PN-ABC-107

STATISTICAL TRAINING DOCUMENT
ISP-TR-4F

Popstan

A CASE STUDY FOR THE 1980 CENSUSES OF POPULATION AND HOUSING

PART F.
Variations for Local Conditions

Prepared under a Resources Support Services Agreement with the U.S. Agency for International Development

U.S. Department of Commerce
Malcolm Baldrige, Secretary
William A. Cox,
Acting Chief Economist

BUREAU OF THE CENSUS
Daniel B. Levine,
Acting Director

BUREAU OF THE CENSUS
Daniel B. Levine,
Acting Director
Meyer Zitter, Assistant Director for International Programs

INTERNATIONAL STATISTICAL PROGRAMS CENTER
Robert Bartram, Chief

Library of Congress Card No. 79-600051

SUGGESTED CITATION

U.S. Bureau of the Census, Popstan: A Case Study for the 1980 Censuses of Population and Housing, Part F, Variations for Local Conditions, Statistical Training Document, ISP-TR-4F, Washington, D.C., 1979.

POPSTAN CASE STUDY

PART A. PROGRAM CONSIDERATIONS, ISP-TR-4A

A-1. Integrated statistical system
A-2. Role and objectives of a census
A-3. Legislation, confidentiality, and publicity
A-4. Planning requirements
A-5. Uses of sampling
A-6. Census content, analysis, and use of data
A-7. Mapping
A-8. Enumeration procedures
A-9. Data processing
A-10. Publication
A-11. Pretesting
A-12. Quality control
A-13. Evaluation
A-14. Organization, personnel, and training
A-15. Equipment and facilities
A-16. Budgeting, scheduling, reporting, and operational control

PART B. PLANNING AND PREPARATION FOR POPSTAN CENSUS, ISP-TR-4B

B-1. Description of Popstan
B-2. Planning decisions and publicity program
B-3. Pretesting, quality control, and sampling
B-4. Mapping operations
B-5. Selection of topics
B-6. Table preparation and publication plans
B-7. Design of questionnaires and administrative forms
B-8. Data collection plans
B-9. Data processing plans
B-10. Coverage evaluation program
B-11. Organization and staffing for the census
B-12. Operational control system and facilities

PART C. PHASE I: COMPLETE COUNT FOR POPSTAN CENSUS, ISP-TR-4C

C-1. Summary of procedures
C-2. Questionnaires and concepts
C-3. Table formats
C-4. Data collection
C-5. Manual editing and coding
C-6. Data keying and computer processing
C-7. Volume IA: Number of Inhabitants
C-8. Volume IB: General Characteristics

PART D. PHASE II: CENSUS SAMPLE FOR POPSTAN, ISP-TR-4D

D-1. Summary of procedures
D-2. Sample design
D-3. Questionnaires and concepts
D-4. Table formats
D-5. Data collection
D-6. Manual editing and coding
D-7. Matching
D-8. Data keying and computer processing
D-9. Estimation
D-10. Volume II: Detailed Characteristics
D-11. Coverage evaluation and special reports

PART E. INTERCENSAL HOUSEHOLD SURVEY PROGRAM FOR POPSTAN, ISP-TR-4E

E-1. Objectives of program
E-2. Program content and scheduling
E-3. Sampling for intercensal surveys
E-4. Administrative structure
E-5. Enumeration plans
E-6. Processing operations
E-7. Publication, analysis, and use of data

PART F. VARIATIONS FOR LOCAL CONDITIONS, ISP-TR-4F

PART G. WORKBOOK AND GLOSSARY, ISP-TR-4G

PREFACE

Popstan: A Case Study for the 1980 Censuses of Population and Housing was developed by the International Statistical Programs Center of the U.S. Bureau of the Census. The publication covers all aspects of planning and implementing population and housing censuses—from determining the topics to be investigated and the method for collecting the information to the tabulation and evaluation of the results. It was prompted by the increasing importance of censuses in developing countries and the need for a set of materials on the interrelated complex of concepts, enumeration and processing procedures, controls, and analyses that are inherent in a census undertaking.

To provide a practical setting for the presentation of the instructional materials, the case study approach was used. A mythical country, Popstan, was created, with administrative subdivisions, social and economic characteristics of the population, housing characteristics, and a national statistical organization which was in the process of planning a decennial census of population and housing. Conceptual and procedural guidelines were developed for this undertaking, utilizing the recommendations of regional and international agencies and commissions and the experiences of various countries in similar programs. The materials are directed primarily to census technicians in developing countries; nevertheless, it is intended that countries adapt the case study to their particular needs. The titles of the seven parts of the case study and the chapters in each part are listed on the previous page.

Part A presents considerations and guidelines for planning and conducting censuses and is directed to census planners; it presents various alternatives and discusses their advantages and disadvantages. Part B introduces the mythical country and describes Popstan's decisions and preparations for its census; in giving the rationale for the decisions, reference is made to the considerations described in Part A. The complete enumeration, Phase I, is discussed in Part C; Phase II, which combines coverage evaluation with detailed characteristics for a sample of areas in Popstan, is the subject of Part D. Part E deals with Popstan's intercensal household survey program to provide current and continuing statistical data. Part F describes adaptations that may be more suitable for some countries. The final document, Part G, contains study exercises, a glossary of terms, and technical notes.

Other case studies prepared by the International Statistical Programs Center include *New Florencia* for the 1970 round of population and housing censuses, *Agrostan* for agricultural censuses, *Providencia* for economic censuses, and *Atlantida* for household sample surveys. Additionally, course materials in mapping for censuses and surveys, basic applied sampling, principles of demography, agricultural economics, and similar topics have been developed.

ACKNOWLEDGMENTS

The Popstan Case Study was developed in the International Statistical Programs Center under the general direction of **Beulah Washabaugh**, Chief of the World Census Staff, to June 1980 and **James Aanestad**, Assistant Chief, Training and Information Services, from December 1980, assisted by **Nina J. Pane Pinto**, Chief of the Census and Survey Methods Branch. Other members of the staff who were responsible for the technical review of the case study and for developing exhibits and supplemental materials include: **William E. O'Leary**, Demographic Statistician; **Janelle Fowler**, Survey Statistician; and **Leo B. Dougherty**, Survey Statistician. **Linda A. Schueter**, Survey Statistician. **Frederick A. Leedy**, Assistant Chief, Training and Information Services, to December 1980, reviewed a number of chapters for technical consistency, particularly those related to the training and control of the field staff.

Chapters on various topics were prepared by former employees of the U.S. Bureau of the Census, as follows: census objectives, organization, operational control, and budgeting by **Robert B. Voight**, former Chief of the Field Division, Chief of Operations in the Population and Housing Division, and Chief of the Data User Services Division; legislation and publicity by **Lawrence A. Marzetti**, former Chief of the Overseas Consultation and Research Office and Assistant Chief of the International Statistical Programs Center; mapping by **Robert C. Klove**, former Census Geographic Adviser; census content, publication, and uses of census data by **Elisabeth S. Nagy**, former Demographer, Population Division; quality control by **George Minton**, former Principal Researcher in Quality Control and Operations Research, Statistical Research Division; pretesting and coverage evaluation by **Herman H. Fasteau**, former Assistant Chief, Statistical Methods Division; data processing by **Dorothy P. Armstrong**, former Assistant Chief, Systems Division, and **Morton A. Meyer**, former Chief of the Demographic Operations Division and Chief of the Geography Division; and table preparation and local adaptions of census guidelines by **Edward P. Swan**, former Foreign Census Adviser. **Robert B. Pearl**, former Chief, Demographic Surveys Division, developed all of Part E, on Popstan's intercensal household survey program. **Howard G. Brunsman**, former Chief, Population Division, reviewed table formats and developed computer specifications for editing and tabulating. **Sigmund Schor**, former Chief, Statistical Procedures Branch, Population and Housing Division, reviewed the chapters on census organization and operational controls. **Charles W. Baker**, former Chief of Training, Field Division, prepared materials on enumerator training. Chapters on sampling were prepared by **Earl E. Houseman**, Sampling Consultant and former Sampling Specialist, U.S. Department of Agriculture.

Important contributions to the case study were made by a number of other persons in the U.S. Bureau of the Census. **Eli S. Marks**, Chief Census Research and Technical Adviser, reviewed and developed portions of the chapters on coverage evaluation; **Edgar H. Elam, Jr.**, Supervisory Statistician in the Decennial Census Division and former Foreign Statistical Adviser, and **Richard R. Storm**, Survey Statistician, assisted in the early developments of the case study. **Barbara A. Garner**, Survey Statistician, assisted with data processing specifications; **Maxwell D. Jeane**, Principal Researcher in Operations and Quality Control Research Staff, and **John F. Powell**, Contributing Researcher, Statistical Research Division, reviewed the materials on quality control; **Anthony G. Turner**, Chief, International Programs Staff, Statistical Methods Division and **Joseph K. Garrett**, Mathematical Statistician, reviewed the chapters on sampling and evaluation; and **J. Gary Van Nest**, Survey Statistician, assisted with the mapping exhibits and the content of the census inquiry. Many other persons in the Bureau of the Census and in outside organizations assisted in the technical review of the materials, including **Samuel Baum**, Assistant Chief for International Demographic Statistics, Population Division; **Mary K. Friday**, Computer Systems Analyst, International Statistical Programs Center; **David V. Bateman**, Assistant Chief for Methods and Development, Statistical Methods Division; and **Maurice Kniceley**, Mathematical Statistician, Statistical Methods Division.

Members of the World Census Staff prepared the publication for printing. **Sandra G. Chapman** edited the text and prepared questionnaires and control forms; **Gladys Rea** designed flow charts, prepared computer printouts of table formats, and typed statistical portions of the text; **Joan A. Godwin**, assisted by **Rhonda L. Muncey** and **Deborah A. Serfass**, laid out and typed the text; **Stella L. Sutton** read the manuscript copy; and **John R. Gibson** drafted table formats. **Dolores E. Adams**, assisted by **Ellen Shifflett**, coordinated operational aspects of preparing the case study.

The case study was prepared under a Resources Support Services Agreement with the Office of Population, Bureau for Development Support, Agency for International Development.

CONTENTS

	Page
LIST OF FIGURES	x
LIST OF EXHIBITS	x
Chapter F-1. GENERAL CENSUS MATTERS	1
Chapter F-2. DATA COLLECTION PROBLEMS	9
Chapter F-3. ONE-PHASE CENSUS	21
Chapter F-4. SAMPLE DESIGN FOR A ONE-PHASE CENSUS	43
Chapter F-5. DATA PROCESSING CONSIDERATIONS	61
Chapter F-6. ALTERNATIVE POPULATION EDIT SPECIFICATIONS	77
APPENDIX	90

Chapter F-1. GENERAL CENSUS MATTERS

1. Introduction	1
2. Inadequate financial resources	1
2.1 Pretesting program	1
2.11 Questionnaires	1
2.12 Experimental census	1
2.13 Data processing	1
2.2 Data collection plans	2
2.21 Number of questions	2
2.22 Precoded questions	2
2.23 Number of questionnaires	2
2.24 Enumeration period	2
2.25 Number of field operations	3
2.3 Equipment and facilities	3
2.4 International assistance	3
3. Lack of personnel	3
3.1 Central office staff	4
3.11 Clerical personnel	4
3.12 Technical personnel	4
3.2 Field staff	5
4. Lack of technical equipment	5
4.1 Cartographic equipment	5
4.2 Other equipment	6
5. Lack of adequate maps	6
5.1 Limited mapping program	6
5.2 Intercensal mapping program	7
6. Lack of transportation and communication facilities	8
6.1 Transportation equipment	8
6.2 Communication facilities	8

Chapter F-2. DATA COLLECTION PROBLEMS

1. Introduction	9
2. Widespread illiteracy	9
2.1 Age or date of birth	10
2.2 Place of birth	10
2.3 Industry	10
3. No common language	10
4. Lack of internal security	11
5. Nomadic populations	12
5.1 Group assembly method	13
5.2 Tribal method	13
5.3 Enumeration area approach	14
5.4 Waterpoint approach	14
5.5 Camp approach	16
5.6 Other approaches	16
5.7 General considerations	17
6. Expatriate population	18
7. Polygamy	19

Chapter F-3. ONE-PHASE CENSUS

1. Introduction	21
2. Data collection forms	21
2.1 Listing sheet	22
2.2 Short form	22
2.3 Long form	22
2.4 Continuation forms	22
2.5 Collective quarters form	22
2.6 Individual census report	23
2.7 EA Book versus booklet format	23
2.71 EA Book format	23
2.72 Booklet format	23
3. Processing	24
3.1 Field office processing operations	24
3.2 Central office processing operations	24
3.3 Long and short forms	24
3.31 Short forms	25
3.32 Long forms	25
3.4 Computer editing and weighting	26
3.41 Editing	26
3.42 Weighting	26
3.43 Diary review	26
3.5 Tabulating	26
3.6 Basic record tapes	27
3.7 Summary tapes	27
4. Quality control of data collection	27
4.1 Preparation of EA Books	27

