, Vegetable Settler, etc.

KIND OF INDUSTRY, BUSINESS OR SERVICE

State clearly, the type of industry, nature of
business, or service carried on in the work
place, eg. Tea Estate, Grocery Shop, Health
Department, Paddy Cultivation

EMPLOYMENT STATUS

- | | | | |
|------------------------|-------------------------|---------------|------------|
| 1. Govt. Employee | 4. Employer | 1. Govt.Emp. | 4. Empr. |
| 2. Semi-Govt. Employee | 5. Own Account Worker | 2. Semi-Govt. | 5. O.A.W. |
| 3. Other Employee | 6. Unpaid Family Worker | 3. O.Emp. | 6. U.F.W.. |

16. PLANNING OF THE 1990 POPULATION AND
HOUSING CENSUS OF THAILAND

Planning of the 1990 Population and Housing Census of Thailand

by

Chintana Pejaranonda

Neramit Dhanasakdi

For Presentation at the Twelfth Population Census Conference

September 5-9, 1988

Organized Jointly by

East-West Population Institute and the State Statistical Bureau of China

The 1990 Population and Housing Census in Thailand

The first national census in Thailand was organized in 1909. The most recent population and housing census was taken in 1980. The 1980 Census was the eighth population and the second housing to be conducted.

Thailand's ninth Population Census and third Housing Census was planned to conduct in April 1990, with April 1 as Census reference date, under the responsibility of the National Statistical Office.

The NSO had planned the 1990 Population and Housing Census since late 1986. Two pre-tests to test the applicability of questionnaires, concepts and field instruction were taken in various provinces in 1987 and 1988. In 1989 a pilot census will be conducted in one province to experiment in every phase of census taking and to find ways and means to solve the problems expected to arise in the actual census.

Planning of the 1990 Population and Housing Census

Coverage of the census

The 1990 Population census will covered as in 1980 census, all persons residing in the country as of the census date, 1 April. To be covered will be all Thai nationals; civilian citizens of foreign countries having their usual residence in Thailand, or those who have resided in Thailand of at least three months; persons having their usual residence in Thailand, but who were at sea or temporarily abroad as of the census dates; and government officials, both military and civilian, including Thai diplomatic personnel, and their families located abroad.

Persons who will not be enumerated included nomadic groups, i.e. hill tribes and fishing groups having no fixed place of residence; foreign military and diplomatic personnel and their families located in Thailand; and citizens of foreign countries temporarily visiting or travelling in Thailand for less than three months and refugees or illegal immigrants located in camps provided by the government.

The basic unit of enumeration will be household. All of the households and persons residing in each household as usual residence were enumerated. All persons falling within the scope of the census were enumerated on a de jure basis with the exception of students, who were enumerated on a de facto basis.

Enumeration Procedure

As in the previous census, the information for each individual is to be collected and entered on the questionnaire by a census official designated to perform this operation in a specified area.

Census sample will be again used for the 1990 census in order to obtain richer information and to reduce the cost of census taking.

Enumeration will be divided into three types according to the administrative areas as following :-

- 1) Bangkok Metropolis : Two phases of interviewing will be taken

Phase I : All persons and households in each enumeration district will be listed first, using Listing Form I

Phase II : Step 1 After the first phase is completed for an enumeration district, Listing Form will be used as the frame for sample selection. Only private households will be selected (20% sample) in this step. The sample private households will be selected by crew leader. The sample design adopted will be systematic with random start.

Step 2 Private households will be interviewed using a Short Form but for those falling in the sample, Long Form will be used for interview.

For collective households, Long Form will be used only on detailed information on population items, but housing items will not be asked.

- 2) Other municipal areas : Two steps will be performed :

Step 1 : All persons and households in each enumeration district will be listed, using Listing Form I.

Step 2 : After the first step is completed for an enumeration district all persons and households will be enumerated on the Long Form on detailed informations of population and housing. However questions on housing will not be asked for collective households.

3) Non-municipal areas : Two phases will be performed :

Phase I : In each village or enumeration district, all persons and households will be listed, using Listing Form I.

Phase II : Step 1 : Completed Listing Form will be used as the frame for sample selection. In this step, only private households will be selected 20%. The sample selection will be performed by crew leader.

Step 2 : Private households will be interviewed using a Short Form, however, for selected sample household Long Form will be used.

For collective households, Long Form will be used only on detailed information on population items, but housing items will not be asked.

Budget

The total budget for the 1990 Population and Housing Census is estimated to be about 237 million baht (\$ us 9.3 million). Of this amount, 24 million baht will be allocated for the 1989 fiscal year (covering October 1988 through September 1989) for Pilot Census to be held in Nakhon Prathom in April 1989. The remainder of the budget will be allotted for costs of field work of the 1990 Population and Housing Census and for the tabulation and publication of reports.

Committee Structure

The implementation of the 1990 Population and Housing Census is technically the responsibility of the NSO, particularly under the Population Survey Division. However, since the population census is the most important statistical project, involving other divisions within the NSO and various government agencies, a committee was formed. Advisory Committee on the 1990 Population and Housing Census was established in 1987 by the Council of Ministers. The Advisory Committee was composed of prominent experts in various fields as well as representative from other government departments.

Crew Leaders and Enumerators

In the previous censuses of Thailand, crew leaders and enumerators were selected from local school teachers and appointed by the Secretary General of the NSO. The criteria for selection of the crew leaders and enumerators were as follows : Crew leaders had to be senior school teachers not over 40 years of age and thoroughly familiar with the locality they were assigned to work. Enumerators had to be between the age of 18 and 35 and familiar with the locality in which they were assigned to work. In some areas, however, the crew leaders and enumerators were also recruited from students with at least high school graduate.

Each district also had one or two technical supervisors who were responsible for technical aspect of field supervision, during the training phase, they also had to instruct crew leaders and enumerators on enumeration procedures.

However, in the next census, some criteria for selection of crew leaders, enumerators and field supervisors will be changed. Experience of field works gained from the last Business and Trade Census held by NSO in 1988, indicated that, temporary employments of those who completed the Bachelor Degree or Vocational Level performed a good quality of field works. But, since population and housing census might involve a large number of employments, therefore this proposed plan is being studied. About 35,500 of crew leaders and enumerators and 5,900 supervisors will be recruited to work in the census period.

Census Map

As for the census map, NSO is now preparing a completely new set of base maps for the 1990 census. Maps will be made for Bangkok Metropolis, all municipal areas and large villages. There is two types of maps, enumeration district map showed its boundaries, street, lanes, canal and other land marks, location of houses and block boundaries within an ED. Size of enumeration district will be approximately 200-300 houses. A large enumeration district in municipal area may consist of one or more blocks.

Master maps will showed the location of all enumeration districts in each municipal area or each district in Bangkok Metropolis, Sanitary District and a large village which consists of more than one ED. Since maps is the most important for field works that show the boundary of ED. Therefore base map must be most up-dated. An actual work for up-dating the census map will be started from 1988.

Tabulation and publication plan

The tabulation and publication of census data will be divided into three stages:

- 1) A preliminary report (by hand tabulation) of population by sex and households by province, region and the whole kingdom. This report will be published within 6 months after the end of enumeration period. Population growth during the inter-censal period, 1980-1990, will also be published for each province.

2) Advance Report of basic characteristics on a specific percent sample by region and whole kingdom.

3) Reports of population and housing data by province, region and whole kingdom. These series should be completed within 2 years after completion of field work.

Special tabulation : In 1990, special tabulation are also planned for a series of monograph reports in such subjects as migration, fertility, etc. A sample size is not yet designed, but will be large enough to represent municipal and non-municipal areas and regions.

Census evaluation

Census of Thailand is followed by a post enumeration survey since 1970 in order to evaluate the completeness and the quality of census.

The plan for the 1990 PES is not yet concretely formalated. But it is considering the same method as was done in 1980.

Beside the PES, an evaluation by using demographic technique is also planned for investigating the coverage of the census.

Topics to be covered in the 1990 Census

In the 1990 Population and Housing Census of Thailand, the following topics are planned to be included in the census schedule :-

A. Population Topics

1. Type of households (Private or Collective)
2. Name
3. Relationship to head of household
4. Sex
5. Month, year of birth, age
6. Highest grade completed
7. Marital Status
8. Occupation of last week
9. Reason of not work
10. Religion
11. Place of birth
12. School attendance
13. Literacy
14. Duration of residence in this village or municipal areas
15. Place of previous residence if the individual migrated into his present residence less than 5 years.
16. Whether the individual's previous residence was in municipal or non-municipal area.
17. Reason of migrants
18. No. of children born alive in this household
19. No. of children born alive who live elsewhere
20. No. of children born alive who died
21. Contraceptive practice
22. Method used
23. Occupation of last year
24. Industry of last year
25. Work status of last year
26. Physical disability and type of disability.

Housing topics to be asked for private household only

1. Type of structure of the private house (single unit structure, multi-unit structure)
2. Type of construction materials
3. Is house used for commercial purpose too?
4. Tenure of household living quarter.
5. Monthly rent.
6. Advanced payment
7. Ownership of land (for owner)
8. Ownership of rented dwelling (for renter)
9. Number of rooms regularly used for sleeping
10. Water supply system
11. Type of lighting
12. Toilet facilities
13. Place used for cooking
14. Type of fuel used for cooking
15. Consumer durable goods : radio, television, electric fan, bicycle, motor cycle etc.

The 1989 Pilot Census of population and housing

Two pretests of the Population and Housing Census of Thailand have been conducted in 1987 and 1988 in various heterogeneous areas selected on a purposive sample basis. The objectives of these pretests are as follows :

1. To test the applicability and appropriateness of listing form, the enumeration form, instructions for enumerators, crew leaders,
2. To test the appropriateness of concepts, wording and format of census forms.
3. To study various problems arising from field operations and find appropriate solutions.

After the pretest, a Pilot Census of Population is planned to be taken in April-May 1989, in the province of Nakhon Prathom, an "average" province with 616,000 inhabitants, some 50 kilometers from Bangkok. However, it is planned that only listing will be done for the whole province, but enumeration will be done only for some selected Amphoe (district).

The primary objective of the Pilot Census was (I) to test the field organizational structure and effectiveness of control, (II) to test the effect and applicability of sampling selection, (III) to test enumeration procedures and some of the programs and other computer techniques to be used in census tabulation and (IV) to test the quality of temporary employment who earns the Bachelor's Degree and Vocational Level as supervisors and enumerators for the census field works.

Census Calendar

1990 Population and Housing Census of Thailand

Steps	Year						
	1987	1988	1989	1990	1991	1992	1993
Preparation and Planning of the Census							
1.1 Coverage and Procedure of the census							
1.2 Payment and Budget							
1.3 Preparation of questionnaire and instructions for enumerators							
1.4 Planning for Data processing (Instruction for manual and machine editing, data entry and tables to be tabulated)							
1.5 Developing Program and tabulation							
1.6 Preparation of maps							
1.7 Publicity							
1.8 Conducting Pilot Census							
1.9 Planning post-enumeration survey (conducting PES and evaluating results)							
Field Enumeration							
2.1 Recruitment of field personnel							
2.2 Training field personnel							
2.3 Census enumeration							
2.4 Conducting post-enumeration survey							
Data Processing							
3.1 Manual edit, coding and verify							
3.2 Data Entry							
3.3 Machine edit							
3.4 Analysis data							
3.5 Evaluating and analysis results of the post-enumeration survey							

Census Calendar

1990 Population and Housing Census of Thailand

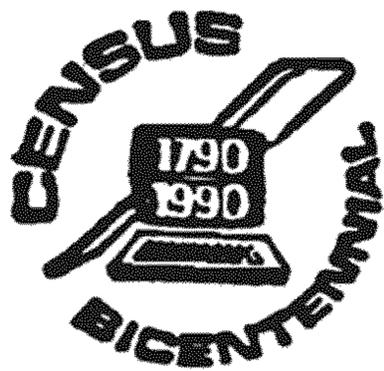
Steps	Year						
	1987	1988	1989	1990	1991	1992	1993
Presentation of Census Results 4.1 Preliminary report 4.2 Advance report 4.3 Final report 4.4 Analytical report				[] []	[]		

17. STATUS REPORT ON THE 1990 CENSUS OF
POPULATION AND HOUSING
OF THE UNITED STATES

STATUS REPORT ON THE 1990 CENSUS
OF POPULATION AND HOUSING OF THE UNITED STATES

PETER A. BOUNPANE
ASSISTANT DIRECTOR, DECENNIAL CENSUS
BUREAU OF THE CENSUS
WASHINGTON, D.C. 20233

TWELFTH POPULATION CENSUS CONFERENCE



BEIJING, CHINA
Sept. 5-9, 1988

INTRODUCTION

The 1990 census will be the 21st decennial census of the United States. It also will be the United States' Bicentennial census. The first census was conducted in 1790 under the direction of the Secretary of State, Thomas Jefferson. George Washington was President. At the time of the first census, the city that bears his name and serves as the capital of the United States had not yet been planned.

So the census is practically as old as the United States itself. Article 1, section 2 of the U.S. Constitution requires a decennial census to provide population counts as a basis for apportioning seats in the House of Representatives. Over time, census data have come to be used for many other important purposes. These include drawing legislative districts, allocating government funds, and planning at all levels of government and in the private sector.

Census Day for the 1990 census is April 1, 1990--less than 2 years from now. Major plans are complete and we have begun preparatory operations. Our current cost estimate for the 1990 census is \$2.6 billion over 10 years.

This paper will discuss five topics: (1) the status of planning for the census, (2) the preparatory operations that have begun or will begin soon, (3) plans for increased automation in the census, (4) plans for promoting the census and building public support, and (5) some critical issues for the 1990 census.

PLANNING

We are now completing our Dress Rehearsal censuses, which are the final dry runs of planned census operations for 1990. We began formal planning for the 1990 census in October 1983. We have conducted more major tests than previously. The Dress Rehearsal is the capstone of our planning efforts and was preceded by 5 years of consultations with data users and formal tests of alternative procedures and questionnaire content.

We consulted with a wide range of data users, including minorities, planners and academics, business leaders, representatives of private organizations, state and local officials, and Federal agencies. We have also kept members of Congress informed of 1990 census plans through more than 30 hearings since 1984. We conducted seven major test censuses and a number of smaller tests. Based on these consultative efforts and tests, we designed a 1990 census plan that we implemented in the Dress Rehearsal.

We conducted the Dress Rehearsal censuses in three sites representative of the kinds of enumeration conditions in which we will conduct the 1990 census.

St. Louis City, Missouri, is a large urban area with a high concentration of minorities, low-income persons, multi-unit housing, and so on. Fourteen counties in East Central Missouri contain one mid-size city (Columbia), a number of smaller cities and towns, rural areas, colleges, a major military base, and resort areas. Eight counties in eastern Washington contain some small cities and towns, but also remote, sparsely populated areas and two American Indian reservations.

Census Day for the Dress Rehearsal was March 20, 1988, precisely 2 weeks before Easter as will be April 1 in 1990. Most data collection and follow up work has been completed, but data processing operations continue.

The Dress Rehearsal has proceeded fairly smoothly. There have been no major problems to date that would cause us to change our plans for 1990. We have, as expected, identified various refinements to operations and systems.

PREPARATORY OPERATIONS

Major preparatory operations for the 1990 census have already begun and will accelerate this year and next.

Addresses

One of the major preparatory activities is the compilation of an address list that we will use to control the delivery and return of questionnaires in 1990. We will enumerate about 95 percent of the housing units by mail-back procedures. With these procedures, we compile an address list in advance of the census, mail or deliver questionnaires to each address, and follow up on those addresses for which questionnaires are not returned. This way, we can complete most of the census by mail and concentrate our resources on the hard-to-enumerate follow-up cases. The success of these procedures depends in large part on an accurate address list.

For the more urban areas of the country, we purchased about 56 million addresses from commercial vendors under contracts awarded in February 1988. We have assigned geographic codes to these addresses by computer. We are now listing addresses for the more suburban and rural parts of the country, where lists either cannot be purchased or cannot be assigned geographic codes by computer. This operation, which we call "prelist," will be conducted in two phases. The first phase, to be completed by the end of the year, will list about 32 million addresses. The second phase, which we will conduct in the fall of 1989, will list another 11 million addresses.

Our experience shows that updates by census enumerators and postal workers are required to produce a complete address list for the census. Thus, we contract with the U.S. Postal Service to conduct several checks of the lists for completeness and accuracy and, in addition, have our staff conduct checks. One of these checks is the "precanvass" operation, an updating of the addresses in urban areas, which we will conduct in 1989.

We expect that when we are through with all of our address-compilation operations that the list will contain an address for almost every housing unit in the areas where we will conduct the census by mail.

For the balance of the country (remote, sparsely-populated areas containing 5 percent of the population and 40 percent of the land area), mail carriers will deliver the questionnaires to housing units. But we will ask householders to complete their questionnaires and hold them until an enumerator visits to pick them up, rather than mail them back.

Geography

A second major preparatory activity, which has been ongoing since 1984, is the creation of an automated geographic support system. We call this system TIGER, which stands for Topologically Integrated Geographic Encoding and Referencing.