Page

4.11 Selection of 20 percent sample..... 27
4.12 Checking..... 27
4.2 Field enumeration..... 28
4.21 Checking EA boundaries..... 28
4.22 Prelisting a sample of HU's in each EA..... 28
4.23 Observation of enumeration..... 28
4.24 Check on the 20 percent sample..... 28
4.25 Number of housing units..... 29
4.26 Kind of housing units..... 29
4.27 Persons in regular HU's..... 29
4.28 Development of crew leader inspection chart to check sample number of persons..... 30
4.29 Persons in CQ's..... 32

5. Operational control of a one-phase census..... 32
5.1 Processing of booklet style forms..... 33
5.11 District office operations..... 33
5.12 Central office operations..... 33
5.2 Processing forms in EA Book style..... 36

6. Evaluation plans..... 36
6.1 Methodology..... 36
6.2 Sample design..... 36
6.3 Questionnaires and other forms..... 37
6.4 Data collection..... 38
6.5 Data processing and matching..... 38
6.6 Estimation of coverage..... 39

Chapter F-4. SAMPLE DESIGN FOR A ONE-PHASE CENSUS

1. Introduction..... 43

2. Decisions affecting sample design..... 43
2.1 Sample collection versus sample tabulation..... 43
2.2 Choice of final unit of sampling (HU)..... 44
2.3 Exception: A sample of persons in collective quarters..... 44
2.4 Sampling enumeration areas..... 44
2.5 Sampling and data collection plans..... 45
2.51 Basic and supplementary questionnaires..... 45
2.52 Two or more questionnaires of equal length..... 46
2.53 Decision to use short and long forms..... 46

3. Sample size..... 47
3.1 Important factors..... 47
3.2 Determination of sample size..... 48
3.3 Examples..... 51
3.4 Stratified, not simple random sample..... 52
3.5 Clustered sample..... 52

4. Error estimates based on analysis of data..... 53
4.1 Need for more exact error estimates..... 53
4.2 Random group method..... 53
4.3 Assignment of random group codes within provinces..... 53
4.4 Estimates of aggregates for random groups..... 55
4.5 Variance of estimates of aggregation..... 56
4.6 Other measures of variability..... 56
4.61 Estimated standard error..... 56
4.62 Estimated relative standard error..... 57
4.7 Standard errors of differences..... 57
4.8 Estimates of percentages and ratios, and their variances..... 57
4.9 Combining province area variances for larger areas..... 58
4.10 Obtaining variances as a by-product of census tabulations..... 58

Chapter F-5. DATA PROCESSING CONSIDERATIONS

1. Introduction..... 61

2. Alternative methods of data entry..... 61
2.1 Planning..... 61
2.2 General recommendations..... 61
2.3 Methods of data entry..... 62

3. Optical mark readers (OMR's)..... 62
3.1 Advantages..... 62
3.11 Speed..... 62
3.12 Cost..... 62
3.13 Less personnel and space..... 63
3.14 Verification unnecessary..... 63
3.2 Disadvantages..... 63
3.21 Error rate..... 63
3.22 Services..... 63
3.23 Questionnaires..... 63
3.24 Processing..... 64
3.3 Options..... 64

4. Data entry edits..... 65
4.1 Types of data entry equipment..... 65
4.11 Keypunch equipment..... 65
4.12 Key-to-tape and key-to-disk equipment..... 66
4.13 Optical scanning equipment..... 66
4.14 Other equipment..... 66
4.2 Types of data entry editing..... 66
4.21 Character and field checks..... 66
4.22 Record checks..... 67
4.23 Batch edits..... 67

5. Computer facilities..... 67
5.1 Small computer systems..... 67
5.2 Large computer systems..... 68
5.3 Power failures..... 68

6. Software development..... 69

CONTENTS—Continued

	Page
7. Computer editing	69
8. Software packages	70
8.1 CENTS and COCENTS	70
8.2 CONCOR	71
8.21 Computer requirements for installation	71
8.22 Major editing capabilities	71
8.23 Input data file constraints	71
8.24 Edit diary	72

Chapter F-6. ALTERNATIVE EDIT SPECIFICATIONS

1. Introduction	77
2. Editing considerations	77
3. Population edits	78
3.1 Use of hot deck techniques	78
3.2 Preliminary edits	79
3.21 Check for valid records	79
3.22 Duplicate line numbers	79
3.23 Inmate	79
3.3 Relationship	80
3.4 Sex	80
3.5 Marital status	81
3.6 Age	82
3.7 Line number of husband	83
3.8 Mother living and mother's line number	84
3.9 Migration	84
3.91 Country of birth	84
3.92 Years lived in district	85
3.10 Education and literacy	86
3.11 Fertility	86
3.12 Economic characteristics	87
4. Serial number	87

LIST OF FIGURES

Figure		Page
F-3a	Simplified flow chart of census processing operations	25
F-3b	Crew leader inspection chart for acceptance range of number of persons enumerated in sample HU's	31
F-3c	Expected number and acceptable range of number of persons enumerated on long form in an EA	32
F-3d	Flow chart of processing operations for one-phase census-booklet style forms	34
F-3e	Flow chart of processing operations for one-phase census style forms in EA Book	35
F-3f	Post enumeration survey questionnaire	39
F-4a	Notations	48
F-4b	Factor to be applied to standard errors	52
F-4c	Worksheet: Computation of province area variances for the enumeration sample, Form PH-409	54
F-4d	Worksheet: Combining variances for the enumeration sample, Form PH-410	59
F-5a	Edit statistics report--count of edits by variable	73
F-5b	Edit statistics report--count of edits by type of edits	74
F-5c	Edit statistics report--count of edits by questionnaire	75
F-5d	Edit statistics report--control level summary	76

LIST OF EXHIBITS

Exhibit		Page
F-3-1	EA Book short-form questionnaire for 1980 Census of Popstan	90
F-3-2	EA Book long-form questionnaire for 1980 Census of Popstan	91
F-3-3	Booklet short-form questionnaire for 1980 Census of Popstan	93
F-3-4	Booklet long-form questionnaire for 1980 Census of Popstan	97
F-5-1	OMR Form	103
F-6-1	Flow charts for population edits	110

OTHER EXHIBITS FOR REFERENCE

Exhibit C-2-1	Individual census report	133
Figure D-7e	Matching form	135

Chapter F-1. GENERAL CENSUS MATTERS

1. *INTRODUCTION*

The purpose of Part F is to present further discussion and details on alternatives to the Popstan plans and on special problems faced by many countries. In order to present a more complete discussion of some of the topics, parts of the previous units have been repeated.

In the planning and conduct of a national census, the planners in some countries may be faced with problems that are relatively unique to their country or region. The issues involved are often basic to the whole census program and must be resolved during the initial planning stages. In many cases, there are no simple or clear-cut solutions to these problems. Chapters 1 and 2 of this unit discuss some alternative solutions for these problems. The solutions presented are based on country experiences and research conducted by various international agencies. As in all other matters, each country must study its own situation and select the procedures best suited, or most feasible, and adapt them according to the prevailing conditions.

2. *INADEQUATE FINANCIAL RESOURCES*

Few countries are fortunate enough to have unlimited financial resources available for the planning and execution of their census programs. For each operation of the census, the resources--personnel, equipment, facilities, etc.--required should be estimated as precisely as possible. Then, a detailed budget for each operation should be prepared. In most countries, the estimated budget will exceed the amount that the funding authorities will be willing to spend on a census. Using the detailed plans, the planners must then review each operation and the resource requirements to find ways to eliminate or reduce the cost.

2.1 *Pretesting program*

One way to reduce expenditures is to plan a more modest pretesting program. However, pretesting should never be eliminated for any operation. The pretesting program and experimental census are vital to conducting a successful census.

2.11 Questionnaires.--A possible method for reducing the costs of the census is to limit the pretests of the questionnaire. One way to accomplish this is to only pretest the questions or concepts which are new or different from those of previous censuses. Questions and concepts from previous censuses which respondents and enumerators reported having difficulty in understanding should also be pretested. In addition, it is essential to pretest questions from the previous census that had high response error rates. The possible exception to this is age, which even the most experienced data collection organizations have difficulty collecting accurately.

2.12 Experimental census.--To save money the experimental census could be conducted only in areas that have been identified as possible problem areas. Probably, a better plan would be to conduct the experimental census in fewer areas, but areas that still represent a wide variety of conditions. Again, areas or population groups that were problems in the previous census should be included.

2.13 Data processing.--All programs and procedures must be pretested. However, more

than one test of the programs can be made at the same time, thus reducing computer costs.

2.2 *Data collection plans*

The field operation is usually the most expensive part of a census program. Therefore, this area offers the greatest potential for cost control or reduction. Unfortunately, most countries prefer to continue using their traditional methods without investigating alternatives. In an effort to find ways to reduce costs, new and different data collection methods should be investigated. Refusing to discard inefficient ways of conducting the field operations may result in excessive costs which could cause the curtailment of other parts of the census program. The amount of money spent on collecting the data may be wasted if the processing or publication plans must be cutback as a result of the refusal to adopt better procedures.

2.21 Number of questions.--One way to reduce the cost of the field operation is to reduce the number of questions enumerators are required to ask or, more specifically, the amount of information collected. Items that are not needed for future planning or that produce information for a limited number of users should be eliminated if possible. The number of questions on a given topic, such as migration, fertility, mortality, etc. may be reduced. Data which are not needed on a small-area basis can be collected by surveys instead of the census. However, data needed for small geographic areas should be retained because these are better collected by a census.

Reducing the number of questions or data items will have a substantial effect on reducing the costs of the census. Interviewing time will be reduced, thus allowing the enumerators to cover more households in a shorter period of time. The total size of the field staff required can be reduced. A shorter questionnaire will also reduce the data processing workload as well as the cost of publishing the final results.

2.22 Precoded questions.--The use of precoded answer categories can save interviewing time as well as data processing time. With precoded answers, the manual coding operations can be reduced. Research has proven that precoding also reduces errors resulting from coding clerks being unable to read the handwriting of the enumerators.

2.23 Number of questionnaires.--Using a household questionnaire instead of using individual person questionnaires will reduce costs. Not only will there be a savings as a result of reducing the amount of paper needed, but the costs of controlling the questionnaires will be reduced. Often, the use of a household register and individual questionnaires means that multiple keying operations are required, thus increasing the processing costs. Household questionnaires as illustrated in the Popstan model will reduce processing costs and, more importantly, reduce keying errors since the data do not have to be keyed from separate forms. Processing errors, or for that matter all errors, add to the cost of a census.

2.24 Enumeration period.--Most countries try to collect data in the shortest amount of time possible. However, changing from a de facto to a modified de jure census and extending the enumeration period could result in large savings. Even though the enumerators will be employed for a longer period of time, the smaller field staff required to conduct the enumeration will reduce the costs of recruitment, training, supervision, and administration. Before any final decision is made, the reduction in these costs should be weighed against the costs of keeping the field offices open for a longer period of time.

2.25 Number of field operations.--Some countries take their population and housing censuses at separate times. Combining the two censuses will greatly reduce costs. In addition, combining the censuses will facilitate the cross tabulation of the data, allowing for more meaningful analysis.

The Popstan plan of having a two-phase census requires two large scale field operations. A one-phase census with a small post evaluation survey can be conducted at significantly reduced field costs. Chapters 3 and 4 of this unit discuss a one-phase census.

In many countries, due to a lack of maps, or because of a forthcoming survey, a major prelisting operation is conducted. Other countries traditionally conduct a large prelisting operation. Prelisting should be eliminated if other information is available that can serve the same purpose. That is, if the prelisting was planned to furnish a base for delineating enumeration areas or for sampling, a simple count of housing units in each block, without listing of addresses and names, could serve the same purpose, at least in urban areas. This could be done at minimum cost with the cooperation of school authorities by having students perform the operation as part of civic courses.

If a prelisting operation is absolutely necessary, the amount of information gathered during the operation could be reduced. Too often, countries try to gather too much information during this operation. Most of the time, the information is not very useful and only increases the cost of the operation. Using a smaller, better trained and supervised field staff to make sketch maps and housing unit counts is less costly, and is just as useful in most cases as a complete listing of each housing unit.

2.3 *Equipment and facilities*

While a certain level of funding is required for basic cash expenditures on equipment and facilities, there are other sources of these items within most government structures. If investigated thoroughly and utilized effectively, these alternative sources can greatly reduce the amount of cash outlay for a census.

Facilities that may be available include existing national and local government office space that can be used for training. Classrooms may be available after regular hours or during school vacation periods. Scheduling the census during school vacations can provide not only free training facilities but free office space and furniture.

Renting or using the equipment of other government agencies will also reduce costs. Sharing the cost of procuring equipment with other agencies, who may need the equipment for future programs, will greatly reduce cash outlays. Designing and making furniture that can be used for more than one operation is another way to reduce cost.

2.4 *International assistance*

Funds for census expenditures are often available from international agencies or through bilateral agreements with individual countries. This is particularly true, if the country has shown a willingness to mobilize its own internal resources and keep expenditures to a minimum. These agencies and countries can also provide technical assistance in planning and implementing the census to make it more efficient and less costly.