Geographic materials are essential to a successful census for two reasons: First, having correct and legible maps helps our enumerators find every housing unit so that we have a complete count; and second, having correct boundaries and geographic information helps us assign each housing unit and the people who live there to the appropriate land area.

Three major tools provide the geographic support for the census: maps, address reference files, and geographic reference files. Maps guide data collection, depict the geographic areas for which we tabulate data, and, in some cases, present data. An address reference file associates addresses with their geographic location. For example, it references 801 Main Street to block 101 because it knows that all the odd numbered addresses in the street segment 801-899 Main Street are in that block. Even-numbered addresses 800-898 on Main Street would be assigned to another block. This task of assigning geographic codes is essential when taking a census by mail. Geographic reference files catalog the various geographic areas for which data are tabulated in the census and define their relationships to one another so as to allow the presentation of data by geographic area in an orderly, understandable fashion.

For example, the assignment of numbers to blocks allows them to be identified when grouped into the next higher level of geography, the block group, which also is assigned a number. Block groups also can be grouped to form census tracts, and so on. Thereby, a specific block can be related to all higher-level geographic units, up to the state level, such as counties or voting districts, for which data is tabulated.

For the 1980 census, we prepared these three major geographic tools in independent operations involving a great deal of clerical work. This process was slow and error-prone, leading to delays in production and errors and inconsistencies in some of the products. The creation of an automated geographic support system--TIGER--is a major step forward for the 1990 census.

The key word in the TIGER acronym is "integrated." TIGER will integrate into one file all the geographic information that was produced in separate operations in 1980. This will allow us to produce the geographic products and services for 1990 from one consistent data base, and will help us avoid some of the delays and inaccuracies we experienced in 1980. Having computer-driven plotting devices generate high-quality maps that match the geographic areas in our tabulations will be a big improvement over the clerical operations of the 1980 and earlier censuses.

TIGER has a large data base with a computer-readable or "digital" map as its foundation. As a start to creating the TIGER file, we and the U.S. Geological Survey entered into an interagency cooperative agreement for the creation of the initial cartographic data base. Then, we updated this initial digital map base with information about new streets and current boundaries that was gathered from state, county, and local officials to make the maps useful for Census Bureau activities. While we had some delays earlier, the computer-driven map plotters we purchased earlier this year are producing the maps our enumerators are using now in the "prelist" operation.

Questionnaires

A third major preparatory operation is the printing of census questionnaires. As in recent censuses, we will have two primary questionnaires--a short form and a long form. The short form will contain the basic population and housing subjects we ask of all persons and housing units. The long form will contain these same basic subjects plus additional population and housing items asked of only a sample of persons and housing units. We refer to the basic subjects that appear on both the short and long forms as the 100-percent subjects. We refer to those that appear only on the long form as the sample subjects.

As required by U.S. law, we reported to the Congress on March 31, 1988, on the proposed content of the questionnaires for the 1990 census.

The 1990 short form will contain the following population subjects: name, household relationship, sex, age, marital status, race, and Hispanic origin. And the following housing subjects: the number of units in the building; whether the housing unit is owned or rented (tenure); number of rooms in the housing unit; a screening question to determine if the value of the unit is affected by its being on 10 or more acres or containing a business or office; value of the unit, if owned; and, if rented, amount of rent paid and whether meals are included in the rent. The short form also will contain coverage questions designed to determine if the respondent listed all appropriate persons on the questionnaire. In addition, it will include several questions about vacant housing units that will be answered primarily through observation by our enumerators.

The long form will contain the previous items plus detailed social, economic, and housing inquiries. (See attachment A.)

Questionnaire printing is one of our most important activities for the census. We cannot conduct the census without questionnaires. We will require approximately 250,000,000 short and long forms, including Spanish-language and other special-purpose questionnaires. Quantity is not our only concern. The printing must meet quality standards so that the questionnaires, when filmed, can be read by our optical scanning devices. About 100,000,000 questionnaires will be assembled, along with instruction guides and return envelopes, into address-labeled mailing packages.

Questionnaire printing must begin in January 1989. Delays would push back other related preparatory activities and could mean that we would not be able to deliver census questionnaires to housing units as scheduled. Obviously, that would seriously jeopardize our ability to meet the legal mandate of delivering state population totals to the President by December 31, 1990.

Space Acquisition

A fourth major preparatory activity is the acquisition of space for temporary offices to manage census data collection and processing. In all we will require about 6 million square feet of space for 13 regional census centers, for 449 district offices that the regional census centers will manage, and for 7 processing offices. The regional census centers and district offices are responsible primarily for collecting the data. In the processing offices, the census questionnaires will ultimately be converted to computer-readable format so that we can produce data products.

AUTOMATION

The 1990 census will be far more automated than the 1980 census or any previous decennial census. A principal aim of planning for 1990 has been to automate many of the time-consuming and labor-intensive clerical tasks of the 1980 and previous censuses; and to begin converting data on the questionnaires into computer-readable format several months earlier than we did for 1980.

Traditionally, census data collection and much of the census data processing have been paper- and people-intensive tasks. The use of automated equipment can help us deal with the mountains of paper and the thousands of clerical tasks in a much more accurate and controlled way.

The TIGER System, which I described earlier, is an important element in our automation plans. It is doing all the things we expected of it for the 1990 census.

We also are automating our address list so that we can key updates to it on our computers concurrent with the data collection operation. With an automated address control file, we can print unique bar-codes on the questionnaires and keep track of them using electronic equipment.

In this way, we can readily ascertain the enumeration status of a housing unit and print out a list of addresses for which no questionnaire has been returned. We will follow up on these housing units.

Since the automated address control file allows us to keep track of individual questionnaires, we can process the questionnaires on a flow basis. This means that questionnaire processing will start concurrently with data collection, at least 4 months ahead of the 1980 schedule. In 1980, the conversion of data to machine-readable form did not begin until the district offices completed all enumeration, edit, and followup activities and shipped all their questionnaires at one time to a processing center. The earlier start for 1990 census processing should make it possible to release data products sooner.

We are automating several other important census operations. For some areas of the country, we will use computer review to replace the clerical check of questionnaires for accuracy and completeness. We are automating the operation that assigns computer-readable codes to written answers on the questionnaire. We are developing an automated management information system that will help us meet critical deadlines in planning the census and monitor the cost and progress of census operations. And, we will have computers in the local temporary offices to report cost and progress, to keep track of payroll, to organize job applicant information, and to control the field follow-up work.

To conduct an automated census under the scenario I have just described requires a good deal of computer equipment. There are dramatic changes involved in moving from a census with large-scale clerical operations to one using significantly increased automation. We have had to balance the need for adequate time to plan and test these new departures with the need for adequate procurement lead times.

Our procurement activities are now progressing as planned and we foresee no major problems. In June 1987, we completed award of a contract for a "family of minicomputers." The first computers have been delivered to support our census operations, including the 1987 Economic and Agriculture censuses, the automated geographic support system, and the 1988 decennial census Dress Rehearsal and prelist. In all, we expect to purchase over 500 minicomputers under this contract.

PROMOTION

Public cooperation is the cornerstone of a successful census. Without public cooperation, we could not have a complete count and be able to produce accurate data in a timely manner.

A creative and ambitious promotion campaign is essential if we are to maintain, and hopefully increase, public cooperation. One of the important goals of our promotion campaign will be to achieve a high mail-return rate.

The other important goal is to encourage everyone to include themselves in the census so that it will be as complete and accurate as possible. Our promotion campaign will be designed to emphasize the importance of being in the census and to ease concerns about census confidentiality.

For the 1990 census, we plan an extensive, multi-faceted promotion campaign.

Advertising

As for each census beginning with 1950, we have again signed a contract with the Ad Council to pursue a public service advertising campaign. The Ad Council regularly handles major mass-media advertising campaigns for U.S. government and other nonprofit programs. Using the Ad Council and its chosen advertising firm (Ogilvy and Mather, Inc.) led to a very successful promotion campaign in 1980. The Ad Council has again chosen Ogilvy and Mather, Inc. to develop the national promotional messages for television, radio, newspaper advertisements, and such "out-of-home" media as billboards and transit posters. For the first time, we have also asked the Ad Council to choose minority advertising firms to tailor the national message or develop new messages to reach minority populations. The Ad Council has chosen the Mingo Group, which has experience in developing messages for Blacks, and Castor Spanish International, which has experience developing messages for Hispanics. The Ad Council is also seeking to recruit firms that have experience in marketing to Asians, Pacific Islanders, American Indians, and Alaska Natives.

Organizations

We will work through key national and community organizations to get the active support of their members.

New for 1990, we are asking major national organizations to adopt a pledge to actively support the census. We work with community organizations through one-on-one contacts between our community awareness specialists and the leaders and membership of the community groups. We are increasing the number of census awareness specialists and hiring staff earlier than for the 1980 census.

Governments

Outreach to other Federal agencies, the Congress, state governments, and local governments will be a key element of our promotional program. Under the Mayors Cooperation Program, we are holding a series of one-on-one meetings with the mayors of some 350 key U.S. cities to discuss mutual concerns about the 1990 census and possible joint efforts for encouraging a city's population to be counted in 1990. To assure ongoing contact between the cities and the Census Bureau staff, we will ask the mayors to appoint high-level liaisons. We are committed to working with the cities to prevent foreseeable problems in the 1990 census and to be able to solve any unforeseen ones as they arise.

We will ask the highest elected official in every jurisdiction in the country to set up a committee of local leaders or use existing city means to generate local publicity about the census. We believe that this local publicity is an essential complement to the other programs and projects we are initiating because different themes or activities will be effective in different areas.

Private Sector

The private sector can also play a key role in promoting the 1990 census. We will seek the involvement of businesses and nonprofit organizations in publicizing the census to their employees or on their products. In some cases, we will seek funding from businesses to carry out selected promotional projects.

Religious Organizations and Schools

Religious organizations will be asked to encourage their members to participate in the census or, in some cases, to provide assistance in filling out the questionnaire. We plan to involve all denominations and national and local religious organizations in this project. We have already held important meetings with the Bishops Committee on Hispanic Affairs of the U.S. Conference of Catholic Bishops, with leaders of the National Council of Black Churches, other Black denominations, and Black fraternal orders.

We are developing a school project that will provide each elementary and secondary school in the country with a set of reproducible lesson plans for classroom instruction. These lesson plans are geared toward making students aware of the history and importance of the census and how to use census data. The ultimate goal is to have students take their knowledge about the census home with them. In some cases, older students may be the only ones in a household with the language skills to complete a questionnaire.

MAJOR ISSUES

Because the census is so important and the stakes--in terms of allocating political power and money--are so high, there is great interest in our plans. Some of our decisions for the 1990 census have been controversial and, in some cases, have led to law suits or bills introduced in Congress to change the plans. Thus far, no census plans have been changed in the courts or in Congress. Three areas of controversy are adjustment of census counts, enumeration rules about who to count, and the content of the census questionnaires.

Adjustment of Census Counts

One bill introduced in Congress would require the Census Bureau to adjust the census counts to correct for a potential undercount. Beginning with the 1950 census, we have included coverage measurement evaluation in each census. The results of these evaluations have shown a steady decline

in overall census undercount from about 4 percent in 1950 to about 1.4 percent in 1980. But there persists a difference in the undercount rates between Whites, on the one hand, and Blacks and Hispanics on the other. It is because of these different undercount rates that some people favor using the results of the coverage measurement evaluations to adjust the census counts.

In October 1987, the Commerce Department (parent cabinet department of the Census Bureau) announced that we would concentrate on producing the most accurate enumeration possible and that we would not adjust the 1990 census counts.

The reasons for this decision are too complex to discuss in detail here, but these excerpts from a letter from the Commerce Department general counsel to Congress will shed some light:

"Adjusting the population counts from the census may create more problems than it solves. No matter how careful and unbiased is the adjustment process, it will raise suspicions in the public mind about the reliability and integrity of the census and of the Federal statistical system.

In addition, adjustment may cause confusion about which are the "right" numbers. Adjusting the census will produce at least two sets of numbers, the enumeration and the adjusted counts. Statistical techniques are available to adjust, but there are questions about the validity of their results. There is no single system generally accepted by the professional statistical community. Different statisticians would employ different models and get different results."

Who to Count

We base our decisions about who should be counted on the concept of usual residence. Usual residence, a principle that was established by the act for the first census in 1790, is defined as where a person lives and sleeps most of the time. Usual residence is not necessarily the same as legal residence, voting residence, or the place where a person is found on Census Day. If the usual residence at the time of the census is in the United States, we count the person. That includes noncitizens, whether they are in the United States legally or illegally. If the legal residence is outside the United States, we do not count the person. That includes U.S. citizens living abroad, whether U.S. military, civilian employees of the government, or private citizens.

A lawsuit has been filed and bills have been introduced in Congress that would change our rules about who to count. They would have us exclude illegal aliens from the population counts used to apportion seats in the House of Representatives (one of the two bodies of the Congress) and include in the apportionment counts all Federal employees (civilian and military) stationed overseas and their dependents living overseas.

The Commerce Department and Census Bureau oppose these efforts because they would be inconsistent with the concept of usual residence. In addition, there are operational problems in determining legal status of those residents in the U.S. or in determining the proper U.S. address of U.S. citizens overseas

at the time of the census. Furthermore, the Department of Justice (another cabinet department) has advised us that it would be unconstitutional to exclude aliens who claim the United States as their residence. The Justice Department has also advised us that it has doubts about the constitutionality of any legislation that would seek to include for apportionment purposes military personnel, Federal civilian employees, or their dependents, when such individuals live outside the United States.

Questionnaire Content

The plans for census questionnaire content have also been controversial and there are bills in Congress that would change or add some questionnaire items.

Part of the controversy stems from the review of the census questionnaires (short form and long form) by the Office of Management and Budget (OMB). The OMB is an executive branch office that reports directly to the President. It has responsibility for, among other things, coordinating Government statistical activities and reviewing Government paperwork (such as census questionnaires) to reduce the public's reporting burden.

After several months of discussions, the Census Bureau and OMB reached an agreement that represented an appropriate balance between the Nation's need for information and the reporting burden the census imposes on individual respondents. The OMB suggested changes to the Census Bureau's proposed questionnaires that involved moving some questions from the short form to the long form and deleting some questions altogether.

A bill in Congress would move two questions--on plumbing facilities and condominiums--back to the short form and restore one question that was deleted (that on heating equipment).

Another bill before Congress concerns the race question, specifically our proposals for identifying Asians and Pacific Islanders. For 1990, we plan to list a category "Asian and Pacific Islander" and ask respondents to write in their specific group. The written entries would be coded for computer processing. The proposed legislation would have us revert to the 1980 census procedure of listing 9-11 specific groups and provide for write-ins for the balance. We oppose this legislation because the results of tests we conducted since 1980 show that the new version we plan for 1990 will provide more accurate and consistent reporting for racial groups, particularly for the Asian and Pacific Islander population.

CLOSING

Despite these controversies, there remains widespread support for the census. In fact, the controversies stem from the recognition that the census is a vitally important undertaking. Not all of the controversy is new. For example, adjustment of census counts and inclusion of illegal aliens were topics of debate in 1980, too.

Meanwhile, census operations are ongoing. We have taken a number of steps to improve the efficiency and accuracy of the 1990 census. We are concentrating our energy, attention, and resources on efforts to make the census better. We expect these steps will continue the trend of reducing the undercount.

Attachment A

Long Form, 1990

POPULATION

Social characteristics:

Education--enrollment and attainment (including degree)
Place of birth, citizenship, year of entry
Ancestry/ethnic origin
Language spoken at home/proficiency
Migration
Disability/personal care
Fertility
Veteran status

Economic characteristics:

Employment and unemployment
Occupation, industry, and class of worker
Place of work, commuting to work, travel time
Work experience and income in 1989

HOUSING

Heating fuel
Source of water and method of sewage disposal
Autos, light trucks, and vans
Kitchen facilities
Year structure built
Year moved into residence
Number of bedrooms
Farm residence
Shelter costs, including utilities
Condominium status
Plumbing
Telephone

USE OF SAMPLING
IN CONJUNCTION WITH THE
CENSUS OF POPULATION

by

Vijay Verma

For presentation at the Twelfth Population Census Conference, Beijing,
China, September 5-9, 1988.

18. USE OF SAMPLING IN CONJUNCTION WITH
THE CENSUS OF POPULATION

INTRODUCTION

1. In most countries today the decennial census of population constitutes the primary source of geographically detailed information on basic demographic and related characteristics of the population and its various components. It is generally accepted that in a population census it is "not necessary to gather all demographic and housing information on a 100 percent basis; a good deal of time and money can be saved through the use of sampling. Furthermore, in certain circumstances, only the sampling method yields data of acceptable accuracy".¹ As elsewhere, the pattern in Asia and the Pacific is increasingly to (a) collect information on size, age and sex composition, geographic distribution and certain other basic demographic and socio-economic characteristics of the population on the basis of a 100 percent or complete enumeration; and (b) to supplement this by collecting, as a part of the census operation, information on a larger range of variables - such as migration and employment, fertility and health and so on - on a sample basis. For example, among the 18 countries of the region for which information is at hand concerning the latest (1980s) census round, 11 (or 60%) employed complete and sample enumeration in combination, while in the remaining seven only complete enumeration was used.² It may be expected that in the 1990s round, the use of sampling to supplement complete enumeration will be even more wide-spread.
2. The objective of this combination of the census with a sample is, of course, to exploit comparative advantages of each of

¹ United Nations. Handbook of Population and Housing Census Methods. Part VI: Sampling in connexion with Population and Housing Censuses. Studies in Methodes, series F, no 16, 1971.