3. LACK OF PERSONNEL

A major problem that confronts many countries in planning and implementing a census is

the lack of personnel. This problem may be especially acute in countries that have never taken a census before, have a high illiteracy rate, and/or cannot utilize major portions of the population (for example, women). All statistical offices must enlarge their staff for a major census. Even countries which have a long history of census taking have to increase the central office staff and, of course, hire a field staff.

3.1 *Central office staff*

The type of personnel required for a census varies from filing clerks to sampling technicians. Usually the more technical the job, the more difficult it is to find a qualified person. However, if there is a limited source of educated persons in the environs of the statistical office, it may be difficult to hire persons for even the less technical positions.

3.11 Clerical personnel.--Clerical personnel includes file clerks, typists, editing-coding clerks, key operators, and the like. Next to the field staff, this is the largest personnel group that is required for a census. In many countries, this is also the group that has the lowest job retention rate; thus, replacements have to be hired constantly.

If there is a shortage of qualified persons for these positions on a full-time basis, it may be necessary to consider hiring persons on a part-time basis. This may be more feasible if arrangements can be made with local schools to institute a work-study program. The students could work part time for the statistical office; thus, they could earn academic credit and money at the same time. A program like this may also result in a higher personnel retention rate. To improve the retention rate of full time employees it may be necessary to pay higher salaries and provide more benefits for the duration of the census program.

3.12 Technical personnel.--Developing a technically qualified staff can only be done by training and experience, both of which take time. If the census planning begins far enough in advance--perhaps three to four years--before actual enumeration, then there is time to provide training and experience. Unfortunately, most countries do not have this much time to train personnel.

One possible solution is to use former employees on a part-time or loan basis if they are working for other government agencies. These persons could be made supervisors or act as advisors. They could be especially useful in training new staff. New staff that have theoretical training but no practical experience could be sent to neighboring countries having similar characteristics to observe or participate in census planning and implementation.

International agencies or other countries offer technical assistance. Experts may be requested to help plan the census and provide training. Training grants are also available for staff members to receive training abroad on a short-term basis.

When an expert is brought in from outside the country, on-the-job training and experience can be given to a relatively large number of staff members, but the training will necessarily be limited to the expert's field of specialization. On the other hand, if individuals are sent abroad for training, they may cover a wider variety of fields, but the training will be limited to a relatively small part of the total staff. In most cases, it is desirable to approach the problem using both methods--to bring in one or more experts to work with the local staff, and to select several individuals for training abroad.

University programs could be instituted to provide the theoretical background while

on-the-job training is provided by an experienced staff member. If qualified applicants cannot be found, less qualified personnel may have to be hired. Experienced staff members would do the more difficult or technical aspects of the operation while the new personnel would do the more routine tasks under the direction of the more experienced person. For example, an experienced computer programmer would design and layout the programs; then, the less experienced programmers would be assigned specific parts of the programs.

Any theoretical training that is obtained by staff members should be applied in practical situations as soon as possible. For example, the planning and conduct of the experimental census can provide useful experience. If the experimental census is still too far in the future, there may be other census and survey programs being conducted which could provide this experience.

3.2 *Field staff*

Many of the jobs in a census field operation do not require high levels of technical expertise. These include most of the jobs in the field offices. Most of the training for the field staff normally can be provided by the central office staff. Possible sources of these employees are the hundreds and thousands of people at all levels in the various government departments. School teachers, agricultural extension agents, public health employees, technicians, administrators and executives, chauffeurs and mechanics, clerical personnel--these are just a few of the types normally found in government departments whose services possibly could be utilized on a short-term basis for a census of population. Of course, special arrangements would have to be made to release these people from their regular duties and to organize and train them for a census.

In some countries, employees are granted leave from their regular jobs, and a small stipend is offered in addition to their regular salaries for participating after regular hours in census training programs and for expenses incurred in carrying out their census duties. In many countries, the census has been scheduled to coincide with a school vacation period and teachers have served as enumerators. This arrangement has been particularly effective in countries where school principals and superintendents have also been recruited and served as crew leaders and supervisors. Many of the teachers have responded to the needs of government out of a sense of patriotism, while others have been attracted by the opportunity to earn extra money during a vacation period. School teachers and certain other types of government employees are especially well suited for census work because they have experience in following written instructions and dealing with the public.

4. *LACK OF TECHNICAL EQUIPMENT*

The planning and conduct of even a minimum census program requires special equipment for such aspects of the program as mapping, data processing, and publication of the census results.

4.1 *Cartographic equipment*

Most censuses are organized on the basis of area maps which are designed to indicate clearly the limits of each enumerator's territory. The drafting and reproduction of these maps can be a formidable task, even in a relatively small country. For example, let us assume a population of 3,000,000, and let us further assume that each enumerator is expected to cover about 500 persons. A total of 6,000 enumerators would be required to cover the entire population. Each enumerator should have a map or other boundary indicator, such as

a written description of the exact limits of his/her area.

In addition to the office drafting operations, the preparation of these maps can require considerable work in the field, locating boundaries and drawing sketch maps. Drafting instruments and supplies, special photographic equipment, and map reproduction facilities must be provided for the mapping operation.

If the statistical office cannot afford to buy all of this equipment, there are alternatives. For example, if there is an agency responsible for topographic mapping in the country it may be able to provide assistance. This assistance could range from doing the entire job of census map preparation to allowing the statistical office to use its reproduction equipment. It should be remembered that census maps do not require the same high accuracy needed for topographic maps. This means they will not take as long or be as costly to produce as topographic maps.

Another alternative is to buy the basic drafting equipment to use in the statistical office for preparing the base maps. The photographic copying, enlarging, and reducing work can be done by another government agency such as the government printing office or by a private company. To produce the copies of the base maps for use by the enumerators, it may be possible to rent a diazo or ammonia process machine.

4.2 *Other equipment*

The publication of the census results requires printing and binding equipment. There are many countries who do not have a government printing agency; therefore, other means of printing the reports must be found. Census reports are one of the most important ways to publicize the work of the statistical agency and build support for future censuses. Hence, well designed reports and a good printing job are an investment in public relations.

There are usually private printing establishments within the country. In many cases, contracts are simply rewarded on the basis of the lowest cost. The statistical agency should award printing contracts on the basis of estimated cost and the quality of previous work in order to receive the best quality for their money.

If good quality printing is not available within the country, then the agency may have to go outside the country to have the reports printed. There are obvious disadvantages to this arrangement. The most important of which is that quality control procedures cannot be easily instituted.

5. *LACK OF ADEQUATE MAPS*

In some countries, census operations have historically made little or no use of maps. Because of this, these statistical offices do not have a map inventory which can be updated for the next census. Preparing up-to-date maps for the enumerators to use can take from 2 to 3 years even if the maps from the last census are available. Obviously, if there is no previous work to build on, the task of mapping the entire country will require much time and resources.

5.1 *Limited mapping program*

The importance of maps in a census is discussed fully in Part A, chapter 7. The various steps and operations in a mapping program are also shown. One of the first steps in setting up such a program is preparing a list of areas to be mapped and assigning each a priority for completion. If time or funds are too short for a full mapping program, the establishment of priorities is even more important. Even a limited mapping program will be useful and can

form the basis of a continuing effort to map the entire country during the intercensal period. Remember that the maps used in the census will not be thrown away after the census but will be kept and updated for the next census. They can also be used for sample surveys and by other government agencies for such things as planning transportation, utilities, and health programs. For this reason, other government agencies may be willing to share the cost of the mapping program.

A limited program should concentrate on mapping those areas with the highest population density, such as cities. The reason for this is that maps are used to prevent coverage errors and this type of error is most serious in densely populated areas. Another type of map that should have a high priority is one showing the boundaries of the areas supervised by the lowest level of field offices, those directly supervising the crew leaders and enumerators. In addition to contributing to complete coverage, such a map is useful in planning staff allocation, assigning work, and monitoring progress.

Once the areas to be mapped are chosen, the procedures outlined in Part A, chapter 7, are followed. If maps and air photos are not available for use as source materials in preparing the census maps, then other methods are necessary. Teams of cartographers could be sent into the field to prepare the maps. An alternative is to have another government agency or the military provide aerial photography for the selected areas. Maps can then be traced from the photographs. This is usually less expensive, faster, and more accurate than sending out field teams. If time does not permit maps to be traced from the air photos, then they can be given to the enumerators and used as maps themselves. The photos as well as the finished maps can be useful for other projects besides the census.

It may also be possible to use satellite imagery to prepare small scale maps of the entire country. The scale of map that could be produced from this method would be most useful for the rural or less populated areas of the country.

Consideration should be given to making arrangements with universities, colleges, and secondary schools to institute mapping courses for credit. The students could do the field sketching and the central office staff the final maps. In some instances, students of the higher educational institutions could also do the final maps. This type of arrangement could not only be useful for the immediate needs of the statistical agency but may provide potential employees for the statistical agency who will be trained in census mapping.

5.2 *Intercensal mapping program*

As mentioned earlier, producing census maps for the entire country without being able to build on any previous work would be a long and costly undertaking. For this reason, it is very practical for a country to gradually build up its map coverage over a period of many years. The mapping program should not be forgotten after the census is completed. To do so risks losing the valuable investment even a limited mapping program requires.

A mapping program should be supported during the time between censuses. The objectives of this program would be to gradually expand the census map coverage and update existing maps. It should be pointed out that maps become outdated at different rates. Maps of expanding metropolitan areas become outdated very quickly while those of rural areas may be quite useful for many years.

6. LACK OF TRANSPORTATION AND COMMUNICATION FACILITIES

A census of population extending, theoretically at least, to every household in the country, may make extraordinary demands on the limited transportation and communication facilities available within the country. Except for processing equipment, transportation and communication equipment are usually the most expensive items in a census budget.

6.1 *Transportation equipment*

In all censuses some transportation must be provided over a long period of time--perhaps one or two years--for preparatory work in the field, including mapping, contacts with local leaders, recruitment and training of field personnel, advance publicity for the census, etc. For a shorter period of time--perhaps one to three months--a large number of vehicles may be required to transport enumerators and supervisors to their assigned areas; to distribute questionnaires, training materials and other supplies to the census field offices; and to collect the completed questionnaires and unused supplies after the data collection operation is finished.

In addition, it may be necessary, particularly in rural areas where public transportation is very limited, to provide some means of transportation for enumerators and supervisors who must travel long distances from one house to another, or from one village to another. The type of transportation will depend, of course, on the terrain, but may include jeeps, bicycles, motor scooters, horses or mules, boats, and aircraft such as helicopters.

Sometimes enumerators arrange for their own transportation and the Central Office pays a fixed allowance for the rent of a vehicle, horse, boat, etc. The primary advantage of this system is that it simplifies the work of the Central Office and places all responsibility for arrangements on the enumerator.

In other cases, it may be necessary for the Central Office to arrange for the rent or loan of vehicles or other means of transportation from another government agency. The military agencies are often the best source of trucks and jeeps outside the major cities. However, vehicles are sometimes available from other government departments or international agencies conducting programs of various types in the field.

In one country, various government agencies ordering vehicles for their own use arranged an earlier delivery date and then loaned them to the statistical office for the duration of the enumeration period.

Another solution, which has been effective in countries with high import duty on vehicles, is to buy the vehicles duty-free for government use on credit; then after the census, the vehicles are auctioned off, still exempt from duty. This procedure may require a special decree or law. However, this solution may be well worth the effort since in most cases the auctions have more than covered the original cost of the vehicles.

6.2 *Communication facilities*

Lack of communication facilities can also present a problem, particularly during the period of enumeration when unexpected situations arise in the field and questions need to be answered quickly. Two of the best sources of assistance on the problem of communications are the military and police. Practically every military outpost or police station has a telephone or some means of communicating with headquarters. If arrangements can be made to utilize these facilities during the enumeration period, the problem of communication will be largely resolved. Other government agencies, such as the Ministries of Education, Health, or Agriculture may have their own communication network which may be made available to the census staff.

Chapter F-2. DATA COLLECTION PROBLEMS

1. *INTRODUCTION*

In some countries, the National Statistical Office is faced with special conditions in the area of data collection. These conditions arise because of certain characteristics of the country's population which make it especially difficult to conduct efficient data collection operations.

The following is a list of some of the most common conditions encountered by developing countries in the planning and conducting of national censuses of population and housing:

(1) Widespread illiteracy

(2) Lack of a common language for all segments of the population

(3) Lack of internal security, particularly in remote rural areas

(4) Large segments of the population which are nomadic

Other items could be added to this list, such as villages with no recognizable street patterns, houses with no house numbers, large segments of the population with the same family name, and frequent name changes.

The discussion which follows is not intended to be exhaustive. Its purpose is to focus attention on these problems during the early stages of planning, so that alternative solutions may be considered and tested in the field before the census enumeration gets underway. Some lines of action which may be helpful in dealing with the various problems are suggested but no standard solution is proposed.