² Cho and Hearn (ed.). Census of Asia and the Pacific: 1980 Round. East-West Population Institute, East-West Center, Honolulu, Hawaii, 1984.

the two schemes. These relative advantages tend to be complementary: one method is strong where the other is weak, and vice versa.

A complete census is generally indispensable for obtaining information for small domains and local areas, and also for obtaining politically important data which - for legal or other reasons - must be seen to be free from sampling variability. Furthermore, complete enumeration can often secure a better coverage of the population than that usually achieved with samples. These are among the fundamental reasons for using sampling in combination with, rather than as a substitute for, complete enumeration. However, enumeration of the entire population requires mobilization of financial and human resources on a large scale, and this cannot be sustained for a prolonged period nor repeated too frequently. The need to deploy a large fieldforce - with the concomitant difficulties in maintaining high quality of training and supervision - means that the type of information appropriately collected in a census, while extensive in coverage, must be simple and brief in content. Simplicity is also necessary to keep the volume of the data to be processed within manageable limits. Therefore, a complete census is typically confined to obtaining a detailed and complete picture of the number (size) and basic structural characteristics of the study population, and to providing as much detail as possible for small domains and especially for local areas. This applies to all types of censuses: whether of human population, economic establishments, or agricultural holdings.

By contrast, enquiries confined to samples of the population can, by virtue of their smaller size, be designed to obtain (and obtain more quickly and more frequently) a wider variety of more complex data. Such data are not gathered in a complete census: attempts to do so would result in very high costs, and what can be even more important, in low quality and serious delays. Relatively speaking, sample surveys can be much cheaper and much more timely and, furthermore, can be tailored to fit a variety of needs and methods of collection.

But generally they cannot yield reliable data for small domains and especially for local areas.

Ofcourse the contrasts sketched above between complete censuses and samples tends to be more striking for larger populations, but less so for smaller countries.

3. Combined data collection on a 100 percent and a sample basis during the census is just one - albeit the most important - aspects of the use of sampling in conjunction with the census. Complete census and sampling can be used in combination or in otherwise related ways in a variety of manners with the objective of capturing the relative advantages of each. These include the following.

- (1) Using sampling in the design and control of census operations, such as in planning, testing, controlling and evaluating the census.
- (2) Using sample enumeration to supplement the items covered in the complete census, as described above.
- (3) Sampling the census results for processing, with the objective of making the results available in a more timely and economical manner.
- (4) Extracting samples of micro-level files of detailed census data so as to facilitate dissemination of primary data more widely for bona fide research and analysis by other users.
- (5) Using the census as a basis for sample surveys through enhancing statistical capability and resources; more specifically, in terms of providing impetus, experience, infrastructure, facilities, base-line data and population controls for estimation and, above all, providing sampling frames and possibly also master samples for surveys in the post-censal period.
- (6) Using the census and post-censal sample survey data in providing post-censal and current estimates for local areas and small domains.
- (7) Finally, it should also be mentioned that there are circumstances in which it is unavoidable, or sometimes even

desirable, to replace complete enumeration entirely by one or a series of sample enumerations.

Below we consider some general sampling issues involved in these combined uses of samples and censuses.³ Some illustrations from recent practices of the countries in the Region are also given.

SAMPLING FOR DESIGN AND CONTROL OF CENSUS OPERATIONS

4. Pilot studies are required to test the adequacy of census questionnaires, instructions, training programs, enumeration procedures, field organization, etc. They serve as practical training for the nuclear staff and supervisors, and provide information on operational aspects (costs, time) of enumeration. For pilots, it is usually difficult to insist on good samples of the entire country: the common practice is to choose areas which are convenient but are also expected to yield a good test of questions and techniques in diverse circumstances. This approach assumes that the diverse circumstances, and especially areas of particular difficulty, are identifiable. Sometimes, but not usually, this approach is supplemented by a "micro-census" based on a large probability sample in advance of the actual census. In any case, to yield full benefits, tests should be carried out through all stages of collection, processing and examination of results. For a large undertaking such as the census, accurate operational information on conditions and requirements of enumeration is essential for proper planning and execution.
5. Statistical quality control techniques are sometimes used to assess and control the work of individual operators and opera-

³ Several of these issues are discussed further in Kish and Verma: Complete Censuses and Samples, Journal of Official Statistics, vol 2, no 4, 1986. I draw liberally on this publication, and acknowledge debt to my co-author.

tions. Various sampling plans have been developed with the object of reducing the cost of verifying the operatives' work, and ensuring at the same time a specified quality level for the outcome. However, "quality assurance" with process control is a common alternative. It is important to emphasize that specific procedures should be developed in each situation, suited to actual field conditions and procedures of supervision.

6. As distinct from quality and operational control measures, evaluation surveys are designed to check the average quality of the census and its major components. In designing samples to check and evaluate census work, a number of options need to be considered; for example:

- . the timing of the supplementary operation(s) in relation to the main census operations of listing and enumeration - these may be all separate or may be combined in various ways;
- . whether the test uses ordinary census enumerators or uses different, specially trained enumerators;
- . whether the (more elaborate) test procedures are additional to or replace the ordinary census operations in the selected sample areas;
- . in the former case, whether the test is conducted independently of or on the basis of the results already obtained;
- . the size and design of the test sample;
- . the relationship between samples for the different operations attached to the census.

Two types of census check operations may be distinguished: checks that help assess coverage errors relating to duplication and omission of units, and checks on content errors relating to the accuracy of responses.

7. Checks of completeness of coverage usually follow the ordinary census as a separate and additional operation, over a sample of EAs (enumeration areas) but enumerating all households, individuals or other ultimate units in each selected EA. They may or may not use special enumerators for the same area. The standards of supervision in the test are usually more strict. Typically, check enumeration is carried out by referring to the first enumeration; i.e., enumerators are given previously prepared lists of units in the sample areas with the object of

finding missed units. However, sample studies using "dual coverage techniques" for estimating undercoverage are also possible where lists of households (or other units) are available from entirely independent sources. Checks of censuses based on analytical, demographic, and statistical methods and models and on data from independent sources, registers and samples are obviously also useful to check coverage.

8. Checks on content tend to be more varied in design. One may use a post-enumeration survey (PES) done after the census, or sample of high quality enumeration done simultaneously with the census. In the latter case, a sample of EAs is covered (possibly after subsampling within EAs) with better methods, better enumerators, longer questionnaires, etc., than those used elsewhere. With this arrangement the additional expense and respondent burden is less than that with double coverage of sample areas; however, only net differences at the aggregate level and with higher sampling variance are measured.
9. Whatever the method, it is generally preferable to use special and improved procedures to identify the nature and source of individual coverage and content errors so as to assess gross errors, including bias, and identify measures necessary to control these errors and minimize their effects. Furthermore, it is worth emphasizing that the objective(s) of the test should be clearly specified and be sufficiently modest to be attainable within given resources. Quality checks have frequently turned out to be of insufficient quality to yield any useful information at all, in spite of the relatively high costs involved.

It is often difficult and expensive to use good samples of the entire country for quality checks. Inferences may have to be made from restricted areas of the population in the expectation that the results can be generalized. This requires carefully controlled sample selection; also, results will be convincing only if observed differences between different areas in the sample, or at least within major domains, are not

too great. In any case, the samples will need to be larger and more objective than those typically required for the pre-census pilot surveys. Furthermore, the sampling requirements are more stringent when the objective is to adjust (improve) the census results, rather than merely to identify the nature and sources of major errors.

The results of the census check operations should be used, to the extent possible, in conjunction with alternative information available from other sources such as past censuses and surveys and administrative data, as well as substantive and analytical relationships.

USING SAMPLE ENUMERATION TO SUPPLEMENT COMPLETE CENSUSES

10. Samples attached to the census differ from ordinary sample surveys in that their connection with censuses gives them special functions and special advantages. They are often substantially larger than ordinary surveys and share the basic census objectives of providing detailed data for relatively small domains. The large size is facilitated by the relative cost-effectiveness of these operations and the advantage they enjoy in availability of funds and resources as a result of their connection with the census. Sometimes, however, smaller and more concentrated samples may be used to cover more difficult topics for national statistics and large domains, often using special arrangements and enumerators, but keeping the basic link with the census. In designing samples attached to the census, a variety of interrelated issues need to be considered such as the following.

- (1) Whether it is necessary to add sample enumeration to the census, and if so, how the data to be collected should be divided between the complete and sample enumerations.
- (2) The appropriate sample size; allocation of the sample to different reporting domains.
- (3) Choice of the sample structure, in particular the type of units and sampling stages to be used.

(4) The relationship of sample enumeration to complete enumeration, in terms not only of sampling but also of field organisation and interviewing operations; and the relationship between different samples if more than are used to obtain different types of data.

11. Whether it is necessary to add sample enumeration to the census is of course the most fundamental question in the present context. With a given set of items to be collected, one may begin by asking two basic questions in relation to each item.

(1) Is it necessary to collect information on the item on a 100 per cent basis, in view of precision requirements and the detail (especially geographical) with which the results have to be tabulated? In this context it should be remembered that complete enumeration can also be subject to all sort of non sampling errors; and that certain items are useful and meaningful only if they are estimated for a reasonably large population base (examples are certain demographic rates and absolute numbers of events).

(2) Is it desirable and feasible to collect the information on 100 percent basis? One has to consider possible costs, delays, burden on census staff and on respondents and its effect on data quality, effect of the item on quality of other perhaps more essential items, data processing constraints, and so on.

One would consider an item for complete enumeration only if the answer to both the questions is positive. If either is negative, the item is a candidate for sample enumeration. This may not be the final solution either. It may be possible to reconsider whether or not to include such an item in the census at all. If the substantive content is fixed, one still has to consider whether the saving achieved by restricting a part of the interview to a sample is big enough to compensate for the inevitable cost and complexity of introducing two types of operations. Experience shows that the savings from enumerating less than the complete population are not at all

proportionate to the reduction in the number enumerated, due to the constant costs of planning, selecting, implementing and analysing the sample. A 10 percent sample may for example cost more than half as much as a complete census: though this difference may be much less striking in the case of a sample attached to the census than in this case of surveys conducted more independly. Clearly if the required sampling rates are large or if only a few items require sample enumeration, it may be more reasonable to cover them all in the complete census, rather than intruduce a separate sample enumeration. The actual balance between costs and benefits depends on numerous factors, as is shown by the great (presumably rational) variation in country practices.

12. With sample enumeration, the size and allocation of the sample also depends on specific data requirements, circumstances and available resources, both material and technical, and little of general validity can be recommended. A critical determining factor is the number, type and relative sizes of the domains for which separate estimates are to be produced. Clearly, if estimates are required for many small domains and areas, the attached sample would have to be large. Similarly, great variation in sizes or importance of domains would require disproportionate allocations (unequal sampling rates); at least, it has to be ensured that the sample does not fall below a certain minimum in any domain, irrespective of domain size. An elementary cautionary note should be inserted here: too often choices are discussed in terms of sampling rates; the relevant issue however is the actual sample size rather than sampling rates per se.

13. The sample may be spread to all EAs or a sample of EAs may be selected for complete enumeration within EAs; sampling may also involve two or more stages: selection of EAs and then of households. The choice of the design is influenced by a number of interrelated factors (similar to the ones noted in para 6 concerning evaluation surveys): size of the operation and

required degree of detailed breakdown of the results; nature (complexity) of supplementary information to be collected; nature of EAs; travel conditions; type of enumerators available for the census and required for the attached sample; how often the households can be visited and the related considerations of time, cost and respondent burden; whether the attached survey replaces or is additional to the ordinary census operations in the sample areas, etc. When several samples are used to get different data, there is a conflict concerning spreading the samples over many different units versus concentrating them all in the same units. Spreading the schedules avoids the concentration of respondent burden, but concentrating them reduces costs and yields more information on relations between sets of variables.

14. (a) For simpler items which can be combined with the basic census enumeration - during a single visit, using the ordinary census enumerators - the sample can be more easily spread over all census areas, though even here operational difficulties in an extensive and large scale operation like the census should be remembered. (b) In any case, the more complex and specialized the inquiry, and especially if specialized enumerators are involved, the more it becomes preferable to concentrate the inquiry to a sample of EAs. This can be done more readily when the objective is to produce results at national or major domain levels, rather than at the level of small domains or local areas. Selection of complete (compact) EAs has the advantage of simplicity and lower cost; it is particularly appropriate when specialized procedures and enumerators are used, or when the survey replaces the ordinary census operations in the sample areas. However, this increases the variances of the estimates, and this would be most serious for small geographic domains, though it may not be critical for estimates for major domains and for cross-classes well distributed over different areas. (c) Two-stage selection of households within EAs can bring effective compromises, and would be particularly suitable for samples which form the

basis of a continuing program of post-censal surveys. However, subsampling within EAs may introduce serious selection biases, unless it is operationally separated from prelisting and is carefully controlled. The use of appropriate ratio estimates can overcome a part of the increase in variance due to clustering of the sample, and may remove some of the effects of selection biases when EAs are subsampled.

15. It will be instructive here to briefly review some recent experience of countries of Asia and the Pacific Region in the context of the above remarks. As noted earlier, of the 18 countries reviewed in the East-West Centre publication for the 1980s round, seven (Australia, China, Japan, Fiji, Malaysia, Nepal and New Zealand) did not resort to sample enumeration. Among the remaining countries, which used sample enumeration to supplement complete count, there was a considerable variation in the sample size and designs used. The Indian census of 1981, for instance, used a very large sample: 20% in most major States of the Union and 100% in smaller States and Union Territories, giving a total sample size of the order of 25 million households. (This was the first use of sampling in the census, and judging from the large sample sizes used, India - of all places! - was perhaps a little hesitant in this introduction of sampling.) Apart from the above, sampling rates in the Indian census did not vary by domain. The sample was selected in a single stage: systematic sampling of EAs with no subsampling within EAs, the sample enumeration replacing the latter in the selected EAs. In other relatively large or medium sized countries among those reviewed (Indonesia, Republic of Korea, Pakistan, Philippines, Sri Lanka and Thailand), sample sizes were much smaller than India but quite similar to each other - mostly in the range 1-2 million households. (The overall sampling rates were in the range 10-20%, apart from Indonesia which used 5%.) The major exception to this pattern was Bangladesh, where a much smaller sample 100,000 (1%) was used. Smaller countries or territories (Hong Kong, Papua New Guinea, Singapore) also used sample sizes of

the order of 100,000-200,000, though the sampling rates were rather similar to those used by most larger countries (10-20%). Most countries adopted the simpler and more practical single-stage design, involving direct selection of EAs with no subsampling of households within EAs. The only exception was again Bangladesh, where a two stage design involving subsampling within EAs was used; this is of course not unrelated to the fact noted earlier that the total sample size (number of households) in this case was relatively small, and subsampling within EAs permitted greater geographical dispersion of the sample.

16. Despite great variability between countries, a few remarks may be made also in relation to the substantive items included in the complete and sample enumerations. For brevity we will confine this only to some commonly encountered population topics. (a) Clearly, place of residence, relationship to head of household, sex, age and marital status are covered virtually universally and on a 100 percent basis, irrespective of whether or not sample enumeration is also used. (b) Certain 'legal' items such as registration status (if included which is not common) are also covered on a 100 percent basis irrespective of the presence of sample enumeration. The same applies with some exceptions to variables defining social categories (such as ethnic group, citizenship, religion and language), as well as to the less frequently asked questions on cottage industry and agricultural activity. (c) Rather surprisingly, several countries included questions on physical disability on a 100 per basis, including several who had introduced sample enumeration for other items. Of course a number of countries covered this item on a sample basis. (d) Simple migration related questions such as place of birth, previous residence, duration of residence etc. are by and large covered on a sample basis if sample enumeration has been introduced, though these items are also covered on 100 percent basis in some countries who did not introduce sample enumeration. This applies even more clearly to economic character-

istics such as activity and employment status, occupation and industry which are covered practically in all countries. (India is an important exception to this in that these economic characteristics were covered on a 100% basis, despite sample enumeration for other items.) (e) Children born alive and, to a lesser extent, children currently alive are covered in most countries, invariably on a sample basis where sample enumeration has been introduced and on a 100 percent basis otherwise. (f) With one or two exceptions, age at marriage and births and survivorship of children during past 12 months are covered only on a sample basis, and hence mainly in countries who had introduced sample enumeration. It is plausible that the need to cover such demographic items on fertility, infant mortality and marriage was an important consideration in the decision by these countries to introduce sampling to supplement the complete census.