2. *WIDESPREAD ILLITERACY*

If a substantial part of the population in any given country is illiterate, this places certain restrictions on the content of the census as well as on the method of collecting and processing the data. People who are illiterate generally do not think in the precise terms that we like to have in a census nor do they maintain written records to which they can refer. For example, it may be difficult or impossible to get exact ages, and it may be necessary to use some method of estimating age. Similarly, information on place of birth may not be reported in the exact terms of the census questionnaire; an illiterate person may know the name of the village or hamlet in which he was born, but not the province or district. Again, it may be difficult to obtain precise information on economic characteristics--occupation, industry, hours worked, and income. The concepts used must be defined and expressed in relatively simple terms.

It should be noted that illiteracy does not preclude the use of precoding on the questionnaire. This is especially true when questions have limited answer categories such as male and female or yes and no.

What action can be taken by the Census staff to ensure the best information possible? First, a careful review of the proposed content of the census may indicate that it is not feasible to collect information on some items, such as personal income or age at first marriage. But for other items, such as age, place of birth, and employment, a little patience on

the part of the enumerator and some special aids provided by the central office may yield acceptable data.

2.1 *Age or date of birth*

A number of countries have used a calendar of major events to help the enumerator obtain the approximate age of individuals who were not able to give the exact date of their birth. Events which are likely to be remembered, such as the term of office of a national leader, or the visit of an outstanding foreign dignitary, or an earthquake, drought, or other natural phenomenon, can be used quite effectively as reference points for fixing the approximate date of birth of an individual. It is even more helpful if separate calendars are prepared for different parts of the country. Each calendar would include events that were significant for the local people but not necesarily nationwide. Within a particular household, the order of birth of children may be useful in estimating the age of one child as more or less than that of another.

Remote tribal groups, which have no concept of time or for which calendars of local events may be too difficult to obtain, may need special procedures. If the area in which the tribe resides has definite seasons, only the number of summers or winters an individual has lived may be ascertained. In countries where there are no well-defined seasons, only approximate age groups may be obtained such as "reached puberty" and "has not reached puberty." The tabulations will have to take into consideration this type of data. Sometimes it is not even feasible to use the standard questionnaires, as most of the questions are irrelevant and it is impossible to obtain a separate questionnaire for each household. One country decided to take only a head count by sex and puberty status. To obtain as correct a count as possible, trinkets for each age-sex group were distributed under the auspices of the chief. The enumerator knew the total count for each type of trinket at the beginning of the distribution process. A count of the trinkets left after distribution subtracted from the original count gave the total number of persons for each category.

Another country used a different approach to get a count of its island population. One boat went from island to island requesting the local chief to have the islanders assemble on the beach in four groups: men, women, boys, and girls. Shortly afterward another boat arrived carrying the enumerators to do the actual counting. In this way several islands were enumerated in one day.

2.2 *Place of birth*

For place of birth, it may be helpful to provide the enumerator with a list of villages or barrios showing the district, province, or other civil division in which they are located. If there are several villages with the same name, these can be distinguished by their proximity to some natural feature, such as a coast line, a river, a mountain, or a large city which should be included on the list.

2.3 *Industry*

In some countries, it has been found helpful to provide the enumerator with a list of the most important businesses and industries in the area with a brief statement of what they do. This technique can be especially useful in cases where the name of the establishment gives no indication of the kind of business or activity in which the workers are engaged.

3. *NO COMMON LANGUAGE*

When several different languages or dialects are used in a country, and there is no common language which everyone understands, special measures may be required to collect the desired information. If the problem is

localized in a particular area, the census office should consider the hiring and training of local residents as enumerators. Residents of the locality would probably be familiar with the language or dialect spoken and would be able to ask the census questions in a language which the respondents would understand.

In some countries, where there is no common language, the census questionnaire has been printed in the most common languages or dialects used in the country. This can be useful in the field, but may require special handling in the Central Office. In other countries, the Central Office has provided translations of the census questions which the enumerator could use along with the standard questionnaire. Whether the questionnaire is printed in different languages or a translation of the questions is provided for the enumerator to use, the object is to standardize the questions for all parts of the country and to reduce enumerator variance.

In most countries, there are instances of individuals, recent immigrants, or diplomatic personnel, who are not familiar with the national language. For such cases, the enumerator should be instructed to find some member of the household or a neighbor who can interpret the questions and answers for the enumerator and respondent. If this type of population is scattered throughout the country, a core of special enumerators who have multiple language capabilities may have to be hired. These enumerators could be attached to the central office or positioned at various locations and sent out to the field as needed.

4. *LACK OF INTERNAL SECURITY*

Sometimes, there are segments of the population which do not fully support the national government and which, therefore, present special problems in a census of population. The attitude of the people may vary from open hostility or lack of cooperation to mere indifference. Occasionally, these groups are found in urban ghettos, but more often they are found in remote rural areas which have only loose ties with the central government. (Nomadic tribes which operate as semi-autonomous units within a country or across national borders represent a special problem which will be discussed in a later section.)

So far as the census is concerned, the problem of enumerating these segments of the population is primarily one of security. To meet this type of situation, it is sometimes necessary to employ members of the armed forces or local police to ensure the safety and protection of the census enumerators. Usually, these people simply accompany the enumerators on their journey from one house to another and do not participate in the census interview. In some countries, however, members of the military or local police are employed directly as census enumerators. This practice is generally not recommended because it introduces an element of coercion into the enumeration process and may have an adverse effect on the quality of the census data. An exception to this is in an area which is normally administered by the police or military who would usually carry out all government programs.

If, in a given situation, it is necessary to use members of the armed forces either as enumerators or enumerator aides, special arrangements should be made with the authorities responsible for internal security. It is important, in such a situation, to ensure that any agreements made at the top levels of government be transmitted to the local government authorities with whom the census enumerators will be primarily concerned. Officials

responsible at the upper echelons of government for national security and defense often have a direct interest in the census results and are quite willing to cooperate in the planning and execution of the census program. Unless appropriate steps are taken, however, this support of the census program may not extend to the local level--and particularly to remote rural areas.

Generally, it is not wise to bring in enumerators from outside the area because almost certainly they will be regarded with suspicion by the local residents. Even though they may have superior qualifications from the standpoint of education and technical skills, these qualifications will be of little value if the individuals themselves are not acceptable to the local residents. Enumerators recruited from within the area will have a better knowledge of local customs, will know where to find the people, and will be in a better position to evaluate the information furnished by the respondents. They are also more likely to be trusted by the local people. School teachers, health and welfare employees, agricultural agents, and persons engaged in similar activities make excellent enumerators because they are already accustomed to working with the local people and dealing with local problems.

In addition to any special arrangements which may be necessary to ensure the safety and protection of enumerators, it is important that the necessary steps be taken to gain the cooperation of local leaders and to convince them of the necessity and/or desirability of being included in the national census count. Any relationship that can be established between the census and other programs of the central government aimed at the improvement of local conditions will, of course, be beneficial to the census. A program of agricultural development, for example, or improved health care, or expanded educational facilities can be related directly to the census since all of these programs are people-oriented. It is important, however, that the future benefits resulting from the census be realistically presented. In the past, the credibility of some statistical organizations has been greatly harmed by promising many immediate benefits in housing and jobs which did not materialize. Often a visit to a local chief, or some show of mutual trust and respect, will open the way for the census program to be carried out to the benefit and satisfaction of all concerned.

Whatever measures are necessary to ensure the safety and protection of enumerators and supervisors and the successful execution of their responsibilities in the field, plans must be made and action taken well in advance of the census date.

5. *NOMADIC POPULATIONS*

The enumeration of nomadic tribes, whenever they constitute a substantial part of the total population, requires special attention. Because there are varying degrees of nomadism, and because life styles vary from one tribe to another, there is no recommended standard approach to the problem. Different approaches have been tried with different degrees of success or failure. Accordingly, the best strategy for a country with a significant nomadic population is to study the various methods used in other countries where conditions are similar and to select and test ones that appear most likely to succeed. Much of this section is based on United Nations experience in dealing with this problem.[1]

[1] *"Study on Special Techniques for Enumerating Nomads in African Censuses and Surveys," Economic Commission for Africa, Tenth Session of the Conference of African Statisticians, Addis Ababa, Ethiopia, 1977.*

Of course, before deciding on an enumeration plan it is necessary to know something about the nomad's way of life. It is important to know such things as languages spoken, migration patterns, economic activities, and the hierarchical structure of the tribes. For some enumeration methods, it is also necessary to know the types of animals they own and the amount of influence the tribal leaders have. Possible sources of this information are government agencies responsible for nomads and university instructors of anthropology and sociology. If these sources are not sufficient, it will be necessary for NSO personnel to conduct research about the nomads.

5.1 *Group assembly method*

In the countries of Africa and the Middle East, where nomadism is prevalent, several distinct approaches have been tried for the enumeration of nomadic tribes. Perhaps the easiest one, from the standpoint of the enumerator, is the group assembly approach. Under this method, the nomadic households (or their representatives) assemble at a fixed place and on a fixed date, and the enumerator systematically enumerates the households in some prearranged order. The date and places chosen for assembly should be as convenient as possible for the nomads within the limits of the overall enumeration plan. The place of assembly may be a market place, a watering point, a grazing area, a food distribution center, or other convenient location.

The principal advantage of the group assembly method is that it requires a minimum amount of travel on the part of the enumerator. The main disadvantage is that a large part of the nomadic households may not come to the appointed place at the appointed time, and consequently, the population can be seriously underenumerated. The extent of this may be difficult to determine. It may be necessary to give them an incentive to come, such as small amounts of money, food, or other goods. For example, in one country the field staff enumerating the remote tribes assembled all the families of a tribe at one location. They then bought some food from each family as it was enumerated. After the enumeration of each tribe was completed, a feast was held for the tribe using the food bought by the field staff.

A variation on the group assembly method is to have only the tribal leaders or their representatives come to the assembly point. One possible problem is that they will not know all the necessary information for each household in their tribe. To remedy this, they could be told what data will be needed before they leave for the assembly point. Another difficulty with this approach is the possible under or over reporting by the tribal representatives. The post enumeration evaluation program would have to explore that possibility.

5.2 *Tribal method*

Another approach which is commonly used is the tribal or hierarchical approach. Under this method, the principal contacts are with the tribal chief and he either furnishes the information for all members of his tribe from memory or from tribal records (administrative census), or makes arrangements for the enumerator to interview directly the various households under his jurisdiction. This method utilizes the authority and prestige of the tribal chief to gain the cooperation of his followers. A list of the tribes, their locations, and the approximate number of households in each must be prepared. This information is important to properly plan the enumeration. The tribal leaders should be contacted in advance to enlist

their cooperation and to arrange for the visit of enumerators at the time of the census.

An advantage of the tribal approach is that the respondents are usually cooperative because they respect the tribal chief. Also, the field work can be accomplished more efficiently if the chief schedules the interviewing of the individual households. This may be the only practical method if the nomads have a high level of mistrust or suspicion for the government or the sedentary population.

However, in practice, much depends upon the personality of the chief and on his ability to contact all members of the tribe. The tribe may be spread out and the chief may not know the exact locations of all the families. Furthermore, there may be a problem in defining the population to be enumerated under this procedure since some individuals who have adopted a non-nomadic life style may still be regarded by the chief as members of the nomadic tribe. This can lead to over-enumeration since the sedentary part of the tribe may also be enumerated with the non-nomadic population. If this is true for a significant number of nomads, then all the tribal households must be contacted. Also, questions must be included in the questionnaire which will enable the census enumerator to distinguish between the true nomads and the settled population. This will also help identify double counting.

5.3 *Enumeration area approach*

This approach divides the entire land area into conventional census enumeration areas. The nomadic households would be enumerated in whichever area they happened to be during the census period. Assuming personnel and other resources are available, this method should assure complete coverage of the nomadic population. This approach requires maps to divide the country into enumeration areas and to help the enumerators locate their assignments. Aerial photographs may also be used for this purpose.

In addition to the increased accuracy resulting from better coverage, the reliability of the data will also be improved by the direct interviewing of each household. The enumerator does not have to rely on the memory of the chief nor will the chief's possible tendancy to under or over report introduce error.

One problem with the enumeration area approach is that it requires a great deal of resources if the nomad population is large or widely scattered. Enumeration crews may have to be transported to and maintained in remote areas. Also, to avoid missing migrating groups it is necessary to have a relatively short enumeration period. This, of course, requires a larger field staff. The longer the enumeration period the more likely it is that movement of the nomads will take them across enumeration area boundaries, thus increasing the chance of coverage errors. This is less of a problem if the enumeration is conducted at a time of the year when the nomads are less mobile.

5.4 *Waterpoint approach*

The waterpoint approach takes into consideration the fact that the nomads must graze and water their animals. This enumeration technique would require that a list of waterpoints used by the nomads be prepared and enumerators be stationed at those waterpoints during the census period.

There are several methods of preparing a list of waterpoints. One is to ask the local government officials to provide lists for their jurisdiction. If this is done, a sample of the results should be checked for accuracy by the statistical office field staff. Another

way is to have the statistical office field staff prepare the list from their own canvass of the country. In either case, some guidelines should be prepared to help the field workers properly identify waterpoints. Aerial photography and satellite imagery can also be useful if recently produced and if there are trained interpreters available.