17. As regards sample allocation among different domains, most countries followed proportionate allocation (i.e. uniform sampling rates) in the main part of the sample. However, certain types of exceptions were common:
- (1) In several countries, the long form was used on a 100 percent basis in urban or special areas, for example in India (small States and Union Territories), Papua New Guinea (urban and non-traditional rural areas), Sri Lanka (urban areas), Thailand (municipal areas except Bangkok Metropolis), Republic of Korea (special EAs with institutional population).
 - (2) In some countries (Pakistan, Bangladesh), urban or very small domains were oversampled to ensure the required minimum sample size.
 - (3) Countries differed in the treatment of special EAs containing institutional population as concerns the more detailed enumeration using the long form. For example, Republic of Korea covered such areas on a 100 percent basis; Indonesia covered them with the same sampling rate as used elsewhere (5%), but here selecting directly a sin-

gle stage sample of household rather than of areas; and the Philippines and Singapore excluded such areas altogether from the more detailed sample enumeration.

Only two countries (Republic of Korea and Papua New Guinea) reported a more general variation in sampling rates among different subnational domains. In Papua New Guinea for example, the sample was allocated in proportion to square-root of the domain population, as a compromise between the requirement of national level estimation (suggesting proportional allocation) and the requirement to produce separate estimates for individual domains (basically suggesting equal allocations). Perhaps this or some similar compromise allocation will be used more commonly in future censuses because of the increasing emphasis on producing national as well as subnational estimates for many domains with considerable variations in domain sizes.

18. Few countries report separate arrangements for recruitment, training or fieldwork for the more complex sample enumeration, and none report using multiple samples or multiple versions of the long form to collect different types of information from different sets of households. Typically, the more detailed sample enumeration replaced the briefer census enumeration in the sample areas, but in some cases (Pakistan, Sri Lanka) the former appears to have been undertaken in addition to the latter in sample areas. An interesting case is provided by the 1980 census of Papua New Guinea, which was the country's first proper complete census. Here the requirement to conduct simultaneously two types of enumeration (using the short and long forms respectively) was judged to be too complex. The pilot test revealed various operational problems and indicated that field workers in rural areas could not manage the operation involving two types of questionnaires, two sets of training materials and two field plans; furthermore, people were missed in the border regions between sample and nonsample areas because the boundaries could not be clearly defined. Consequently, the procedure adopted was to (a) use only the long form in urban and non-traditional rural settlements on a

100 percent basis, replacing the short form (which was incorporated into it); (b) to use only the short form in traditional rural areas during the 1980 census itself; and (c) to conduct the sample-based long form operation in these areas during the subsequent year, as an undertaking separate from the main census. It may be mentioned that Pakistan too, in introducing sample enumeration in its 1972 census, did so on a relatively small scale (around 0.25 million households), and in the form of its Housing Economic Demographic Survey conducted separately a year after the main census. With more experience, the 1981 census of Pakistan combined a larger scale sample enumeration with the complete census in a more conventional manner.

SAMPLING FOR CENSUS PROCESSING AND DATA DISSEMINATION

19. Sampling from census data for processing may be resorted to in two situations:

- (1) for quick tabulation of broad census results; and
- (2) for production of detailed tables.

The first situation arises from the need to release some broad results quickly, even in a preliminary form. The objective of introducing sampling for detailed tabulation is to save resources.

For (2) at least, the sample for processing can usually be selected at household level from all the census EAs. (Sampling of individuals can be more cumbersome, and is often also unnecessary due to demographic heterogeneity within households.) However, confining the processing to a sample of EAs can still retain some operational convenience, especially for (1) which may need to be carried out well before the entire census data can be keyed in and edited.

20. For more complex analysis, as well as for release of micro-data for public use, it is usually desirable and possible to select samples from the entire census and avoid their cluster-

ing into basic units of enumeration. Random selection makes analysis simpler and facilitates prevention of identification of individuals in the sample.

For sampling of households, and of individuals if applicable, systematic selection from ordered lists is probably the most convenient and efficient technique. The arrangement or ordering has to be operationally convenient, as well as suitable for a large number of characteristics for which estimations are required. For areas and households, ordering by administrative and geographical location is usually the most appropriate.

CENSUS AS A BASIS FOR SAMPLE SURVEY DESIGN AND ESTIMATION

21. The census is a major operation which can provide great impetus to development of national statistical capability, particularly in the area of sample surveys. The census forms the basis of subsequent surveys in a number of ways: by providing the sampling frame; by providing auxiliary information for improved estimation especially estimation of population totals through regression and ratio estimates; and by mobilising resources for the development of infrastructural facilities for conducting subsequent sample surveys. In addition, large-scale surveys attached to the census can provide a convenient and efficient basis for launching continuing survey programs. Later surveys can be smaller in scale and more varied and complex in content, or they can be specially designed to monitor changes, as for example in multi-round demographic surveys. The larger base-line survey can also provide a 'master sample' for efficient and convenient subsampling and for estimation from subsequent, smaller survey (see below). It will be good if in future rounds of the population census, greater attention is paid to these varied rolls of the census in relation to post-censal sample surveys.

22. Good samples need and are based on census data providing the sampling frame, especially in countries where alternative sources such as population registers are not available. This applies in particular to sampling frames for household surveys of the general population covering a variety of demographic, social and economic topics. However, it should be remembered that the population census can be a primary source of sampling frames for surveys of other types as well, such as agricultural surveys of farming households and holdings, and economic surveys of establishments particularly in the small, informal sector. Special surveys in these areas can also be 'attached' to the census, even if as operations which are organisationally, logistically and temporally separated from the census itself. Such extensions in uses of the population census to non-household sectors depend primarily on utilisation of opportunities available - usually at a relatively marginal or affordable cost - during the census house-listing operation. The Indian census of 1981, for instance, incorporated the compilation of an enterprise list along with the house listing in each area; a similar proposal is under consideration in relation to the coming 1990 census of Indonesia, and I hope that this potentially efficient practice will also be considered by other countries who are interested in developing frames and surveys for small scale establishments. Similarly, the listing operation can be used to identify and distinguish between farming and non farming households - as well as rural landless and agricultural-worker households which unfortunately many conventional agricultural surveys and censuses have failed to cover adequately.
23. In addition to obtaining the sampling frame, it can also be useful to construct master sample(s) from which samples for different surveys or survey rounds in the post-censal period can be drawn readily and with appropriate relationship to each other. The term 'master sample' refers to a more specific and limited form of the sampling frame, specially constructed with some particular applications in mind. It may be useful here

to clarify this term briefly. In any survey with a multi-stage design, for instance, each stage of the sampling process involves the task of frame preparation and sample selection, till finally a sample of the required ultimate or lowest stage units is obtained. For economy and convenience, one or more stages of this task may be combined or shared among a number of surveys. The sample resulting from the shared stages is called a master sample. The objective is to provide a common sample of units upto a certain stage, from which further sampling can be done to serve the needs of individual surveys. More specifically, the objectives are : (a) to economise on costs of developing sampling frames and materials and of sample design and selection, by combining these operations for different surveys; (b) to facilitate substantive as well as operational linkages between different surveys and survey rounds; and (c) to facilitate, as well as restrict and control when necessary, the drawing of multiple samples for various surveys from the same frame.

24. Often there is a (desirable or necessary) close relationship between the large sample attached to the census, and the master sample(s) required for subsequent surveys - and hence a close relationship between the design of these two. For instance, if special attention and resources have been devoted to staff deployment and training or to the development of maps, other materials and data for the areas included in the attached sample, it will be obviously efficient to ensure that the master sample largely overlaps with these areas: and the design of the attached sample will have to take this requirement into consideration. Sometimes it is possible that the attached sample can itself serve as master sample for subsequent surveys: in which case the former should be designed as such, with 'replications' and other features which will facilitate subsampling from it at a later stage. In general, however, because of differing objectives, the two sample systems may not fully coincide. (a) For instance, to provide acceptable estimates for the whole country as well as for

numerous large and small census domains, the attached sample may have to involve greatly varying sampling rates among domains. For ordinary sample surveys, however, the reporting domains will be typically larger, fewer and probably with less extreme variations in size; and consequently more uniform sampling rates will be appropriate for the master sample. (b) For similar reasons another likely feature of the situation is that a master sample may not have to be as large as the attached sample; and therefore it may not be necessary to undertake the taxing and expensive operations of maintaining and updating the full attached sample for future use. Both (a) and (b) imply that in many situations the master sample itself requires sampling - often at variable rates - of areas in the attached sample; and this requirement too has to be considered in the design of the latter. (c) Conversely, it is also possible that the master sample has to be particularly intensive or large in certain areas of special interest (for instance to serve intensive cost-of-living surveys in a few urban centres, or surveys to monitor impacts of intensive projects or programs confined to relatively limit areas), and therefore go outside the areas included in the attached sample. Whether either or both of the sample systems should (or can) be modified to minimise such areas of non-overlap of course depends on specific circumstances. It should be noted that, generally, the census provides sampling frames or master samples only of area units for subsequent surveys. Lists of housing units, households and certainly of individuals from the census are usually not allowed because of confidentiality concerns. Also they are too difficult to arrange and too mobile and unstable to be used for subsequent surveys - with the possible exception of surveys attached to or conducted without too long a delay after the census.

25. Careful consideration must be given to functions of the census as a source of frame for diverse surveys during its planning and execution. The enumeration areas of the census have multiple functions: to partition the population into areas with

clear, stable and identifiable boundaries, with maps and descriptions; to facilitate complete and unique coverage of the units in the population; to create reasonably equitable and feasible workloads; to facilitate organization and control of census operations; to provide a flexible basis for the production of area statistics at various levels of disaggregation; and to provide a basis for scientific and efficient sample selection for subsequent surveys. While these requirements cannot all be satisfied simultaneously, they point to certain desired characteristics of census EAs. EAs should be relatively small (say a few hundred people on the average) as well as reasonably uniform in size. However, the requirement of clear boundaries is more important than uniformity in size, though very large EAs should in any case be avoided. EAs should be proper areal units covering the entire country fully and should be mapped and described for clear identification of boundaries. They should not cut across administrative subdivisions, and should be geographically ordered with proper identification systems, to facilitate the production of results at different levels of aggregation. Information on size and other basic characteristics should be collected, coded and tabulated for individual areas.

26. Apart from the use of census data for stratification, determination of sample size requirements, and sample allocation and selection, etc., census data are also useful in improving the precision of survey estimates. This applies in particular to the estimation of population totals and aggregates from relatively small samples. For example, more precise estimation can often be made by first obtaining sample proportions, means or ratios, and then inflating them with complete count figures from the census, updated from auxiliary information to the extent possible. Similarly, census (and other) information can be used to obtain more precise post-stratified estimates, or "standardized" estimates where the achieved sample is weighted to correspond to a more precise

estimation of the distribution of the population on some related characteristics (e.g., age and sex).

27. The estimation problem becomes more critical in connection with current estimates for small domains in the post-censal period; special estimation procedures are required for this purpose. The need for detailed statistical data has been augmented by increased mobility and diversity in both modern and in rapidly developing societies. Even more varied, rich and current data for local areas and for other small domains are required for planning and administration in many fields. Furthermore, the demands for data have become sharper and more critical because they are used directly for apportioning funds and resources to local areas. Census data have been used where available, but they tend to be obsolete and limited in content. Population registers and other administrative records supply the needed data only in rare cases, and sample surveys cannot generally be made on a large enough scale to yield data for small domains, especially for small administrative areas. Consequently several methods have become available recently and are being developed for providing estimates detailed both in time and for geographic and other small domains, by combining the spatial details of censuses with the timeliness of sample data and/or, less commonly, with registers when available. It is not possible to review here these more topical illustrations of the combined use of censuses, sample surveys and other sources of data, but several good references have become available.⁴

SAMPLING IN PLACE OF THE COMPLETE CENSUS

28. Finally, we must raise the issue whether complete enumeration on a 100 percent basis is indispensable, or whether in certain circumstances it may be replaced by sample enumeration or even

⁴ For example, Platek, Rao and Singh (ed) Small Area Estimation (proceedings of an international symposium at Statistics Canada).

by some other source such as continuous registration of units. Outside the population field, the term 'census' - or even the self-contradicting term 'sample census' - is often used to refer to an operation which is by no means a complete enumeration of all units in the study population. Thus most agricultural 'censuses' are in fact conducted on a sample basis, albeit often - but not invariably - on a large scale. Similarly in a census of economic establishments, while large and medium sized units are covered on a 100 percent basis, it is often reasonable as well as practically unavoidable to cover the numerous small establishments only on a sample basis. These are illustrations in which the sampling method may acceptably replace complete enumeration. Such can be the case if the units are too numerous to be completely surveyed with the available resources; or if the information to be collected is, by nature, too complex and cannot be simplified or reduced in content sufficiently to make it suitable for complete enumeration; or if the objectives are adequately served by sample enumeration despite the sampling variability to which the results are subject; or if the larger coverage errors and variances typically encountered in sample surveys (especially in estimating population totals) can be reduced to acceptable levels by appropriate estimation techniques utilising more accurate control data from some external source; or finally, if the units vary greatly in size and the contribution of the numerous small units for estimating population values of interest is, relatively speaking not so important.

29. While one or more of these considerations may apply to the enumeration of agricultural holdings, economic establishments or other units of that type, these are, in my view, not valid in the context of what is expected and required of a modern census of human population, especially as concerns the total count, distribution at the level of local areas and basic, demographic characteristics of the population. Population census data - even though they can be collected only at relatively infrequent intervals, suffer from various errors aris-

ing from numerous sources and are often published with lamentable delays - remain the most fundamental basis of all planning and perhaps of all statistics on our social existence. Some northern European countries may be in a position to consider using population registers to compile the so called 'administrative record censuses' in place of the 'conventional' censuses, but such alternatives are available to only a few countries, and to practically none among the developing ones - now or in the foreseeable future. And replacing complete count by sample enumeration in a population census - whether by choice or through lack of resources - has been found in experience to yield inadequate results, especially as concerns coverage errors.

30. Perhaps with greater merit, a suggestion has also been made that large scale periodic surveys with largely non-overlapping samples (which can be cumulated over an extended period of time to yield greater geographic detail) may be better suited than, or at least greatly enhance the usefulness of conventional censuses in meeting certain objectives.⁵ These objectives include providing current estimates for local areas and other small domains, monitoring trends and other temporal changes, and providing data which are not only spatially representative but, though averaging over seasonal and haphazard variations, are also more representative (more objectively extrapolatable) in the time dimension. Possibly this is for the future: for the moment one has to contend with the likely high cost and considerable administrative and logistical difficulties which such a scheme, involving long range planning and large-scale mobile operations over a prolonged period, may entail.

⁵ Kish (1979). Rotating samples instead of censuses. Asia and Pacific Census Forum, (6) pp 1-2, 12-13.

19. USE OF OPTICAL MARK READING SYSTEMS
IN THE CENSUS PROCESS

Use of
Optical Mark Reading Systems
in the Census Process

WR/1.9.88

Optical mark reading technology (OMR), by incorporating the use of micro processors, has become a flexible multi faceted data entry/processing system. This flexibility is no where more evident in the OMR systems configurations that are available in the census domain. The options available to statistics bureaus range from totally distributed OMR systems to totally centralized OMR systems. Varying combinations of the distributed/centralized systems are also available.

The various operational configurations are depicted in Figures 1, 2 and 3.

The total distributed system, Figure 1, shows the steps in the operation and the location at which each critical event in the process occurs. The three main operational advantages of this type of OMR system are:

- o Allows the data to be corrected in the field
- o Only valid data is forwarded for final statistical analysis
- o Allows the concurrent processing of data - particularly important in heavily populated countries with a large geographic area.

The limited distributed system (also called the combined system), Figure 2, shows the steps in the operation and the location at which each critical event in the process occurs. The two main advantages of this type of OMR system are:

- o Allows the processing of data to be load sensitive by having distributed scanning stations in heavy population areas and central scanning stations for processing remote or lightly populated areas.
- o Central processing point can process corrected or corrected forms

The total central processing system, Figure 3, shows the steps in the operation and the location at which each critical event in the process occurs. The main advantages of this system are:

- o Allows all the processing to occur at one central location.
- o Production dates easier to predict, monitor and meet.
- o Limits the amount of down or off loading of operational programs.

The total central processing system has distinct advantages in those countries with relatively low populations. The operational realities of work space and time to enter data limitations make the adaptation of the total central processing systems to those countries with very large population extremely difficult.

The seven concerns that statistics offices have in common are shown in Figure 4. Also shown are the relative impact (major/minor) that each of the 3 types of systems has on these concerns.

Heretofore there has been a tendency to consider key entry systems as the only viable input mechanism for census data collection. However, optical mark reading (OMR) technology has advanced and matured, especially since the 1980/81 round, to the point where it must be considered as viable as a method of data entry and in fact in certain countries it is the only method that will allow the timely and accurate method of effective census data entry.

OMR systems can be small, low speed, inexpensive distributed devices that function as inputting devices with a minimum of data validation. They also can be very powerful high speed processing systems that not only verify the presence/absence of data, discriminate between an erasure and a valid response, read both sides of a form on the same pass but because they are now micro processor driven can also perform a large amount of the statistical analysis that was previously done only on the host device.

OMR systems range in throughput capabilities from hand fed system to 15000 A4 sheets per hour (SPH). For purposes of comparison a typical OMR system shall be considered to have a throughput capability of 6,000 SPH. Similarly a typical key entry system shall be considered to have a processing speed of 10,000 unverified strokes per hour.

There are no universal formulae that describe all census, however, some generalities may be assumed: the ratio of population to households is 5/6 to 1; 1 in 5 households are enumerated; the remaining 4 in 5 of the population are enumerated on an individual basis.

While there may be significant variations in any single nations' census process these assumptions serve as a basis to make a sound comparison between the key entry process of data entry and the use of OMR.

Figures 5 and 6 depict the parameter analysis of key entry and OMR methods of data entry. The OMR system is assumed to have a throughput of 6000 two sided sheets per hour and the key entry process has an input of 10,000 keystrokes per hour per person. It can be seen that each scanning station processes 6000 individuals or 750 households per hour and each keystroke station process, on a unverified basis 100 persons per hour or 13 households per hour.

Figure 7 shows the productivity comparison of the two methods. The productivity advantage to the OMR system to a single keystation ranges from 58 to 1 to 120 to 1 depending on whether the keystation is operating in the verified or unverified mode.

Figure 8 shows the costs comparison of the two methods. Assuming a typical cost of 80,000 USD for an OMR station, in the unverified mode of operation and annualizing the costs, if the total compensation of the key station operator is more than 1380 USD/year then the scanning station is more cost effective. Similarly in the verified mode the break even point is 667 USD/year.

Thus it can be seen that if the total key station workload for a census is greater than 116,000 person hours for verified operation or 240,000 person hours for unverified operation then the cost of labor for the key stations is equivalent to the cost of a 6,000 SPH scanning station provided that the annualized compensation to the key station operator does not exceed 1380 USD or 667 USD respectively. Should the compensation paid to the key station operator exceed either 1,380 USD or 667 USD then the total hours to enter data break even point would be proportionately lowered. Table 1 is a compilation of how various compensation figures influence the break even point.

Figure 9 indicates the general assumptions by which it is concluded that, except for the cost of forms, the above method of analysis is valid and that in fact OMR is a mechanism that competes economically very effectively and that in fact at a very low wage rate (667 USD per year) it is possible to justify OMR on a cost basis only.

Optical mark reading forms are more expensive than normal "business" forms simply because they cost more to produce. The printing registrations specifications are tighter, a more chemically balanced ink must be used, a better grade of paper is required and a higher standard of printing in-line quality control must be exercised. All of these items add cost to the end product - the scannable form. This additional cost must be factored into the process. It is typical to find a ratio of 4 to 3 between the cost of OMR forms and business forms when comparing forms of the same size. To account for this additional cost Table 1 has a column headed "with form factor". This column has had the difference in cost of 12,000,000 OMR forms (the number of forms processed in 2,000 hours) and 12,000,000 business forms added to the cost of the scanning station as the basis for break even determination. The difference was calculated as follows:

$$\begin{array}{rcl}
 12,000,000 & \times & 25/1000 & = & 300,000 \text{ USD} \\
 \text{Cost of business forms} & & & & \\
 \underline{300,000 \text{ USD} \times 3} & & & = & 225,000 \text{ USD} \\
 & & 4 & & \\
 \text{Added cost of OMR forms} & = & & & 75,000 \text{ USD}
 \end{array}$$

Table II indicates the number of persons whose data is entered for the census at the various indicated annual compensation levels when compared to a single scanning station. The comparison value used was the one from Table 1 designated "with forms factor". What these data indicate is that when operating in the unverified mode with 12,000,000 people to be entered into the system should the annual compensation of each key station operator be approximately 2,500 USD or greater a scanning station is a more cost effective method of capturing the data including the added cost of the form. The data further indicate that similarly when in the verified mode and the key station operator annual compensation is approximately 1,250 USD or more then a scanning station is the more cost effective method of capturing the data.

Total Distributed Processing

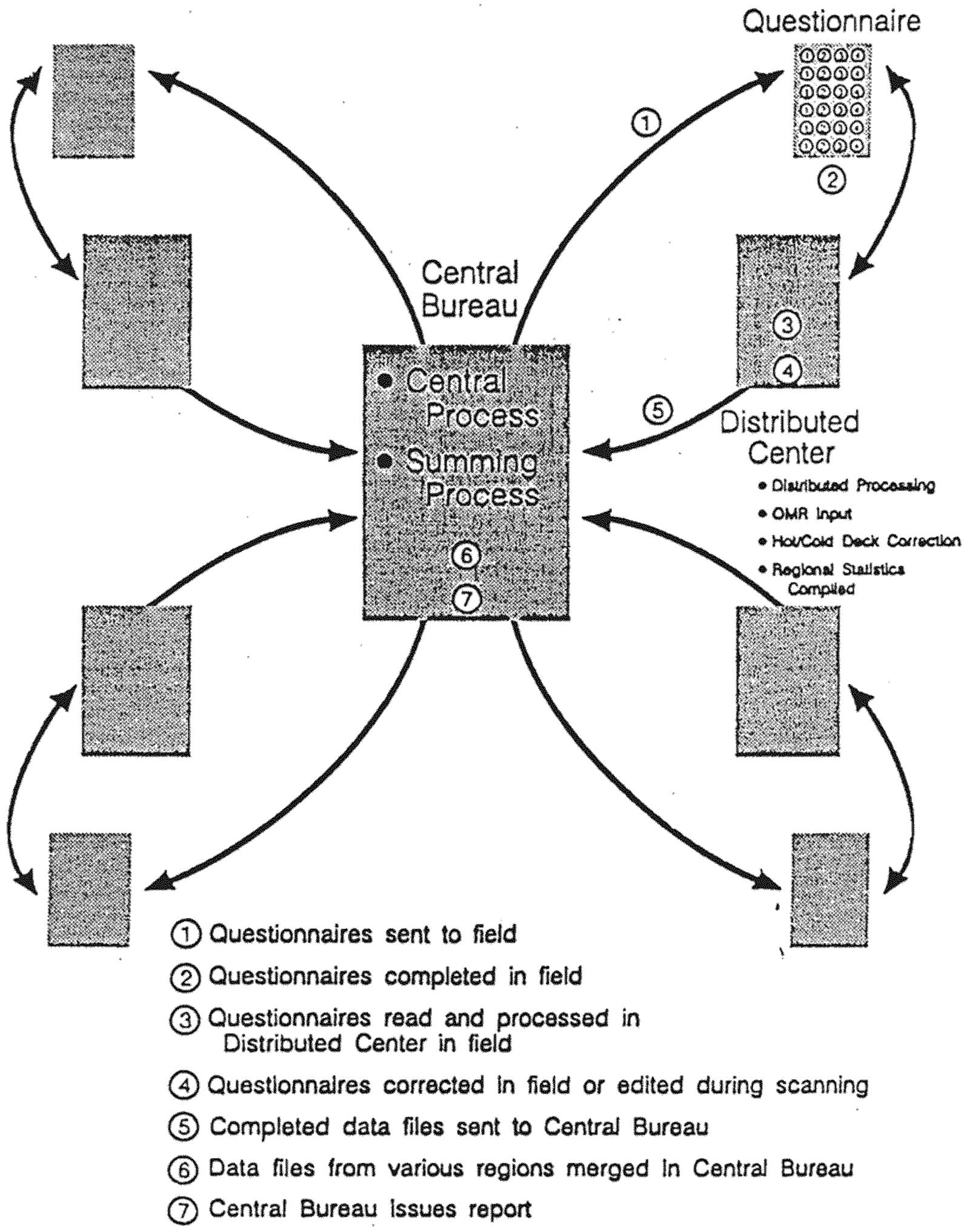
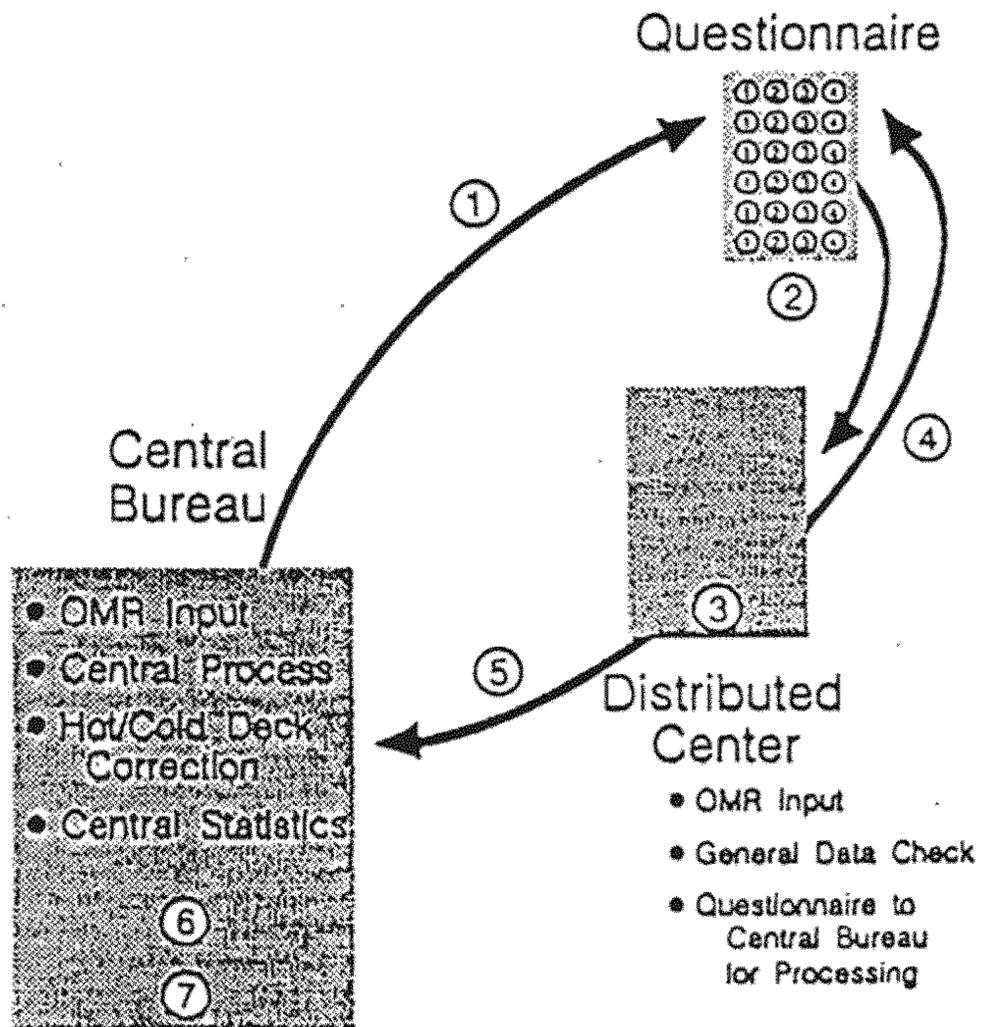


FIGURE 1

Limited Distributed System



- ① Questionnaires sent to field
- ② Questionnaires completed in field
- ③ Questionnaires scanned in field for:
 - missing data
 - obvious out-of-range inputs
- ④ Defective questionnaires returned to field for correction and re-scanning in field
- ⑤ Correct(ed) questionnaires sent to Central Bureau
- ⑥ Questionnaires processed at Central Bureau
- ⑦ Report issued by Central Bureau

FIGURE 2

Total Central Processing

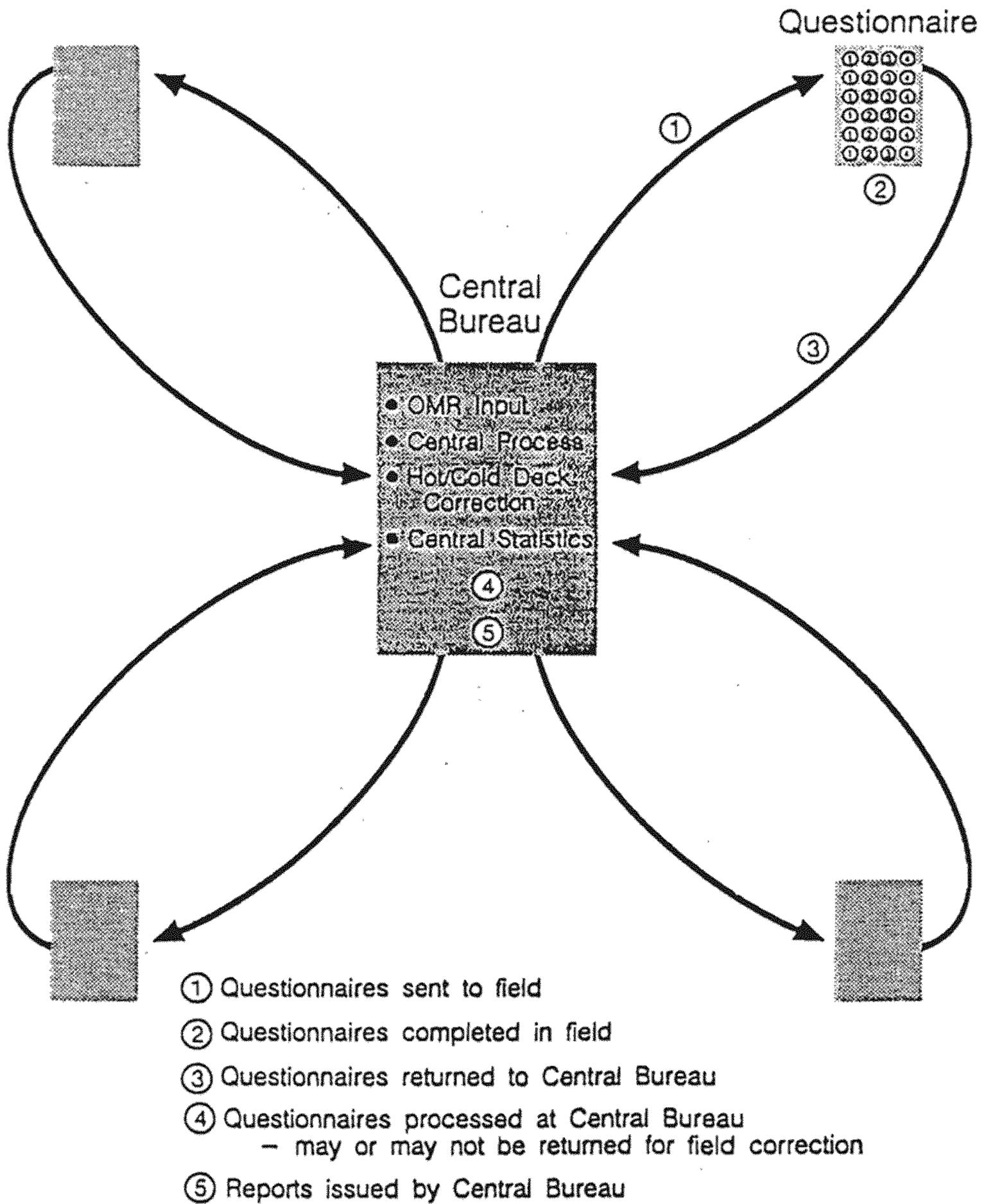


FIGURE 3

CENSUS CONCERNS

OMR SYSTEM ROLE

	<u>Centralized</u>	<u>Distributed</u>	<u>Combined</u>
Input Accuracy	Major	Major	Major
Input Completeness	Major	Major	Major
Timely Inputting of Data	Major	Major	Major
Timely Processing of Data	Minor	Major	Major
Analysis Accuracy	Minor	Minor	Major
Report Accuracy	MI *	MI *	MI *
Timely Report	MI *	MI *	MI *

* Major but indirectly

FIGURE 4

KEYSTROKE PARAMETER ANALYSIS

<u>Parameter Name</u>	<u>Mnemonic</u>	<u>Value</u>
Keystrokes per hour per operator	KHO	10,000
Keystrokes per person in census	KPP	100
Keystrokes per household	KPH	750
Person data entered per hour per operator	PPO	100
Unverified	$\frac{KHO}{KPP} = \frac{10,000}{100}$	
Person data entered per hour per operator	PPV	50
Verified	$\frac{PPO}{2}$	
Household data entered per hour per operator	HPO	13 (.333)
Unverified	$\frac{KHO}{KPH} = \frac{10,000}{750}$	
Household data entered per hour pe operator	HPV	7
Verified	$\frac{HPO}{2}$	

FIGURE 5

SCANNER PARAMETER ANALYSIS

<u>Parameter Name</u>	<u>Mnemonic</u>	<u>Value</u>
Scanner typical throughput sheets per hour	STT	6,000
Census sheets per person	SPP	1
Census sheets per household	SPH	8
Person scanned per hour per scanner	PSS	6,000
$\frac{STT}{SPP} = \frac{6,000}{1}$		
Households scanned per hour per scanner	HSS	750
$\frac{STT}{SPH} = \frac{6000}{8}$		

FIGURE 6

KEYBOARD TO SCANNER PRODUCTIVITY COMPARISON

SYSTEM	INDIVIDUAL				HOUSEHOLD			
	Unverified		Verified		Unverified		Verified	
	People Processed	Ratio OMR/Key	People Processed	Ratio OMR Key	Household Processed	Ratio OMR Key	Household Processed	Ratio OMR Key
OMR System	6,000/hr.	60/1	6,000/Hr.	120/1	750/Hr.	58/1	750/Hr.	107/1
Keyboard	100/hr.		50/Hr.		13/Hr.		7/Hr.	