The waterpoint approach is usually practical only during the dry season when the waterpoints are much fewer in number. If the nomads cluster closely around the waterpoint, then the enumerator would be instructed to cover a specified surrounding area. If they tend to graze at some distance and come to the waterpoint only periodically, then an enumerator would wait there and enumerate each family or group as it arrived.

An important factor that must be considered when planning the enumeration for this second situation is the types of animals raised by the nomads. Various animals have different watering needs. For example, cattle must usually be watered every other day while camels may require watering only every two or three weeks. Therefore, the enumeration period must be at least somewhat longer than the shortest watering period. This is to ensure that the enumerators are at the waterpoint for at least one visit by each family or group. However, using the shortest watering period as the basis for the length of enumeration can lead to some problems. Sometimes, families own several types of animals with differing watering periods and arrangements may be made between several families to share the task of watering. For example, one family may bring the cattle of several other families for watering. At another time a different family would water the camels for these families. If the enumeration period was based on the shorter watering period for cattle, the second family would probably be missed. This illustrates how important it is to fully understand the local customs when planning enumeration procedures. In such cases, the length of enumeration should be based on the longer watering period of the camels. However, even this procedure may not ensure complete coverage as sharing the watering responsibilities may mean a family appears at the waterpoint less often than even the longest watering period.

Another problem occurs when people decide to stay away from the waterpoint because they do not want to be counted. To counter these situations the enumerators should ask those who do come to the waterpoint the names of any families who have not been there since the start of enumeration. As the enumeration progresses those families that eventually show up can be removed from the list. Towards the end of the enumeration period arrangements can be made to find those families who have not been enumerated.

If the number of waterpoints is too large for each one to be covered with the available resources, an alternative procedure would be to select a sample of waterpoints. This of course requires a complete list of all the possible waterpoints from which to select the sample. The enumerators would be sent only to the selected waterpoints for the enumeration period. A problem with having enumerators at only a sample of waterpoints for an extended period of time is that some nomads may avoid the sample places because they do not want to be enumerated. Another method which might avoid this problem would be to enumerate in the sample areas for one day only. Care must be taken when choosing the day of enumeration so as not to bias the results. Certain days may be preferred over others for watering. To avoid this bias, the original sample could be subsampled to have

the visits to the selected waterpoints distributed over several days.

There are several possible drawbacks to the waterpoint system of enumeration. The accuracy of the coverage depends on how complete the list of waterpoints is. Also, in some cultures the younger members of the family are responsible for watering the herd. They may not always be knowledgeable respondents. This situation would require finding and interviewing an adult member of the family. If this is a common occurrence, then the waterpoint approach would not be an efficient enumeration method. Despite these problems, the waterpoint approach has been successfully used in a number of countries.

5.5 *Camp approach*

In many countries nomads live in camps. These are households which live close to each other and form a somewhat stable group. The size and locations of the camps may change from season to season. However, they can form the basis of an enumeration plan if their locations are known and a field staff can be organized to visit and enumerate them.

The first step in using this approach is to prepare a list of camps and their locations. Knowledge of tribal customs and migration patterns will be helpful in knowing the areas where camps might be located. One method of preparing a list of camps is to ask local government officials to report all the camps in their jurisdictions. Also, it would be most helpful for the enumeration planning staff to know the approximate size of each camp. The success of this method of course depends entirely on the motivation of the local officials. It may be a good policy to give these local officials some payment for the extra work asked of them. In any case, it would be prudent to check the accuracy of at least a small sample of these estimates.

A more costly method of preparing a list of camps is to have the statistical office staff go into the field to locate the camps. This may be done by ground transportation or by aircraft. Although the use of aircraft can be expensive, the cost may actually compare favorably with a ground search because of greater speed, reduced personnel needs, and easier access to remote areas.

The advantages of the camp enumeration method are that it is easier to enumerate the nomads when they are clustered together. Also, since a leader is usually present, it is possible to elicit the leader's help in gaining cooperation from the people.

A disadvantage of this method is that an accurate list of the camps is necessary for success. As with the listing of waterpoints, it may be impossible to prepare such a list because of the great number of camps or a shortage of staff and resources may make visits to all of them impossible.

5.6 *Other approaches*

In addition to the five methods discussed so far, there are other techniques that can be used by themselves or in conjunction with these methods. The first is a variation on the waterpoint method. Instead of stationing enumerators at waterpoints, they are positioned along migration routes to enumerate the nomads passing by. This would be an effective technique provided the migration flow was in one direction and the routes followed were well defined and of a manageable number. Unless virtually the entire nomadic population were in the process of migration, it may be necessary to also use one of the other approaches to enumerate those nomads that are temporarily stationary. Care must be taken to avoid double counting in this situation since the stationary nomads may have been included in the non-nomadic count.

If the nomads live in another country for part of the year or movement cycle, this may present some problems for the data collection operations. If they are not in the country during the period which is best suited for the enumeration of the stationary population, a separate field operation will be required. If the migration routes are well defined, the method mentioned earlier of stationing enumerators along the route at the border can be used. An additional point regarding international migration is that the census publications should provide a description of the flow, including volume and timing.

Another technique that could prove useful in counting the nomads is remote sensing, either aerial photography or satellite imagery. Both were mentioned in the discussions of the waterpoint and camp methodologies as aids in preparing lists. Within limitations, they can also provide population estimates based on the number of tents or the land area covered by the encampment. Aerial photography has the advantage of being easier to interpret because it is in a much larger scale than satellite imagery. It is also easier to find a person skilled in airphoto interpretation. The disadvantages are that it is considerably more expensive to have done than satellite imagery and cannot be done as quickly on a countrywide basis.

To be useable, the air photos and satellite imagery must be relatively recent. How recent depends on the stability of the nomadic population. To be visible on satellite imagery the nomads would have to be clustered in moderate size encampments or be accompanied by herds of animals. If they engage in some form of short term agriculture, they might also be visible. A certain amount of preliminary research is necessary in each country before either of the techniques can be used. The expertise for the use of these techniques may be available at a university or at the government agency responsible for topographic mapping.

In addition to the airphotos or satellite images, a field visit to a small sample of the nomads is necessary. This visit would provide the average number of people per tent or settlement and some data on demographic characteristics not available from the photos/images. This information would form the basis of ratio estimating the characteristics of the entire nomadic population.

5.7 *General considerations*

Regardless of which method is used in a given country, a few general observations may be in order. It may be desirable to use a simplified questionnaire covering only those items of information which the respondents may reasonably be expected to supply. In some instances, the content of the census may be limited to just a few items, for example, name of head, number of males, and number of females in each household. In most cases, however, it should be possible to include such items as age (or age group), relationship to the head of the household, marital status, literacy, and education. Questions such as date of first marriage or hours worked last week would probably be inappropriate and might have an adverse effect on the quality of the basic information desired.

It should be recognized that there are two distinct types of nomads: (a) pure nomads--those who constantly wander about with their herds and never engage in agriculture, and (b) semi-nomads--those who are engaged primarily in pastoral nomadism, but who engage in agriculture during certain seasons of the year. In addition to these two main types, there are also transhumants--herdsmen who move about with their herds, but who are attached to a group which is essentially sedentary.

The method employed in the enumeration of these different types should take into consideration the difference in their respective life styles.

A most important element to consider in the enumeration of nomadic peoples is the seasonal nature and extent of their wanderings. Most nomads have a fairly regular schedule for their migrations, depending on the availability of pasture lands and the special requirements of their livestock. Moreover, most of them have a definite attachment to a particular geographic area--sometimes fairly small and sometimes rather extensive. The timing and extent of their migrations is a most important consideration in the planning and conducting of the census enumeration.

It has been found generally more effective to use census enumerators recruited from among the members of the nomadic tribe than to bring in enumerators from outside. Some countries have used government provided teachers who are already residing with the tribe. A greater degree of cooperation can be expected from the respondent if the enumerator is someone whom he knows and who presumably has the support of the tribal chief than if the enumerator is unknown and unfamiliar to the respondent.

Advance publicity for the census is especially important in the enumeration of nomadic populations. It is sometimes said that nomadic people are more secretive and less cooperative than other segments of the population, but experience would seem to show that this is not necessarily true. If they are made to understand the constructive purposes of the census, and are given no reason to fear taxation, military conscription, or other undesirable results from the census, they are generally found to be no less cooperative than the sedentary population.

Experience on the enumeration of nomadic populations is still quite limited, but has been fairly well documented in a number of United Nations publications. Because the preparatory work for this aspect of the census program includes the determination of an appropriate method as well as the usual items of questionnaire design, instructional materials, publicity, recruitment of field staff, etc., it is important that preparations be started early enough to allow time for testing the methodology as well as the content of the census questionnaire.

6. *EXPATRIATE POPULATION*

Every country has at least a small number of expatriates or foreigners living in it. Because of this, population census programs must decide what data will be collected from them. If the expatriate population is small, whatever plan is chosen will have little effect on the overall census program. Some countries, however, have very large expatriate populations and the way the census chooses to deal with them can greatly affect the data collection operations.

In many cases, the expatriates are there to supplement the local work force. Often, their stay in the country is temporary and they intend to return home after their work is completed. Even if they have no plans to return to their native land, some countries will not allow them to become citizens. There may be a tendency to ignore the expatriates in the census because they are not citizens. This would be a mistake if they compose a large portion of the population. Usually, they have an effect

on the economy and make demands on governmental services; therefore, planners should have at least some basic demographic data on them. Questions on age, sex, marital status, and relationship would provide this basic data. Additional information may be desirable for planning purposes such as country of origin, occupation, industry, and employment status. Of course, the decision on how to deal with the expatriate population should be made early along with the other fundamental decisions.

7. *POLYGAMY*

In some countries, the existence of polygamous unions does not present a data collection problem. All the spouses and their children either live under one roof or in a type of compound settlement. In other countries, polygamous unions are the source of data collection problems. In these countries, spouses and their children may live in different administrative areas, different EA's within an administrative area, or far enough apart so that they cannot be considered one household. The main consideration when dealing with this situation is to avoid double counting the husband or not counting him at all.

Countries that have these types of unions must have an item on their questionnaires to identify when the situation exists. That is, married respondents must be asked if their union is polygamous. Once the situation is identified, there are several courses of action depending on the type of census.

If it is a de facto census, the husband is usually counted where he spent the previous night or in the housing unit where he is found at the time of the enumeration. If the husband is not in any of the housing units, he is still counted where he is found. Caution must be taken that in enumerating the households where the husband is absent, the "spouse" is designated as the "head of the household."

In a de jure census the problem is more complex. If at all possible, the husband should be counted with the first wife. A question must be asked as to the rank of the spouse in each household. The husband should be entered as "head" in the first wife's household; in all other households the spouses are designated as the heads. Some double counting of the males will result if the wives do not know that they are involved in a polygamous union or do not know their rank. However, it should not affect the census data seriously.

Chapter F-3. ONE-PHASE CENSUS

1. INTRODUCTION

As mentioned earlier, there are times when it is not feasible for a country to take a two-phase census. If a country decides to take a one-phase census, there are several options that can be considered.

(1) PLAN 1--All the data, basic items and detailed characteristics, can be collected and tabulated for every person and housing unit in the country. Depending on the number of data items, this procedure can be very expensive and time consuming.

(2) PLAN 2--All the data can be collected for every person and housing unit but only a sample tabulated. This procedure is costly and wasteful. The data collection operation is the most expensive operation for most censuses. The sampling procedures and operational control procedures are much more complicated.

(3) PLAN 3--The basic data can be collected and tabulated for every person and housing unit and the detailed characteristics can be collected and tabulated for only a sample of the population and housing units. This is the least expensive of the options. However, the processing and operational control procedures are more complicated than those for a two-phase census.

If this type of data collection plan is selected, there are basically two choices for the sampling plan; an area sample and a simultaneous, systematic sample.

This chapter discusses some of the procedures involved when Plan 3 is selected. For purposes of discussions, it will be assumed that a 1 in 5 sampling plan has been chosen. That is, for every fifth housing unit data for the detailed characteristics will be collected for all the usual residents and the housing unit. For the transient population and persons residing in collective quarters, data on the detailed characteristics will be collected for every fifth person. Chapter F-4 discusses the sampling details for a one-phase census.

Two styles of questionnaires have been used in censuses of population and housing to collect 100 percent and sample data. The first style is the EA Book in which the questionnaires are bound together for an EA. This was the style used in Popstan's Phase I (see Part C, chapter C-2). This method has the advantage that questionnaires are less likely to be lost or pulled out of serial number order. Another style of questionnaire is called the "booklet" style. Questionnaires for an EA are not bound together. There is one questionnaire for an entire household, and this form is used to collect data concerning both the population characteristics for each person within the household and the housing characteristics. The booklet style allows questionnaires of any length whereas the EA Book style restricts the length to some degree.