FIGURE 7

Keyboard/Scanner Cost Analysis

(Based on 2,000 hour work year - i.e., Single Shift)

Keyboard			Scanner	Comparison
Annual Compensation	No. of Operators	Costs/Year	Typical Cost of 6000 SP4 Scanner	
1,000 USD	58	58,000 USD	80,000 USD	Keyboard 22,000 less expensive
1,200	58	72,500	80,000	Keyboard 7,500 less expensive
1,500	58	87,000	80,000	OMR 7,500 less expensive
1,380	58	Breakeven 80,000	80,000	Same Costs
667	120	Breakeven 80,000	80,000	Same Costs
1,000	120	120,000	80,000	OMR 40,000 less expensive

All calculated in USD

FIGURE 8

GENERAL ASSUMPTIONS IN COST ANALYSIS
(Key station vis-a-vis OMR stations)

- o Same power assumed (60 keyboard stations would in fact consume more power).
- o Assume the support required in distributing and collecting the forms to and from 60 key stations is the same as loading and unloading the input hopper of scanner.
- o Same work area (in fact the key stations require significantly more).
- o Storage area is the same.
- o Cost of maintenance of one OMR station shall be the same as the cost of maintaining 60 key stations.

FIGURE 9

TABLE 1

Annual Compensation		Equivalent Hours * of Production	
		<u>Without Forms Factor</u>	<u>With Forms Factor</u>
667	USD	240,000	465,000
1,000		160,000	310,000
1,380		116,000	225,000
1,500		106,000	207,000
2,000		80,000	155,000
2,500		64,000	124,000
5,000		32,000	62,000
7,500		2,000	41,000

* All hours rounded to nearest 500

Basis of Table

$$\text{Equivalent Hours of Production} = \frac{\text{Cost of Scanning Station}}{\text{Annual Compensation}} \times 2000$$

Production Hour Tabulation

TABLE II

<u>Annual Compensation</u>	<u>Unverified Production Hours</u>	<u>Person Per Hour (PPO) Entered</u>	<u>Total Person Entered</u>
667 USD	465,000	100	46,500,000
1,000	310,000	100	31,000,000
1,380	225,000	100	22,500,000
1,500	207,000	100	20,700,000
2,000	155,000	100	15,500,000
2,500	124,000	100	12,400,000
5,000	62,000	100	6,200,000
7,500	41,000	100	4,100,000

<u>Annual Compensation</u>	<u>Unverified Production Hours</u>	<u>Person Per Hour (PPV) Entered</u>	<u>Total Person Entered</u>
667 USD	465,000	50	23,250,000
1,000	310,000	50	15,500,000
1,380	225,000	50	11,250,000
1,500	207,000	50	10,350,000
2,000	155,000	50	7,750,000
2,500	124,000	50	6,200,000
5,000	62,000	50	3,100,000
7,500	41,000	50	2,050,000

Comparison of productivity of
annual key station compensations.

20. THE USE OF DISTRIBUTED, LOW COST
OPTICAL MARK READING (OMR)
FOR CENSUS DATA COLLECTION

THE USE OF DISTRIBUTED, LOW COST
OPTICAL MARK READING (OMR)
FOR CENSUS DATA COLLECTION.

Presented by:- Mr Stephen Stewart
Technical Director
DRS Data & Research Services Plc

Ladies and Gentlemen.

Thank you for giving me the chance to address this conference. I should like to start with one confession and one promise. I confess that I am neither independent nor unbiased. I am the Technical Director of a British company, DRS Data & Research Services plc, that designs and sells Optical Mark Readers, and you represent a significant market for my company's products! That was my confession. My promise is that my talk will be short, to the point and as free from overt selling as my conscience will permit.

The part of the census process I should like to concentrate on is the gap between the data being collected and written down for the first time, and the data being stored in a computer.

There are only three practical methods of getting data written on a piece of paper into a computer. These are to KEY the data, to use Optical Mark Reading (OMR) or to use Optical Character Recognition (OCR).

Before we compare these three techniques, I should first like to review them in turn and select the best way of using each technique for census data collection. After that we can compare the best of keying, OMR and OCR.

Keying.

There are two components to any keying system. The first component is human. This is the component which does most of the work. The second component is the keying hardware.

Twenty years ago most data was keyed on to paper tape or punched cards using mechanical devices. Only very limited validation was performed as the data was keyed. With the introduction of key-to-disk systems the quality of keying improved significantly. However for the last ten years the limits on keying speed and accuracy have been entirely those of the operator. There have been no technical developments in human key punch operators!

The main decision facing you in selecting an appropriate method of keying census data is the degree to which the keying is to be centralised. Ten years ago, in order to gain the speed and validation advantages of a key-to-disk system, it was necessary to centralise keying by using a few powerful computers each with tens or hundreds of key stations. However, with the advent of low cost mass produced 'Personal Computers' the possibility of widely distributed keying, with full validation, has emerged. The general consensus appears to be that distributing the keying and moving the operators closer to the source of the data is a 'good thing'. It is this type of distributed keying which I will use for later comparison with OMR and OCR.

Optical Mark reading (OMR).

Unlike keying, there have been dramatic developments in the performance of OMR readers in the last ten years. Before I select a particular type of OMR reader to compare with keying and OCR I should like to review the range of OMR reader currently available.

Worldwide there are two distinct types of OMR reader. One type of reader works by shining light through the paper, the other type works by reflecting light from the surface of the paper. The first type of reader is rarely seen outside the USA, and in Europe is considered obsolete. The main advantage of a transmission reader is that it is cheap to produce. The main disadvantage is that it requires expensive printing on a type of expensive paper whose supply is limited. Since this type of OMR reader normally requires that its forms be printed in the USA I do not consider it suitable for census use (outside the USA).

Having discarded transmission readers we are left with reflective readers. We can divide these into four broad categories based on their size and their theoretical reading speed. The categories I shall use are:-

**Fast floor standing (10,000 or more A4 forms per hour).
Fast desk top (5000 to 10,000 A4 forms per hour).
Medium desk top (2000 to 5000 A4 forms per hour).
Hand fed (200 to 2000 A4 forms per hour).**

Before considering each of these categories of OMR reader, the term 'reading speed' requires clarification. Most manufactures of OMR readers quote reading speeds based on the timing of a small sample of perfect forms. Since the majority of the world's OMR readers are used for reading American multiple choice tests (which are well printed on good quality paper and completed by motivated students) quoting this theoretical speed may be acceptable in this application. However if we use a fast OMR reader for census data collection the practical reading speed may be as low as 30% of the theoretical speed!

There are two main reasons for this. The first reason relates to mechanical paper feeding problems and the second to the process of correcting errors.

Unless the forms being read are printed on very high quality paper, are completely clean, and flat, then any fast OMR reader will occasionally jam. This will require the operator to remove the sheet and restart the reader. The faster the reader and the lower the paper quality the more often this will happen. If, for example, a jam happens every few hundred sheets and it takes thirty seconds for the operator to remove the jammed sheet and restart the reader, then the speed of a 15,000 form per hour reader is immediately reduced by 40% to 9000 forms per hour. By comparison the equivalent figures for a 1000 form per hour, hand fed reader are a 4% reduction in speed to 960 forms per hour.

The second reason that high speed readers do not achieve their full potential in census data entry is the correction of errors. If a validation error is found by the reader (a pregnant male example) there are two ways of correcting it. The first method is to stop the reader, extract the form, and correct it while the OMR reader waits. The second method is to have the reader out-sort the form for latter correction and rereading.

The first method can reduce the reading speed of a fast OMR reader dramatically. The second method is less disruptive but can be operationally complex with multi-page documents, and it still requires the form to be reread thus reducing the useful throughput of the OMR reader.

Returning to the first of our four categories of OMR reader, **fast floor standing**. I should like to start with what I think is the fastest general purpose OMR reader ever built. This reader (a Westinghouse 650E) was intended to read 39,000 forms per hour! This particular one was built in 1968 and was used by DRS for more than fifteen years. During that time it never read more than 15,000 forms in one hour! At DRS we replaced this machine with two fast desk top readers.

170
62
96

An example of a fast desk top reader is this DRS 4000 series system, with two OMR-80 readers. Each reader has a theoretical reading speed of 8000 forms per hour. Systems like this can be configured with up to four readers, giving a theoretical system reading speed of 32,000 forms per hour. This sort of system is used by most large educational users in the UK. In census applications where a high standard of printing and a reasonable standard of paper is available this sort of system may be cost effective.

The third of my four categories of OMR reader is represented by the DRS OMR-30. This reader has a theoretical reading speed of about 4000 forms per hour. With this slower type of table top OMR reader it is possible to read forms printed on a lower quality of paper than the faster OMR-80. There are currently about twenty of these OMR-30 readers in China. Last July they were used in China to read many millions of forms, all of them printed in China.

Rather surprisingly it is amongst the slowest (and cheapest) type of hand fed OMR reader that the most important developments have occurred in the last few years. In the past only the fastest and most expensive OMR readers were able to apply significant computer power to the problem of separating real marks from poor rubouts and paper faults. This meant that low cost hand fed OMR readers required the user to make laborious large round marks (bubbles) and were very, very, slow.

With the recent dramatic reduction in the cost of computer power it is now possible to produce a low cost hand fed OMR reader that can read simple marks, made on ordinary cheap paper. DRS has developed such a reader specifically for census applications. It is this reader, the CD200 that I shall use as an example to compare keying, OMR and OCR.

Optical Character Recognition.

The third method of getting data from a piece of paper into a computer is Optical Character Recognition (OCR). OCR has been "nearly good enough" for the past 30 years. The current state of the art in OCR is that carefully printed forms, on good quality paper, which are completed by skilled, motivated and practised people are readable. My Company sells expensive OCR readers capable of reading alphanumeric hand print. We would not offer them for general census applications.

This leaves us with only two alternatives - keying and OMR.

Before we consider the merits of keying versus OMR, I should like to suggest four criteria that can be used to evaluate these alternatives. These criteria are:-

Speed Accuracy Cost Flexibility
--

If we now consider the use of keying in relation to our four criteria:-

Since you know roughly how many people your census data will be collected from, and you know exactly how much data will be collected from each person, it should be fairly easy to calculate how many key stations are required to enter your census data in a given period. The crucial figure you require for this calculation is the average keying speed (usually expressed in thousands of key depressions per hour).

The figures quoted for this vary from as high as 15,000 key depressions per hour, down to 3,000 key depressions per hour. The high figure usually represents the rate at which a very good operator can key well prepared documents on good keying equipment after considerable practice (to put this into perspective, 15,000 key depressions per hour means pressing four keys every second, hour after hour!). The lower figures come from the more realistic route of taking the total number of useful key depressions and dividing it by the total number of hours used to key them.

Turning now to accuracy. Providing that the keying is verified, and that the key operators are merely keying clearly written data, the accuracy of keying should not be a major problem. However, census operations usually involve the transcription of some of the original data into a form suitable for keying. This transcription process can be significantly inaccurate and, its inaccuracies are added to the inaccuracies of the subsequent keying.

The cost of keying is directly related to the speed of keying. We can consider the cost of keying in two parts. Firstly the cost of the hardware on which the keying will be done and secondly the cost of the people who will do the keying. The relative costs of these two will vary dramatically from country to country. For example, in Britain you can buy a keying station for about 1,000 US Dollars, and you pay about 8 dollars an hour for key operators. In some countries you might pay four times as much for the hardware and one tenth the cost per hour for the key operators.

In order to arrive at an approximate cost for keying census data, I will make a series of fairly arbitrary assumptions. These assumptions are about right for a country with low labour costs and locally made (cheap) keying hardware. We will assume that the keying equipment, usually an IBM PC or locally available compatible, can be purchased for 1,000 US dollars and at the end of its census use can be sold for 30% of this. We will assume that the equipment is used for 12 months, with two 8 hour shifts, 6 days a week. I will also assume that allowing for training periods, sick leave, equipment malfunction, power cuts etc, an effective keying speed of 5,000 key depressions per hour is achieved. If we assume an operator cost of 80 US dollars per month, we end up with a cost of about 105 US dollars per million key depressions if not verified. If 100% verified, this figure would be 210 US dollars per million key depressions. As we shall see later this is more than 8 times the cost of using OMR.

Moving on to our fourth criteria, flexibility, there is no doubt that keying is exceedingly flexible. Key operators are by far the most flexible and sophisticated data entry machines ever devised!

If we now consider the use of OMR in relation to our four criteria:-

The only reliable method of determining the speed of an OMR reader in a particular application, is to try it. As a guide, speeds of 800 forms per hour can be achieved with a fast hand fed reader, reasonable forms and a motivated operator. To achieve this speed for long periods, two operators per OMR reader are required.

Although I should not say so, all OMR readers occasionally read marks that you did not want them to read. They also occasionally miss marks that you would prefer them to read. What OMR readers do not do (unlike key operators and OCR) is to substitute one answer for another. If you take the time to design your forms and your validation with that in mind, and if you use what we call black-tests, white-tests and skew-tests then OMR is more accurate than keying.

Now let us look at one of the most important criteria - cost. Once again, it is necessary to make a lot of assumptions in order to produce an approximate cost for collecting data via OMR. As far as possible I will make the same assumptions as were made in the keying example earlier. I will assume that each OMR reader costs about 4,000 US dollars and that a PC is purchased to go with it at 1,000 US dollars. I will also assume that the equipment is used for 12 months, for two 8 hour shifts six days a week and that the speed is a realistic 400 forms per hour. Although the operators of this equipment will not need to be keying operators, we will nevertheless assume the same monthly operator cost of 80 US dollars. I will assume two operators per OMR reader per shift.

For the OMR cost calculation we need one additional piece of information, and that is the number of characters that we get from each form. The figure I shall use here is 150. This represents the amount of data from a typical household on this type of form. These assumptions result in a cost of about 25 US dollars per million characters, compared with 210 dollars for keying.

Moving on from cost to flexibility. OMR is an excellent method of entering numeric or category data. It is poor method of entering alphabetic data. Because of this OMR has in the past often been excluded from consideration for census work. However as those of you who have seen the CD200 demonstration will know it is now possible to use OMR and keying together. This approach can give you most of the speed of OMR and all the flexibility of keying.

In summary, I would like to suggest that OMR is a fast, accurate and cheap method for collecting census data.

If OMR is so good,
why is it not
used more often?

There are three interrelated reasons why OMR has not been more widely and successfully used in census work. These are prejudice, the use of inappropriate equipment and the high cost of OMR printing.

The prejudice has arisen because 80% of the world's OMR readers are in the USA, and most of these are used to read multiple choice tests. The problem of reading millions of multiple choice tests in a country with high labour costs has resulted in a firmly held prejudice that for high volumes of forms you need fast OMR readers. This may be true for multiple choice tests in the USA, it is not true for census data in a developing country.

Ideally census data should be entered with intermixed keying and OMR, with keying being used to enter any alphabetic data and OMR being used to enter the numeric and category data. Any validation errors in the OMR data should be corrected immediately by keying. This natural, easy to manage, intuitive approach is possible with hand fed OMR readers, but impractical with large high speed OMR readers. The prejudice that has developed from the processing of large volumes of multiple choice tests has blinded people to the differences between these tests and typical census data. Multiple choice tests are usually single sheet documents, in excellent physical condition, with little or no alphabetic data, and almost no data validation requirements. Census data is often collected in multi-sheet documents, in poor condition, with significant amounts of alphabetic data, and very complex validation requirements.

The single most important reason why OMR is often not even evaluated for census work is the very high cost of printing OMR forms. But why are OMR forms so much more expensive than non-OMR forms? Firstly because OMR forms must be printed on thick high quality paper. Secondly OMR forms must be printed in two colours. Thirdly because the suppliers of OMR readers often make much of their profit by selling the printing!

All this has now changed! The hand fed OMR readers I have with me will read almost any type of paper. I have examples that I can read, that are printed on newspaper. These readers will accept forms printed in a single colour, I have examples with me printed entirely in green. At DRS we believe the way to sell OMR readers is not to try to sell you expensive printing, but to show your printer how to print OMR forms for our readers. I have examples with me printed in China on Chinese paper. Several millions of these forms were read in China in early July.

Thank you for the opportunity of presenting my opinions. I am of course happy to answer any questions.

中國의 人口 및 經濟統計現況

○ 中國의 人口統計現況

指 標	單 位	1986	1987
• 總 人 口	百 萬	1,065.29	1,080.73
• 性 比	女子百名當	107.0	106.2
• 出 生 率	千名當	20.8	21.0
• 死 亡 率	千名當	6.7	6.7
• 市 部 人 口 比 率	%	20.6(1982)	
• 平 均 家 口 員 數	名	-	4.23
• 平 均 初 婚 年 齡 (女子)	歲	-	22.1
• 文 盲 率 (12歲以上人口中)	%	-	26.8
• 合 計 出 產 力	名	-	2.49
• 期 待 壽 命 男	年	66.4(1981)	
女	年	69.3(1981)	
• 少 數 民 族 人 口 比 率	%	6.7(1982)	

註：總人口는 1982年 人口센서스 結果를 基礎로한 推計值임

附錄：中國の人口・經濟統計 現況

○ 中國의 經濟統計現況

指 標	單 位	1986	1987
• 農林漁業就業人口比率	%	-	60.1
• 1 人 當 國 民 所 得	中國元	751 (\$203)	857 (\$232)
• 產業部門別國民所得寄與率 計		100.0	100.0
農 業		34.9	32.5
工 業		45.6	46.9
建 設		6.3	6.8
運 輸		3.6	3.6
商 業		9.6	10.2
• 主要製品生産量			
TV세 트	萬	1459.4	1937.8
石 炭	百萬噸	894	928
石 油	百萬噸	130.7	130.0
電 力	10 億 kwh	449.5	497.0
시 멘 트	百萬噸	166.1	81.3
自 動 車	百萬	0.37	0.47
• 外 換 保 有 額	10 億 달러	10.51	15.24
• 輸 出 入 額	1 億 中國元		
輸 出 額		1082.0 (\$292 億)	1471.8 (\$398 億)
輸 入 額		1498.6 (\$405 億)	1615.1 (\$437 億)
• 年間 1 人 當 賃 金 額	中國元	1329 (\$359)	1459 (\$394)

註：달러 表示는 1988 年の 1 달러當 3.7 元 基準으로 換算한 結果임.

蒐集資料目錄

- 1986年 中國 統計年鑑(824 페이지)
- 1986年 中國 農村의 經濟·社會統計年鑑(289 페이지)
- 1986年 中國 統計要約 報告書(124 페이지)
- 1986年 中國 貿易 및 物價統計 報告書(243 페이지)
- 1986年 中國 都市統計 報告書(492 페이지)
- 中國의 所得 및 家口에 關한 標本調査 報告書(303 페이지)
- 中國 統計法(13 페이지)
- 1987年中 中國 經濟·社會發展現況(13 페이지)
- 中國의 統計業務現況(20 페이지)
- 1988年 中國의 統計(팜프렛)
- 中國 人口統計(팜프렛)
- 中國의 人口地圖(人口分布, 人口密度, 人口增加率)
- 中國의 1990年 人口센서스 準備現況과 基本方案(11 페이지)
- 中國의 1982年 人口센서스 資料의 利用現況(18 페이지)
- 中國의 深層 出産力 報告書(13 페이지)
- 其他 參加 各國의 Country Report(18個)

POPULATION STATISTICS IN CHINA

With its 9.6 million square kilometres of territory, China has a population of 1,080 million living in 30 provinces, municipalities and autonomous regions (not including Hong Kong, Macao and Taiwan Province) of the country. China is a nation with 55 minority nationalities, representing 8 percent of the country's total population. Obviously, population statistics plays an important role in a country with such a large population like China.

The central statistical agency in China is the State Statistical Bureau (SSB). As one of the functional departments of the Bureau, the Population Statistics Department, together with the population statistics divisions in provincial statistical bureaus, is in charge of organizing population statistics of the country. Specifically, the population statistics divisions at provincial statistical bureaus collect, through county and city statistical offices, population data of the province, make provincial tabulations and analysis, and report to SSB, while the Population Statistical Department of SSB tabulates, analyses and publishes national data on population, in addition to its major responsibility of organizing, directing and coordinating population statistics for the whole country. In recent years, the staff contingent of population statistics has been continuously expanded, their quality constantly improved, and the data collection methods and data processing facilities gradually modernized.

Population statistics are collected in China through population census, sample surveys and current population registration. After the founding of the People's Republic, three nation-wide population censuses were conducted in 1953, 1964 and 1982 respectively. The

comprehensive and reliable data obtained through the censuses have provided basis for China's socialist construction and for the formulation of development plans of the national economy. In particular, the 1982 population census has brought the census taking and the population statistics in China to a new horizon, as it was in this census that computers were used, for the first time and with success, for the processing of census data, which covered a population of over one billion (1,031,882,511 persons as indicated by the result of census enumeration), and included 19 census items as compared with the 9 census items for the 1964 census.

Until 1982, population registration comprised the major source of current population statistics in China. Since the population registration system was not comprehensive and up-to-date, the Population Statistics Department of the SSB began to conduct its own annual Survey on Population Change in 1982 on the basis of the census taken in that year. The size of the sample for this survey is 500,000 people, who were selected through scientific sampling procedures, and the data obtained from the survey are used to update the annual population figures.

In recent years, SSB has also conducted, in two phases, the In-depth Fertility Survey in China's 7 provinces and 2 municipalities, covering one-third of the country's total population. Questions asked in the Fertility Survey were those recommended by WFS regarding many aspects such as marital and fertility status of women, contraceptives used, wish and preference in child-bearing, etc., hence the large amount of data collected through this survey have a high degree of comparability with similar data collected in other countries. Meanwhile, the Population Statistics Department of SSB has also provided assistance to other government agencies in the design and the implementation of other nation-wide sample surveys, for

instance, on the disabled population, on the ageing population, etc..

Along with the course of the political and economic reform of the country, great changes have taken place in the composition, distribution and industrial structure of China's population since the 1982 population census. To reflect such changes, the Population Statistics Department of SSB carried out, in 1987, a one-percent population sample survey which is China's largest sample survey in demographic field, and obtained new information on the country's population.

The Chinese government has decided to take regular decennial census and to conduct a mini-census between the two full scale censuses. Following the recommendation by the United Nations, China's fourth census is scheduled in 1990.

Selected Statistics from the Survey on Population Change
per thousand

	1982	1983	1984	1985	1986
Birth rate	21.09	18.62	17.50	17.80	20.77
Death rate	6.60	7.08	6.69	6.57	6.69
Natural growth rate	14.49	11.54	10.81	11.23	14.08

Major Figures from China's Three Population Censuses

	1953	1964	1982
Total population	'000	723,070	1,031,882
Male	%	51.33	51.5
Female	%	48.67	48.5
Urban	%	18.4	20.6
Rural	%	81.6	79.4
Han nationality	%	94.22	93.3
Minority nationalities	%	5.78	6.7

-477-

Selected year-end population of China

	1949	1959	1969	1979	1987	
Total population	'000	541,670	672,070	806,710	975,420	1,080,730
Male	%	51.96	51.91	51.18	51.46	51.50
Female	%	48.04	48.09	48.82	48.54	48.50
Urban	%	10.6	18.4	17.5	19.0	46.6
Rural	%	89.4	81.6	82.5	81.0	53.4
Birth rate	‰	36.0	24.78	34.11	17.82	21.04
Death rate	‰	20.0	14.59	8.03	6.21	6.65

Note: Figures in this table do not include Hong Kong, Macao and Taiwan Province.

Selected Statistics from 1% Population Sample Survey, 1987

Total population	'000	1,072,330
Male	%	51.1
Female	%	48.9
Birth rate	‰	21.2
Death rate	‰	6.4
Average size of domestic household	person	4.23
Employed population age 15+	'000	585,231
Female average age at first marriage	year	22.1
Illiteracy rate (people aged 12+)	%	26.77
Total Fertility rate		2.49

Total first marriage rate of women at child-bearing ages in the past decade

	Beijing	Liaoning	Shandong	Guangdong	Guizhou	Gansu
1977	.713	1.042	.846	.802	.692	.850
1978	.884	1.081	.898	.887	.728	.974
1979	1.151	1.021	1.013	1.008	.974	1.072
1980	1.376	1.165	1.189	1.134	1.176	1.358
1981	1.423	1.173	1.301	1.065	1.103	1.211
1982	1.292	.941	1.190	1.072	1.126	.975
1983	1.163	.814	.980	1.027	1.019	1.010
1984	1.062	.910	.971	.928	.933	.768
1985	1.068	.886	1.007	.854	1.012	.964
1986	.899	.661	.831	.668	.868	.904

Selected Statistics from the Indepth Fertility Survey

Average age at first marriage for women of child-bearing ages in the past
in the past decade (non-standardized)

Year	Beijing	Liaoning	Shandong	Guangdong	Guizhou	Gansu
1977	25.4	22.9	23.5	22.7	22.0	21.1
1978	25.4	23.2	23.9	22.9	22.5	21.6
1979	25.4	23.2	23.7	23.2	22.6	21.3
1980	25.1	23.0	23.6	23.1	22.4	20.5
1981	24.8	22.9	23.5	22.9	22.1	20.3
1982	24.4	22.9	23.3	22.8	21.5	20.2
1983	24.4	22.9	23.1	22.7	21.1	20.5
1984	24.5	22.5	22.9	22.7	21.2	20.2
1985	24.0	22.4	22.6	22.9	21.5	20.4
1986	23.9	22.4	22.5	22.9	21.4	20.6

Total fertility rate in the past decade

	Beijing	Liaoning	Shandong	Guangdong	Guizhou	Gansu
1977	1.45	1.90	2.34	2.83	4.51	2.73
1978	1.41	2.39	2.28	3.12	3.83	3.12
1979	1.43	2.08	2.26	3.33	4.38	3.12
1980	1.58	1.67	1.88	3.21	4.05	2.52
1981	1.66	1.68	2.14	3.28	4.53	2.77
1982	1.92	1.64	2.18	3.16	4.52	3.06
1983	1.43	1.18	1.93	2.98	3.67	2.78
1984	1.42	.89	1.59	2.71	3.41	2.84
1985	1.35	1.04	1.70	2.43	3.33	2.34
1986	1.48	1.29	1.98	2.16	3.69	2.68

Average no. of children born to married women by age of mother

	Beijing	Liaoning	Shandong	Guangdong	Guizhou	Gansu
15-19	.80	—	.43	.47	.38	.28
20-24	.59	.69	.71	1.08	1.01	1.10
25-29	.98	1.11	1.28	1.82	2.13	2.00
30-34	1.28	1.54	1.76	2.52	3.15	2.66
35-39	1.75	2.30	2.43	3.11	3.99	3.42
40-44	2.49	3.19	3.33	3.80	4.97	4.28
45-49	3.04	3.97	4.09	4.46	5.99	5.47
Total	1.63	1.97	2.19	2.67	3.47	2.79

CHINA

STATISTICS IN BRIEF

1988



Compiled by
CHINA STATISTICAL INFORMATION
& CONSULTANCY SERVICE CENTRE

POPULATION

	(million)			
	1965	1978	1986	1987
Total	725.38	962.59	1065.29	1080.73
Male	371.28	495.67	550.75	556.58
Female	354.10	466.92	514.54	524.15
City and town	130.45	172.45	441.03	503.62
Rural	593.93	790.14	624.26	577.11

**BIRTH RATE, DEATH RATE AND
NATURAL GROWTH RATE OF POPULATION**

	(%o)			
	1965	1978	1986	1987
Birth rate	37.9	18.3	20.8	21.0
Death rate	9.5	6.3	6.7	6.7
Natural growth rate	28.4	12.0	14.1	14.1

AVERAGE LIFE EXPECTANCY AT BIRTH

	1957	1978	1981
Average life expectancy at birth for both sexes	57.0	68.2	67.9
Average life expectancy at birth for males		67.0	66.4
Average life expectancy at birth for females		70.0	69.3

**POPULATION OF CITIES WITH MORE THAN 2,000,000
NON-AGRICULTURAL RESIDENTS (end of 1986) (million)**

City	Population	
	Total	Non-agricultural
Total of 11 cities	424.9	351.5
Shanghai	71.0	69.9
Beijing	59.7	52.2
Tianjin	54.6	42.4
Shenyang	42.9	33.4
Wuhan	34.9	30.2
Guangzhou	33.6	26.5
Chongqing	28.3	21.3
Harbin	26.7	22.9
Chengdu	26.4	15.7
Xi'an	23.9	17.8
Nanjing	22.9	19.2

LABOUR FORCE

	(million)			
	1965	1978	1986	1987
Total	286.70	398.56	512.82	527.83
Number of staff and workers	49.65	94.99	128.09	132.14
Individual labourers in cities and towns	1.71	0.15	4.83	5.69
Individual labourers and labourers in collectives in rural areas	235.34	303.42	379.90	390.00

EMPLOYMENT IN VARIOUS OF THE NATIONAL ECONOMY (end of 1987)		(million)
Total		52783
Agriculture, forestry, animal husbandry, fishery & water conservancy		31720
Industry		9342
Geological survey exploration		107
Construction		2419
Transportation, post and telecommunication		1373
Commerce, catering trade, service trade, supply & marketing of materials & warehouses		2656
Real estate administration, public utilities, residential service & consultancy service		540
Public health, sports & social welfare		469
Education, culture, art, radio & television broadcasting		1357
Scientific research & comprehensive technical service		158
banking and insurance		170
Governments, parties and other organizations		925
Others		1025

TOTAL SOWN ACREAGE OF FARM CROPS				
	1978	1985	1986	1987
Total sown area (10 thousand mu)	225156	215439	216306	217435
Composition (%)				
Grain crops	80.3	75.8	76.9	76.8
Industrial crops	9.6	15.6	14.1	14.3

VALUE AND INDEXES OF TOTAL PRODUCT OF SOCIETY AND NATIONAL INCOME				
	1980	1985	1986	1987
Value (RMB billion)				
Total product of society	853.1	1658.7	190.6	2279.3
National income	368.8	703.1	789.4	915.3
National income per capita (RMB)	376	674	751	857
Index (1978=100)				
Total product of society	117.6	199.7	219.8	248.5
National income	113.9	182.3	196.6	214.9

COMPOSITION OF NATIONAL INCOME (%)				
	1965	1978	1986	1987
National income by sector				
Agriculture	46.2	32.8	34.9	32.5
Industry	36.4	49.4	45.6	46.9
Construction	3.8	4.1	6.3	6.8
Transport	4.2	3.9	3.6	3.6
Commerce	9.4	9.8	9.6	10.2
National income available				
Consumption	72.9	63.5	65.4	65.6
Accumulation	27.1	36.5	34.6	34.4

* Figures of 1987 in this table are preliminary.

POPULATION

	(million)			
	1965	1978	1986	1987
Total	725.38	962.59	1065.29	1080.73
Male	371.28	495.67	550.75	556.58
Female	354.10	466.92	514.54	524.15
City and town	130.45	172.45	441.03	503.62
Rural	593.93	790.14	624.26	577.11

BIRTH RATE, DEATH RATE AND NATURAL GROWTH RATE OF POPULATION

	(‰)			
	1965	1978	1986	1987
Birth rate	37.9	18.3	20.8	21.0
Death rate	9.5	6.3	6.7	6.7
Natural growth rate	28.4	12.0	14.1	14.1

AVERAGE LIFE EXPECTANCY AT BIRTH

	1957	1978	1981
Average life expectancy at birth for both sexes	57.0	68.2	67.9
Average life expectancy at birth for males		67.0	66.4
Average life expectancy at birth for females		70.0	69.3

POPULATION OF CITIES WITH MORE THAN 2,000,000 NON-AGRICULTURAL RESIDENTS (end of 1986) (million)

City	Population	
	Total	Non-agricultural
Total of 11 cities	424.9	351.5
Shanghai	71.0	69.9
Beijing	59.7	52.2
Tianjin	54.6	42.4
Shenyang	42.9	33.4
Wuhan	34.9	30.2
Guangzhou	33.6	26.5
Chongqing	28.3	21.3
Harbin	26.7	22.9
Chengdu	26.4	15.7
Xi'an	23.9	17.8
Nanjing	22.9	19.2

LABOUR FORCE

	(million)			
	1965	1978	1986	1987
Total	286.70	398.56	512.82	527.83
Number of staff and workers	49.65	94.99	128.09	132.14
Individual labourers in cities and towns	1.71	0.15	4.83	5.69
Individual labourers and labourers in collectives in rural areas	235.34	303.42	379.90	390.00

EMPLOYMENT IN VARIOUS OF THE NATIONAL ECONOMY (end of 1987)		(million)
Total		52783
Agriculture, forestry, animal husbandry, fishery & water conservancy		31720
Industry		9342
Geological survey exploration		107
Construction		2419
Transportation, post and telecommunication		1373
Commerce, catering trade, service trade, supply & marketing of materials & warehouses		2656
Real estate administration, public utilities, residential service & consultancy service		540
Public health, sports & social welfare		469
Education, culture, art, radio & television broadcasting		1357
Scientific research & comprehensive technical service		158
banking and insurance		170
Governments, parties and other organizations		925
Others		1025

TOTAL SOWN ACREAGE OF FARM CROPS				
	1978	1985	1986	1987
Total sown area (10 thousand mu)	225156	215439	216306	217435
Composition (%)				
Grain crops	80.3	75.8	76.9	76.8
Industrial crops	9.6	15.6	14.1	14.3

VALUE AND INDEXES OF TOTAL PRODUCT OF SOCIETY AND NATIONAL INCOME				
	1980	1985	1986	1987
Value (RMB billion)				
Total product of society	853.1	1658.7	190.6	2279.3
National income	368.8	703.1	789.4	915.3
National income per capita (RMB)	376	674	751	857
Index (1978=100)				
Total product of society	117.6	199.7	219.8	248.5
National income	113.9	182.3	196.6	214.9

COMPOSITION OF NATIONAL INCOME (%)				
	1965	1978	1986	1987
National income by sector				
Agriculture	46.2	32.8	34.9	32.5
Industry	36.4	49.4	45.6	46.9
Construction	3.8	4.1	6.3	6.8
Transport	4.2	3.9	3.6	3.6
Commerce	9.4	9.8	9.6	10.2
National income available				
Consumption	72.9	63.5	65.4	65.6
Accumulation	27.1	36.5	34.6	34.4

* Figures of 1987 in this table are preliminary.