Another style that has been seen, but only rarely, is one that uses one unattached questionnaire for each person. Using individual questionnaires presents control and processing problems. If at all possible, individual unattached questionnaires should not be used when a simultaneous sampling is used for data collection.

2. DATA COLLECTION FORMS

A one-phase census requires multiple data collection forms. In this respect, the printing costs will increase. There will also be a need for stricter quality

PREVIOUS PAGE BLANK

control procedures for printing, assembling of forms, and the data collection operation. A minimum of seven data collection forms are usually required.

(1) Listing sheet
(2) Short form
(3) Long form
(4) Short form continuation
(5) Long form continuation
(6) Collective quarters form
(7) Individual census report

2.1 *Listing sheet*

This form will be used to list all housing units and collective quarters within an EA. It is designed to provide both a record of the enumerator's canvass of the area, and a running total of the persons enumerated. The enumerator must be as thorough as possible when completing the listing. The location or description of the housing unit or collective quarters should be stated as precisely as possible in order to facilitate any return visits by an enumerator.

2.2 *Short form*

The short form is one of the basic documents used in the population and housing census. This form is used to enumerate 4 out of 5 housing units. The population information obtained is restricted to name, relationship to head, sex, age, marital status, line number of husband, mother living, and line number of mother. The population items are the same for the EA Book and booklet style forms with the exception of father living and line number of father which is included on the long form. Housing information includes type of living quarters, number of housing units, construction of walls, construction of roof, construction of floors, number of rooms, ownership, amount of rent, vacancy status, and condition of vacant units. The questionnaire is designed to take a minimum of the respondent's time. The population questions on this form are concerned with basic population characteristics and will be asked of all persons enumerated in the census. (See exhibits F-3-1 and F-3-3.)

2.3 *Long form*

The long form will be used for 1 in 5 housing units, giving a 20-percent sample for more detailed characteristics. The basic items on the long form are the same as those on the short form. Sample items that may be asked for the population are place of birth, residency, school attendance, highest grade attended and completed, literacy, number of children ever born, number of children living and dead, and children born alive in last 12 months. A number of economic items can be asked, such as economic activity last month, looking for work, hours worked, occupation and industry, and status of employment (see exhibits F-3-2 and F-3-4).

2.4 *Continuation forms*

The function of these forms is to provide additional space for households with more than 10 persons. The housing items are not needed and, thus, are excluded from the continuation forms. Other than these changes, the forms are the same as the intital long and short form questionnaires. These forms will not be illustrated in this manual.

2.5 *Collective quarters form*

This form is used to enumerate homeless persons and transients, in addition to persons in collective quarters of 10 persons or more. The form has the same population items and format as the regular long forms. However, no specific order is required for listing names of persons in item P-1. Also, there is no housing information obtained on this form. Sample population information will also be

obtained on this questionnaire for persons living in collective quarters (and homeless or transient persons). However, instead of obtaining sample information for 1 in 5 housing units (households), the information will be obtained for 1 in 5 persons. In some places such as mental institutions or military installations, information may have to be transcribed from official records rather than by interview.

The relationship codes will be changed on the questionnaire and all housing items will be eliminated. Shading will be used to indicate for which person listed the sample data are to be collected.

2.6 *Individual Census Report*

The Individual Census Report form is used in situations where a person is not available to be interviewed and no one else can supply the needed information. The form is left at the housing unit and the absent person is asked to complete the form so that the enumerator can pick it up at a later date. This form is designed so information can be entered for one person. The information on the form is restricted to the five basic population items. It will have instructions on how to fill it, and also a space to indicate the best time for the enumerator to return in case additional information is necessary. The form is also designed so it can be folded and sealed in such a way that the questions are not visible, thus maintaining the confidentiality of the information. (See exhibit C-2-1, for an example of this type of form.) The information contained on the completed forms will be transcribed onto the regular form following the procedures set forth in Part C.

2.7 *EA Book versus booklet format*

As discussed earlier in the chapter, either the EA Book or the booklet formats can be used. Both styles have their advantages and disadvantages.

2.71 EA Book format.--If the EA Book format is to be used, the questionnaires must be assembled in a definite order. Using the simultaneous sample design which employs a systematic sample of every fifth household, the pattern of the questionnaires in the EA Book would be, using S for short, L for long, S, S, S, S, L: four shorts, one long, four shorts, one long, etc. Failure to bind the questionnaires in this order would seriously undermine the sampling procedures. (The questionnaires in exhibits F-3-1 and F-3-2 have been designed for this sample design.) Note: stricter quality control procedures to ensure the correct printing and binding of the questionnaires and EA Book must be instituted. As can readily be seen, if a continuation form is necessary due to a large household, the enumeration cannot be continued on the following form. This would destroy the sampling pattern. Some enumerators may even continue a household started on the fourth form (short) on the fifth form (long). In order to avoid this problem, the continuation forms, short and long, are bound in the rear of the EA Book. This may present processing problems, since the key operators will have to find and key the rest of the records for a household before going on to the next sequential questionnaire. However, strict control procedures can usually remedy this situation.

2.72 Booklet format.--The booklet format does not cause the special printing and assembling problems that an EA Book format does (see exhibits F-3-3 and F-3-4). However, even stricter data collection quality control measures are needed to ensure that the enumerators follow the sampling pattern.

The collective quarters questionnaire would probably cause the enumerators the most

problems. Every fifth line on the household roster pages will have to be emphasized so that the enumerator remembers for which persons to collect the sample data.

3. *PROCESSING*

Most of the considerations for processing a two-phase census apply equally to processing a one-phase census. Manual processing and machine processing for either a one- or two-phase census are very similar. These subjects have been covered in Part C, chapters C-5 and C-6 for processing the 100 percent data and in Part D, chapters D-6 and D-8 for the sample data. A rereading of that material is helpful in providing detailed explanations for points that are only briefly described here. The differences between the two types of census and the consequences those differences will have on processing are treated in this section.

3.1 *Field office processing operations*

Data processing in the field offices consists of receiving the questionnaires, checking them in against the listing sheets, checking to make sure the sample design has been followed, checking the entries on the questionnaires for completeness, checking the geographic codes on the forms, and tallying the number of persons and housing units for the preliminary figures to be reviewed and announced locally. The questionnaires, with control documents, are then boxed by designated control unit--EA, crew leader area, subdistrict, etc.--and shipped to the processing center, usually the central office. When a booklet type format is used, the field offices must make sure that all booklets for an EA are present and together.

3.2 *Central office processing operations*

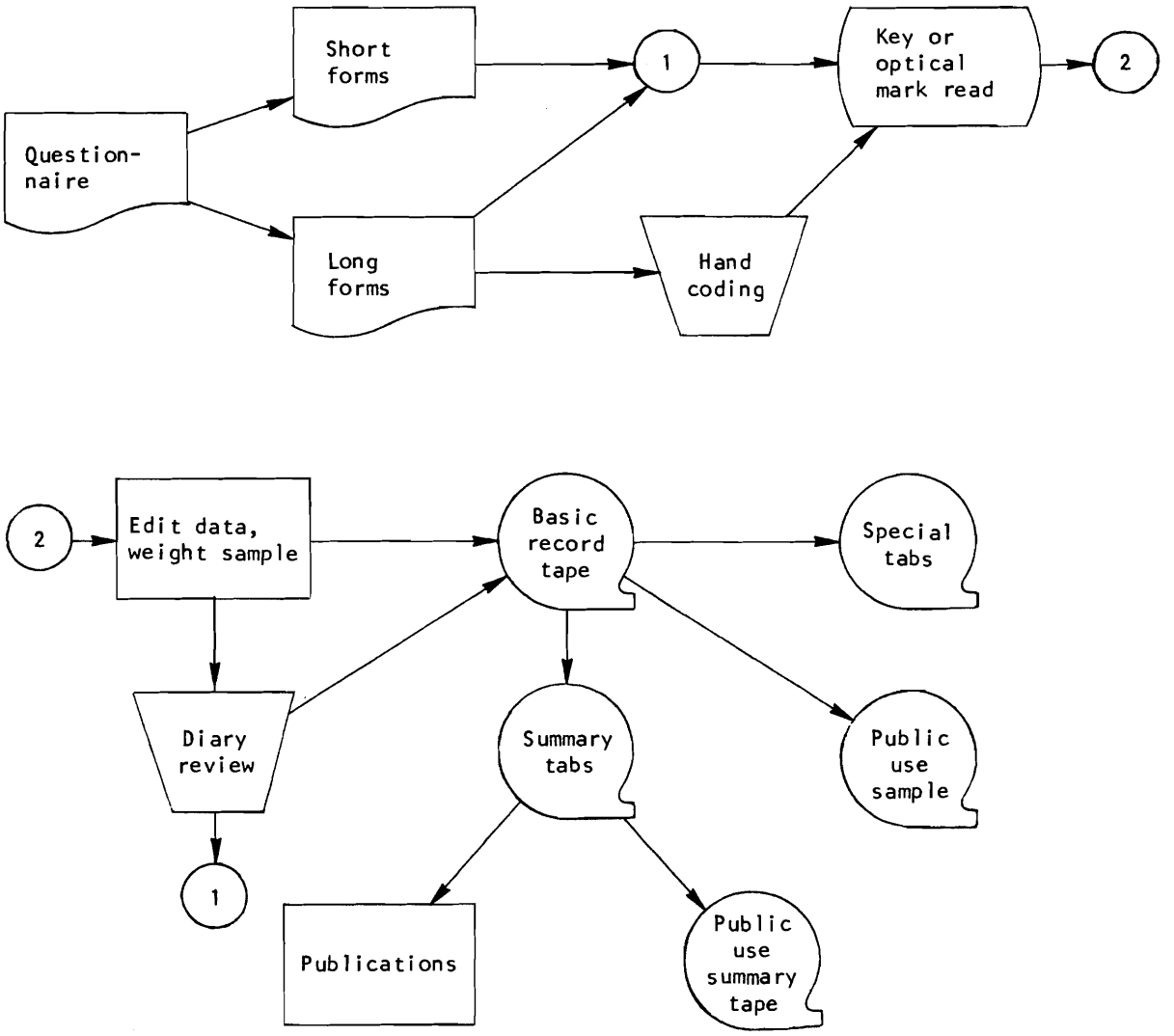
Personnel at the NSO are responsible for receiving, sorting, and storing the boxes and listing sheets. They must clerically code written responses on the sample questionnaires. The questionnaires are keyed (or scanned by optical mark readers) and this keyed data is entered on magnetic tape. The records are then edited. After passing edit, they are stored on the basic record tapes which contain the records for all the housing units and persons in a tabulation area; in Popstan, these are province tapes. The data is tabulated. After the tabulations have been reviewed by the professional staff, reports and computer tapes containing summary statistics are generated. Figure F-3a shows the flow of processing operations for a one-phase census.

3.3 *Long and short forms*

In the central office the processing of the census questionnaires occurs in two separate but overlapping stages. The first involves the processing of the 100 percent questions on all the questionnaires. These are the population and the housing items asked of all persons and housing units found on the short form and the initial pages or columns of the long (sample) form. The second stage involves processing only the information on the long form (used at a sample of housing units--20 percent). The second stage begins when the manual operations and keying of the 100-percent data for each control (work) unit are completed.

Within each processing stage, the tabulation and publication of statistics generally occurs on a flow basis by province. As the data for the geographic levels within a province are tabulated, they undergo review by NSO staff to ensure that the confidentiality of individual information has not been violated, and that the data have been accurately tabulated. Following this review, the appropriate publication tables and reports can be prepared, printed, and distributed to the public.

Figure F-3a. SIMPLIFIED FLOW CHART OF CENSUS PROCESSING OPERATIONS



3.31 Short forms.--The processing of the 100 percent data are given priority because the government requires this information rapidly for local or regional planning. To hold up publication until both 100 percent and sample data are ready would delay publication months and possibly years. The 100-percent data are basic to much subsequent analysis and are used in establishing weighting controls for inflating the sample responses to reflect the total population. The 100 percent processing can be done without the time-consuming hand coding of responses that some of the sample questions require.

3.32 Long forms.--As mentioned above, the basic 100 percent items on the long forms will be processed at the same time as the items on the short forms. Only after the keying operation for the 100 percent items is completed will the sample items be processed. The sample questionnaire shown in exhibit F-3-4 has four questions which are not precoded. These four, on occupation and industry of the respondent, require coding by clerks into the proper categories. A booklet of occupational categories and corresponding codes will be used; a similar manual for industry codes will aid the clerks. These coding

manuals will be based on the codes set forth by the United Nations and International Labor Organization. This is a time consuming task. Since some human judgment is involved, a careful quality control procedure has been established for reviewing the work (see Part D).

After hand coding, the long forms are keyed (or scanned if optical mark readers are being used). Although the 100 percent items on the long form have already been entered, these items will be rekeyed (or rescanned). These items are needed for cross tabulations. This re-entry appears wasteful but the alternative would be to match the sample items on the long form with the previously entered 100 percent items. A matching operation of this magnitude would be very costly and troublesome. Since there are only a few items which need rekeying, the cost of this duplication will be minimal.