**INDEX AND COMPOSITION OF GROSS
AGRICULTURAL OUTPUT VALUE ***

	1978	1985	1986	1987
Value (RMB billion)	139.7	362.0	401.3	467.9
Index of production (1978 = 100)	100.0	161.8	167.3	177.1
Composition of gross agricultural output value				
Crop cultivation	76.71	62.99	62.26	60.66
Forestry	3.44	5.21	5.01	4.80
Livestock	14.98	22.02	21.77	22.78
Fishing	1.58	3.48	4.10	4.81
Sideline production	3.29	6.30	6.87	6.95

* Figures in this table are at current price, calculated according to new coverage.

OUTPUT OF MAJOR AGRICULTURAL PRODUCTS

	(million tons)			
	1965	1978	1986	1987
Grain	194.53	304.77	391.51	404.73
Cotton	2.10	2.17	3.54	4.24
Oil-bearing crops	3.63	5.22	14.74	15.50
Sugarcane	13.39	21.12	50.22	47.86
Beetroots	1.98	2.70	8.31	8.16
Jute, ambary hemp	0.28	1.09	1.42	1.14
Mulberry silkworm cocoons	0.07	0.17	0.34	0.35
Tea	0.10	0.27	0.46	0.51
Cured tobacco	0.37	1.05	1.37	1.64
Fruit	3.24	6.57	13.48	16.68
Rubber	0.02	0.10	0.21	0.21
Walnuts	0.05	0.11	0.14	0.15
Pork, beef, mutton	5.51	8.56	19.17	19.86
Milk		0.88	3.33	3.30
Aquatic products	2.98	4.66	8.24	9.53

**GROSS INDUSTRIAL OUTPUT VALUE
AND ITS COMPOSITION ***

	1980	1985	1986	1987
Value (RMB billion)	514.3	971.7	1119.4	1380.6
Composition (%)				
Light industry	242.5	461.0	533.0	661.2
Heavy industry	271.8	510.7	586.4	719.4

* Figures in this table are at current price, calculated according to new coverage.

INDEX OF GROSS INDUSTRIAL OUTPUT VALUE

	(1978 = 100)			
	1980	1985	1986	1987
Industry	118.9	209.8	234.3	275.9
Light industry	130.7	246.2	276.3	326.1
Heavy industry	110.1	182.7	203.3	238.5

OUTPUT OF MAJOR INDUSTRIAL PRODUCTS

	1965	1978	1986	1987
Cloth (billion metres)	6.3	11.0	16.5	16.7
Chemical fibre (thousand tons)	50.10	284.60	1017.26	1156.9
Wollen piece goods (million metres)	42.40	88.85	251.87	259.54
Silk (thousand tons)	9.10	29.70	47.16	50.80
Machine-made paper and paperboard (million tons)	1.73	4.39	9.99	10.08
Sugar (million tons)	1.46	2.27	5.25	5.11
Cigarettes (million cases)	4.78	11.82	25.96	28.81
Bicycles (million)	1.84	8.54	35.68	40.91
Sewing machines (million)	1.24	4.87	9.89	9.60
Wrist watches (million)	1.01	13.51	73.17	61.42
Television sets (10 thousand)	0.44	51.73	1459.40	1937.84
Cameras (thousand)	17.20	178.90	2025.40	2394.20
Household washing machines (million)			8.93	9.92
Household refrigerators (million)			2.26	3.98
Coal (million tons)	232	618	894	928
Crude oil (million tons)	11.32	104.05	130.69	130.04
Electricity (billion kwh)	67.60	256.60	449.54	497.00
Rolled steel, final products (million tons)	8.81	222.08	40.58	43.91
Timber (million m ³)	39.78	51.62	65.02	68.43
Cement (million tons)	16.34	65.24	166.06	81.28
Sulphuric acid (million tons)	2.34	6.61	7.63	9.62
Soda ash (million tons)	0.88	1.33	2.15	2.37
Caustic soda (million tons)	0.56	1.64	2.52	2.74
Chemical fertilizers for agricultural use (million tons)	1.73	8.69	13.97	17.03
Metal cutting machine tools (million)	0.04	0.18	0.16	0.15
Motor vehicles (million)	0.04	0.15	0.37	0.47
Tractors (million)	0.01	0.11	0.03	0.04
Walking tractors (million)		0.32	0.77	1.10

INDUSTRIAL ENTERPRISES

	1965	1978	1986	1987
			(thousand)	
Industry	158	348	499	494
Light industry	129	204	282	275
Heavy industry	29	144	217	219

ENERGY PRODUCTION AND ITS COMPOSITION

	1965	1978	1986	1987
			(equivalent to standard fuel at 7,000 cal / kg)	
Energy production (million tons)	188.24	627.70	881.35	891.18
Composition (%)				
Coal	88.0	70.3	72.4	72.0
Crude oil	8.6	23.7	21.2	21.5
Natural gas	0.8	2.9	2.1	2.1
Hydropower	2.6	3.1	4.3	4.4

ENERGY CONSUMPTION AND ITS COMPOSITION

(equivalent to standard fuel at 7,000 cal / kg)

	1965	1978	1986	1987
Domestic energy consumption (million tons)	189.01	571.44	808.82	858.15
Composition (%)				
Coal	86.4	70.7	76.03	76.14
Crude oil	19.3	22.7	17.06	17.03
Natural gas	0.6	3.2	2.26	2.15
Hydropower	2.7	3.4	4.65	4.68

VOLUME OF PASSENGER TRAFFIC

(billion person / km)

	1965	1978	1986	1987
Total	69.7	174.3	489.7	542.7
Railway	47.9	109.3	258.7	284.3
Road	16.8	52.1	198.2	220.4
waterway	4.7	10.1	18.2	19.3
Civil aviation	0.3	2.8	14.6	18.7

VOLUME OF FREIGHT TRAFFIC

(billion ton / km)

	1965	1978	1986	1987
Total	346.3	982.9	2014.8	2192.7
Railway	269.8	534.5	876.5	947.2
Road	9.5	27.4	195.8	240.7
Waterway	67.0	377.9	864.8	941.6
Petroleum and gas pipeline		43.0	61.2	62.5
Civil aviation		0.1	0.5	0.7

POSTAL AND TELECOMMUNICATIONS SERVICE

	1965	1978	1986	1987
Service revenue of post office and telecommunications (RMB million)	628	1165	3286	3884
Letters through the post office (million)	2176	2835	4959	5479
Newspaper and magazine circulation (million)	56.21	112.50	287.31	310.05
Telegrams (million)	52.77	127.48	211.09	250.27
Long distance telephone calls (million)	88.69	185.74	423.03	515.25

TOTAL INVESTMENT IN FIXED ASSETS

(RMB billion)

	1986	1987
Total	302.0	351.8
1.State-owned units	197.9	226.2
Investment in capital construction	117.6	132.4
Investment in technical innovation and transformation	61.9	74.3
Others	18.3	19.5
2.Collective-owned units	39.2	48.1
Urban	14.6	18.2
Rural	24.5	29.8
3. Individuals in construction	64.9	77.6
urban	7.5	7.9
Rural	57.5	69.7

**COMPOSITION OF INVESTMENT IN CAPITAL
CONSTRUCTION BY STATE-OWNED ENTERPRISES (%)**

	1965	1978	1986	1987
Of the total: Agriculture	13.9	10.6	3.0	4.8
Light industry	3.9	5.8	7.0	5.1
Heavy industry	45.6	48.7	38.2	46.4
Of the total: Productive	80.6	79.1	60.6	65.9
Non-productive	19.4	20.9	39.4	34.1
Of the total: Residential building	5.5	7.8	16.1	13.4

GOLD AND FOREIGN EXCHANGE RESERVES

	(Year-end figures)			
	1984	1985	1986	1987
Gold (Million oz.)	12.67	12.67	12.67	12.67
Foreign exchange (U.S. \$ billion)	14.42	11.91	10.51	15.24

TOTAL VALUE OF RETAIL SALES

	(RMB billion)			
	1965	1978	1986	1987
Total value of retail sales	67.0	155.9	495.0	582.0
Of which:				
Urban	33.9	74.8	209.4	246.2
Rural	33.1	81.0	285.6	335.8
Of the total:				
Consumer goods	59.0	126.5	437.4	511.5
Means of agricultural production	8.0	29.4	57.6	70.5

RETAIL SALES OF CONSUMER GOODS

	1965	1978	1986	1987
Grain (million tons)	36.82	47.50	94.41	96.96
Edible vegetable oil (million tons)	0.74	0.88	3.65	4.19
Pork (million tons)	2.78	4.68	9.59	9.80
Fresh eggs (million tons)	0.34	0.46	2.24	2.34
Liquor, wine and beer (million tons)	0.94	2.46	9.57	11.38
Cloth* (billion metres)	4.4	7.70	11.78	11.99
Bicycles (million)	1.76	8.10	35.09	35.92
Television sets (million)		0.55	21.98	23.26
Wrist watches (million)	1.89	13.88	56.83	57.69

* Both chemical fibres and cotton cloth are included in cloth figures.

TOTAL VALUE OF IMPORTS AND EXPORTS

	(RMB 100 million)				
(Customs statistics)	1980	1981	1982	1986	1987
Total value of imports and exports	570.0	735.3	772.0	2580.6	3086.9
Exports	271.2	367.6	414.3	1082.0	1471.8
Imports	298.8	367.7	357.7	1498.6	1615.1
Balance	-27.6	-0.1	+56.3	-416.6	-143.3

NUMBER OF TOURISTS AND INCOME FROM TOURISM

	(thousand)			
	1980	1985	1986	1987
Total number of tourists	5702.5	17833.1	22819.4	26902.3
Foreigner	529.1	1370.5	1482.3	1727.8
Of which:				
from Japan	169.3	470.4	483.5	577.7
from USA	67.8	239.6	291.8	315.3
Overseas Chinese	34.4	84.8	68.1	87.1
Chinese compatriots from HK and Macao and from Taiwan Province	5139.0	16377.8	21269.0	25087.4
Total foreign exchange income from tourism (US \$ 100million)	6.17	12.50	15.31	18.40

PRICE INDEXES, 1986 (%)

	General index of retail prices	General index of cost of living of workers and staff	General index of purchase prices of farm and sideline products	General index of retail prices of industrial products in rural areas
1950=100	184.5	208.0	386.1	126.1
1952=100	165.1	180.8	317.5	114.8
1957=100	152.1	164.1	264.2	112.5
1965=100	136.8	149.6	205.5	106.5
1970=100	140.2	150.9	197.9	112.7
1978=100	135.8	143.6	177.5	114.7
1980=100	125.6	131.2	135.8	113.7
1985=100	106.0	107.0	106.4	103.2

PEOPLE'S LIVELIHOOD

	1978	1986	1987
Income of rural and urban inhabitants (RMB)			
Per capita annual income of peasants	134	424	463
Per capita annual income available for living expenses of families of workers and staff	316	828	916
Per capita annual consumption of major commodities (kg)			
Grain	195.5	254.8	251.4
Edible vegetable oil	1.6	5.2	5.4
pork	7.7	14.3	14.5
Per capita living space (s.q)			
Urban	4.2	8.0	8.6
Rural	8.1	15.3	16.0
Savings deposits			
Total bank savings of rural and urban inhabitants at year-end (RMB billion)	21.06	223.76	307.53
Urban	15.49	147.15	206.74
Rural	5.57	76.61	100.79

WAGES AND SALARIES OF WORKERS AND STAFF

	1978	1986	1987
Total wages (RMB billion)	56.88	165.97	188.11
of which:			
Bonuses, piece wages and overtime payments (RMB billion)	1.13	24.25	31.57
Bonuses, piece wages and overtime payments per capita (RMB)	12	194	245
Wages and salary per capita (RMB)	614	1329	1459

MAJOR FIGURES FOR STAFF AND WORKER HOUSEHOLDS *

	1981	1986	1987
Number of households surveyed	8715	31126	32855
Average number of persons per household	4.24	3.82	3.74
Average number of employees per household	2.39	2.12	2.09
Average number of persons supported by each employee * *	1.77	1.80	1.79
Monthly per capita income (RMB)	41.70	75.83	84.35
Monthly per capita living expenditures (RMB)	38.07	66.58	73.70

* Figures in this table are from a sample survey of household income and expenditures of staff and workers.

* * Employee himself is included.

COMPOSITION OF LIVING EXPENDITURES OF HOUSEHOLDS OF URBAN STAFF AND WORKERS

	1986	1987
Monthly per capita living expenditures (RMB)	66.58	73.70
Proportion of different living expenditures to the total percentage		
Purchase of commodities	91.95	91.50
Of which: Food	52.43	53.47
Clothing	14.15	13.69
Articles for daily use	11.13	11.37
Articles for culture and recreation	6.82	5.61
Books, newspapers and magazines	0.87	0.85
Medicine and medical products	0.95	1.00
Fuel	1.50	1.39
Non-commodity expenditures	8.05	8.50
Of which: Rent	0.90	0.88
Water and electricity	1.10	1.17
Culture and recreation	0.36	0.34

**BASIC INFORMATION AND NET INCOME
OF PEASANT HOUSEHOLDS ***

	1978	1986	1987
Number of households surveyed	6095	66836	66912
Permanent residents in the households surveyed	34961	338736	334970
Permanent residents per household	5.74	5.07	5.01
Average number of able-bodied and semi-able-bodied labour force per household	2.27	2.95	2.95
Average number of persons supported by each member of able-bodied labour force	2.53	1.72	1.70
Net per capita annual income (RMB)	133.57	424.76	462.55

* Figures in this table are from a sample survey of household income and expenditures of peasants.

**COMPOSITION OF LIVING EXPENDITURES
OF RURAL PEASANT HOUSEHOLDS**

	1978	1986	1987
Per capita living expenditures	116.06	356.95	398.29
Proportion of different living expenditures to the total percentage			
Purchase of commodities	97.3	96.9	94.9
Of which:			
Food	67.7	56.3	55.2
Clothing	12.7	9.5	8.6
Fuel	7.1	5.2	4.8
Housing	3.2	14.4	14.5
Daily articles and others	6.6	11.5	11.8
Cultural articles and services	2.7	3.1	5.1

**PRINCIPAL DURABLE CONSUMER GOODS
PER HUNDRED HOUSEHOLDS**

	Peasant Households			Staff and worker households		
	1983	1986	1987	1983	1986	1987
Sewing machine	38.07	46.99	49.76	76.21	73.85	74.88
Wrist watches	91.44	145.06	161.24	268.24	298.96	314.91
Bicycles	63.41	90.31	105.51	159.93	163.45	176.53
Radio sets	56.82	54.24	53.01	104.55	68.71	67.01
TV sets	3.99	17.28	24.38	83.15	92.83	99.40

STUDENT ENROLLMENT BY LEVEL

	(million)			
	1965	1978	1986	1987
Primary schools	116.21	146.24	131.83	128.36
Regular secondary schools	9.34	65.48	48.90	49.48
Specialized secondary schools	0.55	0.89	1.57	1.87
Institutions of higher learning	0.67	0.86	1.88	1.96

**STUDENT ENROLLMENT IN ADULT EDUCATION
BY LEVEL**

	(10 thousand)	
Higher education for adults	185.6	185.8
Radio and TV universities	60.4	56.6
Universities for workers & staff and universities for peasants	34.1	33.9
Correspondence and evening universities	59.5	64.6
Colleges for management cadres	5.6	5.6
Pedagogical colleges	26.0	25.1
Secondary education for adults	806.6	1047.0
Secondary specialized schools	151.2	168.2
Middle schools	213.0	142.5
Technical classes	442.4	736.3
Primary education for adults	1261.4	1351.8
Technical classes	676.0	930.3
Courses for Primary education	232.1	173.7
Literacy courses	353.3	247.8

PUBLIC HEALTH

	1965	1978	1986	1987
Number of health institutions (thousand)	224	170	203	205
Of which:				
Hospitals (thousand)	43	64	60	60
Hospital beds (thousand)	766	1856	2297	2405
Beds per thousand population	1.06	1.94	2.18	2.25
Medical personnel (thousand)	1532	2464	3507	3609
Of which:				
Doctors (thousand)	763	1033	1444	1482
Senior and junior nurses (thousand)	235	407	681	718
Number of doctors per thousand population	1.05	1.08	1.37	1.39

*** All figures of the pamphlet do not include Taiwan Province.**