3.4 *Computer editing and weighting*

The data for the basic and sample items are edited and tabulated separately. The processing by computer of the sample data can begin as soon as all the records for an area have been keyed and the weights calculated.

3.41 Editing.--Editing includes a check to make sure that the record represents a person and not a keying error (or in case of mark sensing, stray marks on a questionnaire page). Each response or record is checked for and in some instances changed to eliminate inconsistent answers. In addition, an allocation procedure is used when missing or inconsistent information cannot be corrected on the basis of other entries in the same record.

Two sets of edit specifications are required, one for the 100 percent data and one for the sample data. Based on the edit specifications of subject matter specialists, two edit programs will be written to make the data consistent and to impute missing data values.

3.42 Weighting.--Following the edit procedures, the sample data go through a procedure that determines weighting factors or multipliers. A weight is inserted in every population and housing record and used to inflate the sample data for tabulation. In the 20 percent sample where each sample person or household represents approximately five persons or households, the weights should be close to five.

3.43 Diary review.--An important adjunct to the editing procedure is the diary review. A computer editing program examines the data and discovers inconsistencies and missing data. It makes the corrections or allocations and then prints out a diary of these transactions. This diary must be reviewed by professional staff members to determine that the program is functioning as intended and identify any serious problems with the data. Some data may have to be returned for additional processing.

3.5 *Tabulating*

The tabulating of the basic 100 percent items will be done separately from the tabulating of the sample items. The basic items are the basis for the tables of simple distribution. The data from the sample questionnaires are the basis for the detailed cross-tabulations. Two sets of table outlines and specifications must be prepared.

The tabulation programs will have to be written to use the weight to inflate the sample data. Existing software packages are designed to handle this procedure. When considering a package, this capability should be one of the criteria for selection. If the statistical office is considering writing its

own software, very specific instructions on how the weighting is to be done must be given to the programmers. The sampling technicians must work closely with the programmers.

3.6 *Basic record tapes*

Once the above processes are complete, edited data about individuals and housing units, together with associated geographic information, are stored on basic record tapes. All 100 percent and sample tabulations are made from these tapes. Although the basic record tapes do not contain names and addresses, they do have detailed geographic codes and household data that could lead to the identification of individuals. Therefore, these tapes are confidential and may be used only by NSO employees in preparing statistics. From these tapes a sample can be drawn which, after careful screening to ensure confidentiality, can be offered to the public for research purposes. This is the origin of the public use sample.

3.7 *Summary tapes*

Data summaries are prepared on computer tapes from the basic record tapes for many geographic areas, regions, provinces, cities, and hamlets. These internal summary tapes are used to produce sets of tabulations for printed reports and the public use summary tapes. By means of the public use summary tapes, the public can produce any tabulation or cross-tabulation desired even if the NSO has not produced the particular tabulations. For many businesses and industries, this is a popular product of the census.

4. *QUALITY CONTROL OF DATA COLLECTION*

Quality control must be an integrated part of a census. This aspect of census-taking has been stressed in earlier parts of the Popstan case study, although earlier discussions referred to a two-phase census. The basic quality control operations required for a one-phase census are no different from those required for a two-phase census. It is the procedures for each of the census operation that will be different. The quality control procedures most likely to differ the greatest amount are those that are affected by the sampling plan. Most notably, the quality control procedures for the data collection operations.

4.1 *Preparation of EA Books*

The discussion in this section assumes that a sampling plan of 1 in 5 has been selected and the EA Book format is being used. The NSO will enter the geographic codes on the cover and mark the books so that the enumerators do not bias the sample before the books are sent to the field.

4.11 Selection of 20 percent sample.--A random number between one and five will be selected for each EA, and from zero to four of the first questionnaires (PH-2) in the EA Book will be crossed out. The same random number will be used to cross out from zero to four lines on the collective quarters (CQ) forms.

Since the number of HU's (and persons enumerated in CQ's) in an EA is expected to be a random variable, the sample selection procedure should result in a sample of long-form questionnaires in each EA which is approximately 20 percent of the total number of questionnaires (and hence a 20 percent sample of persons will be obtained). Over a whole district, or a province, the proportion should be very close to 20 percent.

4.12 Checking.--Both the identification of the EA and the designation of the questionnaire on which the enumeration is to start are extremely important. The latter step

determines which HU's will be in the sample. The work involved in both steps should be verified 100 percent.

4.2 *Field enumeration*

The responsibilities of the crew leader during the enumeration include controlling the quality of work of the enumerator, insofar as possible. Some of these tasks must be performed on a 100 percent basis and some can be done on a sample basis. In the case of the latter, the sample may be a scientifically selected one for some operations and a spot-check for others.

4.21 Checking EA boundaries.--Immediately after training, the crew leader should check the boundaries of the EA's which are to be assigned to the enumerators. The crew leader can do this more accurately and efficiently than the enumerators.

4.22 Prelisting a sample of HU's in each EA.--Before assigning an EA to an enumerator, the crew leader should indicate on the map the point where the enumerator should start and suggest the starting direction of travel from this point. In addition the crew leader should list about 10 consecutive HU's in each EA starting from an easily identifiable point, to be used as a check on the enumerator's work. In Popstan this short check list will be sent to the DCO for independent checking against the enumerator's list for the same HU's. (In some census situations in other countries it might be feasible to have the crew leader reconcile the two lists and take corrective action during the principal census period.) The sample for the check list need not be a scientific sample; insofar as possible it should include a fairly typical part of the EA. The crew leader should not tell the enumerator the location of the sample listing but should mention that a check is being made.

4.23 Observation of enumeration.--As early as possible after the start of enumeration, the crew leader should visit each enumerator and observe the enumeration of at least one household. The crew leader is asked to check some portions of the enumerator's work entirely (for example, to see that the listing of HU's is being done in the prescribed order, that no lines are skipped on the Listing Sheet or on the Collective Quarters Form, and that no short form or long form questionnaires are skipped). For other portions of the enumerator's work, the crew leader is asked to check a sample of pages, lines, or items. Instructions on how to select the sample, how to record findings, and when to take remedial action will be provided.

4.24 Check on the 20 percent sample.--In a census like that of Popstan, where some information is obtained from all units and some from a sample of persons and HU's it is possible to use sampling theory as an aid in checking the sample. To some extent such checks can serve a control purpose, and to a greater extent they can serve the purposes of evaluation or analysis of the sample. The theory is similar and the difference depends on whether it is feasible to make the checks early enough to affect subsequent field work. This in turn depends partly on the extent to which it is believed practical to entrust such checks to the crew leader. In Popstan, the crew leader is asked to make one simple and precise check to compare the number of sample HU's in each EA with the expected number. The crew leader is also advised on how to make a few other optional checks based on sampling theory. The purpose of these checks is to detect possible sample bias (deviations in the number or kind of sample HU's which are too great to be due to sampling fluctuations).

4.25 Number of housing units.--In each EA, the number of HU's enumerated on the long form should be almost exactly one-fifth of the total enumerated on long and short forms. If the latter number is exactly divisible by 5, the dividend should be exactly the number of sample HU's. If the total number of HU's is not exactly divisible by 5, the number of sample HU's should at least be equal to the whole number in the dividend and it may be one greater than that number, depending on the remainder and on the random number from one to five which was selected as a starting number. Some examples are as follows:

Total HU's	Total HU's ÷ 5		Procedure		Number of HU's in sample should be exactly--
	Whole number	Remainder	Random start	Cross off--	
(1)	(2)	(3)	(4)	(5)	(6)
157	31	2	1	1 short form	31
			2	2 short forms	31
			3	3 short forms	32
			4	4 short forms	32
			5	0 short forms	31
170	34	0	1	1 short form	34
			2	2 short forms	34
			3	3 short forms	34
			4	4 short forms	34
			5	0 short forms	34

These facts are reduced to a simple rule for use by the crew leader. Take 20 percent (one-fifth) of the total number of HU's on short forms and long forms; compare with the number of sample HU's on long forms; if the difference is as much as, or more than, one (1.0), investigate and explain the discrepancy. The DCO would decide whether remedial action should be taken. The crew leader would, however, be better able to train this enumerator or others on future assignments.

4.26 Kind of housing units.--While the number of all sample HU's should check out within a narrow range of error (plus or minus one HU), the numbers of any particular kind of HU are subject to a sampling variability having a wider range. In addition to sampling variability, there may be a tendency of enumerators to reduce the burden of interviewing by putting vacant HU's on listing lines that result in their being enumerated on the long form (or they may do the opposite and enumerate vacant HU's too often on the short form). These mistakes can be effectively controlled (prevented) only by close checks on the order of listing and the order of completing questionnaires. However, the crew leader is advised that if such a tendency is suspected, a rough inspection of a completed or partially completed enumeration book to see whether either tendency is evident must be made. Remedial action for already completed work might be very costly and hence should only be undertaken on instruction from the DCO.

For reasons similar to those affecting vacant HU's, there may be a tendency to obtain too many short forms from large households. This can best be detected by comparing the number of persons enumerated on short forms with the numbers on long forms or by comparing the average size of sample households with that of all households.

4.27 Persons in regular HU's.--If there is no sample bias affecting the size of sample HU's, they should have about the same average number of persons as all HU's. Another way to say this is that, assuming sample HU's are almost exactly one-fifth of all HU's, the total number of persons in sample HU's

in an EA should be about a fifth of the number in all HU's in that EA. A check on these expectations can best be made after the EA has been completely enumerated, and the necessary summary information has been entered on the cover of the enumeration book. Or it can be made as part of the analysis of the sample, after all field work has been completed, with the intention of evaluating rather than of controlling quality.

In Popstan, the crew leader is allowed to decide whether or not it is necessary to make this check, perhaps depending on whether he/she plans to assign additional EA's to the same enumerator. Figure F-3b provides a simple graphic method of making the comparison and interpreting the result.

In order to use the chart the crew leader finds the total number of persons enumerated in regular HU's and the corresponding sample number. (The EA Book cover will contain a section for these entries.) Entering the chart at the total number along the x (horizontal) axis, the crew leader measures vertically up to a point representing the sample number on the y (vertical) axis. The crew leader enters a large dot and identifies it by also entering the EA number near the dot. If the dot for an EA falls between the upper (maximum) and lower (minimum) lines on the chart, there is no reason to suspect a sample bias affecting the size of households. However, if it falls outside these lines, a bias may have occurred, since such a result has only about one chance in 20 of occurring from sampling fluctuations. Most likely, a point falling outside the acceptable range, will fall below the minimum line of the chart, indicating that vacant HU's or HU's with fewer than average persons have been enumerated too often on the long (sample) form. Even if all the EA's in a Crew Leader Area are within the acceptable range but all or nearly all are near the lower line, there is probably a bias. The crew leader usually cannot take remedial action in case a bias appears likely. Without additional field work, a check of the order of the filling of questionnaires against the order of listing can be made, but a check of the order of listing cannot be made without field work. Here again, the checking may only influence the quality of future work and the evaluation of the census.

4.28 Development of crew leader inspection chart to check sample number of persons.--The chart in figure F-3b was based on facts which may be different in countries other than Popstan. Therefore, the sampling statistician should know how the chart was derived and how to adapt it to the situation in another country. The lines in the chart are based on the average number of persons per HU and on the amount of variability in the number of persons per HU. In Popstan, the 1970 Census data show that the average number of persons was about 5.0 per HU and the standard deviation of the number of persons in regular HU's was about 2.6 persons--for both urban and rural areas. Assuming these values have not changed significantly since 1970, they were used to obtain the data in figure F-3c, from which the chart was drawn, as follows:

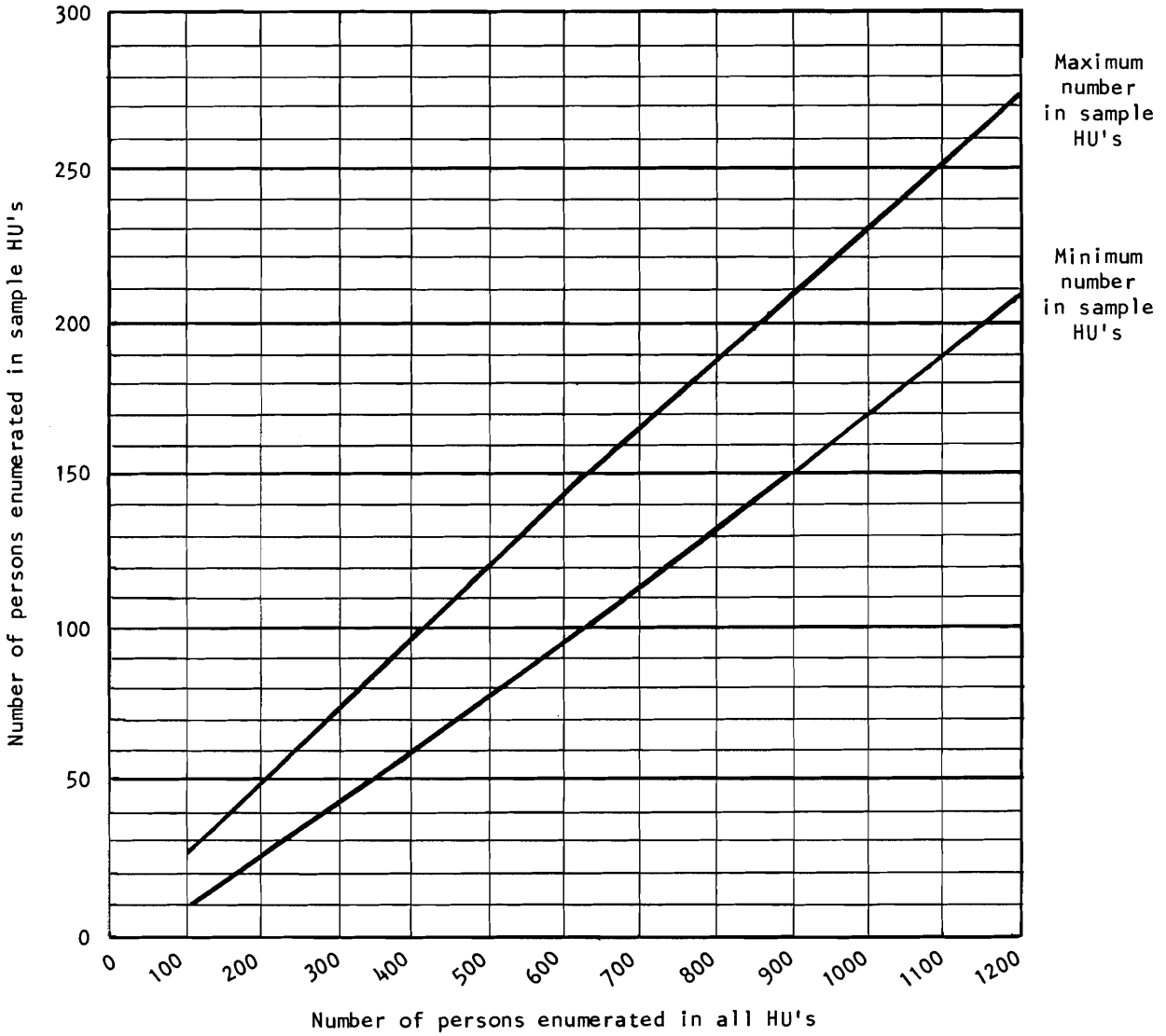
(1) Enter values in column 1, going by steps from the smallest EA that can be expected to be found, to the largest (100 persons to 1200 persons by steps of 100 for Popstan).

(2) Compute the expected number of persons on long forms as one-fifth of the column 1 entry, and place in column 2.

(3) Compute the expected number of HU's on long forms as one-fifth of the column 2 entry, and place in column 3.

Figure F-3b. CREW LEADER INSPECTION CHART FOR ACCEPTANCE RANGE OF NUMBER OF PERSONS ENUMERATED IN SAMPLE HU's

(95 percent probability)



(4) Compute the expected sampling error in the expected number of persons on long forms in two steps:

(a) Expected sampling error in the average number of persons per HU,

$$s_{\bar{y}} = \sqrt{\frac{N-n}{N} \times \frac{s^2}{n}} = \sqrt{.8 \times \frac{(2.6)^2}{n}}$$

$$= (2.6)\frac{(.9)}{\sqrt{n}} = \frac{2.34}{\sqrt{n}}.$$

Example, for $n = 4$, $s_{\bar{y}} = \frac{2.34}{2} = 1.17$.

(b) Expected sampling error in total persons on long form,

$$s_y = ns_{\bar{y}}.$$

Enter in column 4. Example, for $n = 4$ and $s_{\bar{y}} = 1.17$, $s_y = 4.7$.

Note: This result can also be obtained in one step by using

$$s_y = .9 \times s\sqrt{n} = 2.34\left(\sqrt{n}\right).$$

(5) The acceptable range is $y \pm 2s_y$, where y is from column 2, (and s_y is from column 4). This gives a confidence level corresponding to 19 chances out of 20, or a 95 percent probability.

Note: In addition to the sampling error included in this analysis, there is a possible maximum deviation of about 5 in the

Figure F-3c. EXPECTED NUMBER AND ACCEPTABLE RANGE OF NUMBER OF PERSONS ENUMERATED ON LONG FORM IN AN EA

Total number of persons enumerated in HU's (x)	Expected number of persons on long form (1) ÷ 5 (y)	Expected number of HU's on long form (2) ÷ 5 (n)	Expected sampling error in y ($s_y = 2.34\sqrt{n}$)	Acceptable range in number of persons on long form ($y \pm 2s_y$)
(1)	(2)	(3)	(4)	(5)
100	20	4	4.7	11 to 29
200	40	8	6.6	27 to 53
300	60	12	8.1	44 to 76
400	80	16	9.4	61 to 99
500	100	20	10.5	79 to 121
600	120	24	11.5	97 to 143
700	140	28	12.4	115 to 165
800	160	32	13.2	134 to 186
900	180	36	14.0	152 to 208
1000	200	40	14.8	170 to 130
1100	220	44	15.5	189 to 251
1200	240	48	16.2	208 to 272

sample number of persons, resulting from allowing a departure of 1.0 in the expected number of sample HU's. This could cause a very small EA to fall outside the expected range with slightly greater than 5 percent probability.

4.29 Persons in CQ's.--Since individual persons are the unit of sampling in collective quarters, and sample information is obtained from every fifth person on the list of all persons, the sample should contain almost exactly one-fifth of all persons in CQ's. It is unlikely that the order of listing will be distorted to put persons of a particular type on sample lines, since the list will probably be copied from administrative records and there would be no advantage to the enumerator in distorting the sample. However, the crew leader should inspect the list and designation of sample persons for possible misunderstandings or mistakes. The questionnaire summary on the cover of the EA Book should show the number of persons in CQ's on short and long forms (and their total). The long form number should be within 1 unit of exactly one-fifth of the total number.

5. *OPERATIONAL CONTROL OF A ONE-PHASE CENSUS*

Using the principles presented in Part A, chapter A-16, and adapting the procedures and system described in Part B, chapter B-12, the Popstan NSO would have to revise the operational control system and create a new system to accommodate the differences found in a one-phase system. This section will briefly describe the critical points in a one-phase census at which the transfer of documents or data tapes requires careful control to ensure that no data are lost and that the data are transferred on schedule and at a reasonable cost.

As described in the section on questionnaire design, there are two methods that could be used in designing questionnaires. One is to make the questionnaires individually separate from one another. This is called the booklet style. The other method binds the short and long questionnaires for one enumerator assignment into a volume called an EA Book. The basic processing procedures for

the two types of formats are identical. However, the use of a booklet type format makes it necessary to introduce several additional procedures (see figures F-3d and F-3e).

As mentioned earlier, the processing of a one-phase census will be done in two stages. All items on the short forms and the 100 percent items on the long form are keyed at the same time (see figures F-3d and F-3e). When the keying for the 100 percent information has been completed, the forms are sent on to the coding section of the NSO. The computer processing of the 100 percent items will be started while the coding of the sample items is being done. At this time the coding of the industry and occupation questions will take place and be verified. After this, ALL items on the long form, both 100 percent and sample, will be keyed. Keying the 100 percent items a second time is a duplication of effort. However, experience has shown that a matching operation which would be required to avoid rekeying is much more difficult and costly than keying the few 100 percent items a second time.

The operational control procedures and forms must be designed to ensure that the two-stage processing operations for all EA's have been completed. Procedures for controlling data tapes will also be instituted to ensure that the data tapes are passed through the correct edit and tabulation programs. Each tape for a district or province will be labelled as to whether it contains 100 percent or sample data.

5.1 *Processing of booklet style forms*

The processing of booklet style short forms is handled differently than that for long forms, not only because the short forms are not as lengthy and take less time, but because the long forms involve several additional processing operations. Figure F-3d shows the flow chart of the central office operations for both the short forms and long forms in booklet style.

5.11 District office operations.--When the enumerators return the questionnaires from the field, the forms must be sorted into proper EA's and sorted by serial number; the forms may not be in serial order. Experience indicates that this is a common occurrence. The forms should be checked against the listing sheet to ensure that there is a form from each serial number and/or each HU listed. It is important to do this at the district office level as additional field work may be required. The district office will also review the sample using the same procedure as the crew leader (see section 4). At this stage, a pre-edit scan of the questionnaires should be undertaken to ascertain whether the forms can or cannot be processed. Some forms with barely legible entries or which have been ripped or destroyed in some way will have to be pulled and transcribed. A manual edit is then performed to check a few items such as the geo-codes. The geo-codes must be entered on each of the booklets.

5.12 Central office operations.--The NSO will perform the same operations as the district offices. In addition, a check will be made to ensure forms have been received for each EA. Hand count and processibility operations will be performed as discussed in Part D for both the long and short forms.

The operational procedures will be more elaborate than for processing the EA Book type format. Rigid control procedures will be needed to ensure that individual questionnaires for an EA remain together throughout the processing. Procedures must be instituted to ensure that for each EA the long forms are reprocessed and then sorted back into serial number order with the short forms.

Figure F-3d. FLOW CHART OF PROCESSING OPERATIONS FOR ONE-PHASE CENSUS-BOOKLET STYLE FORMS

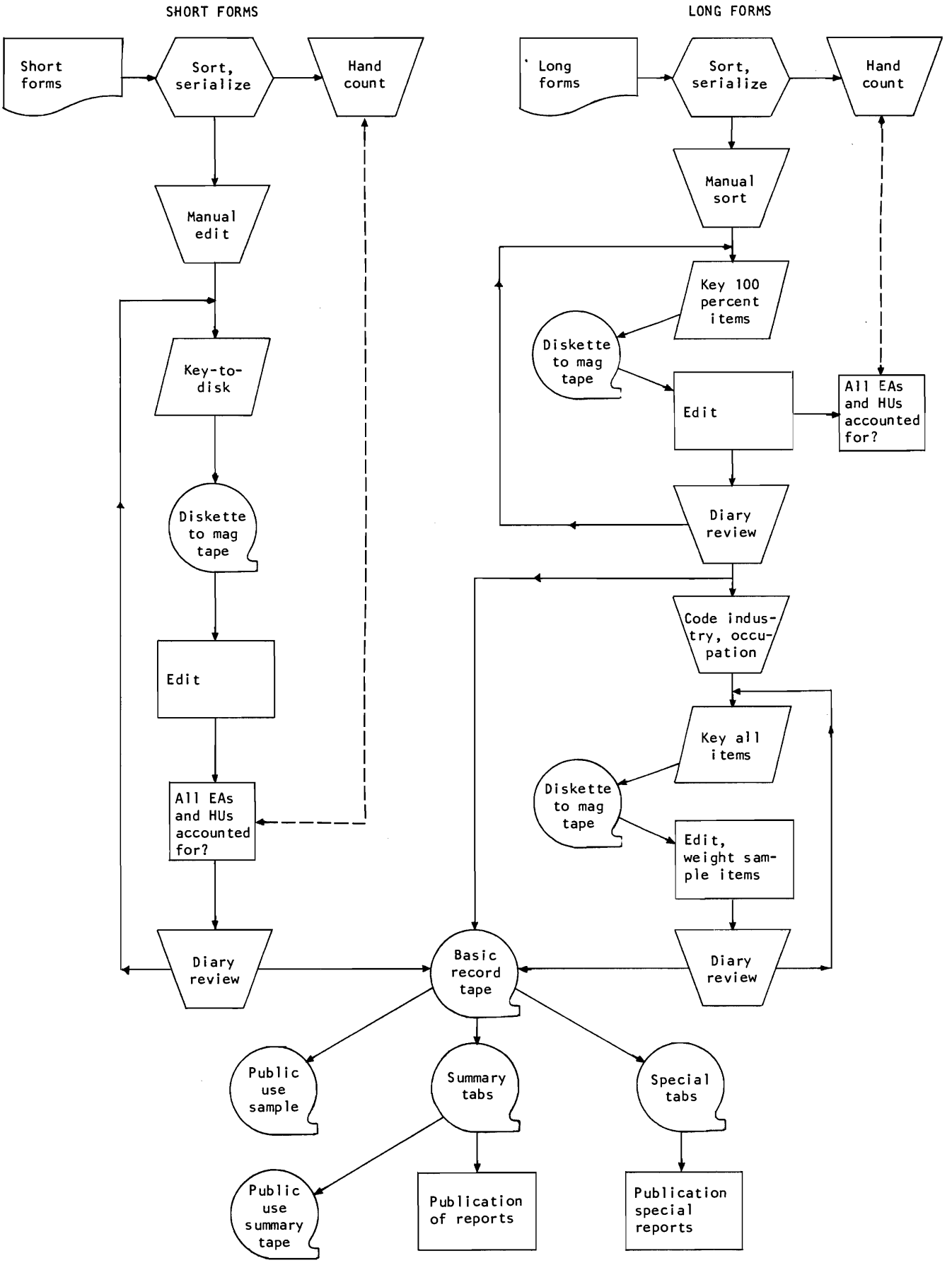
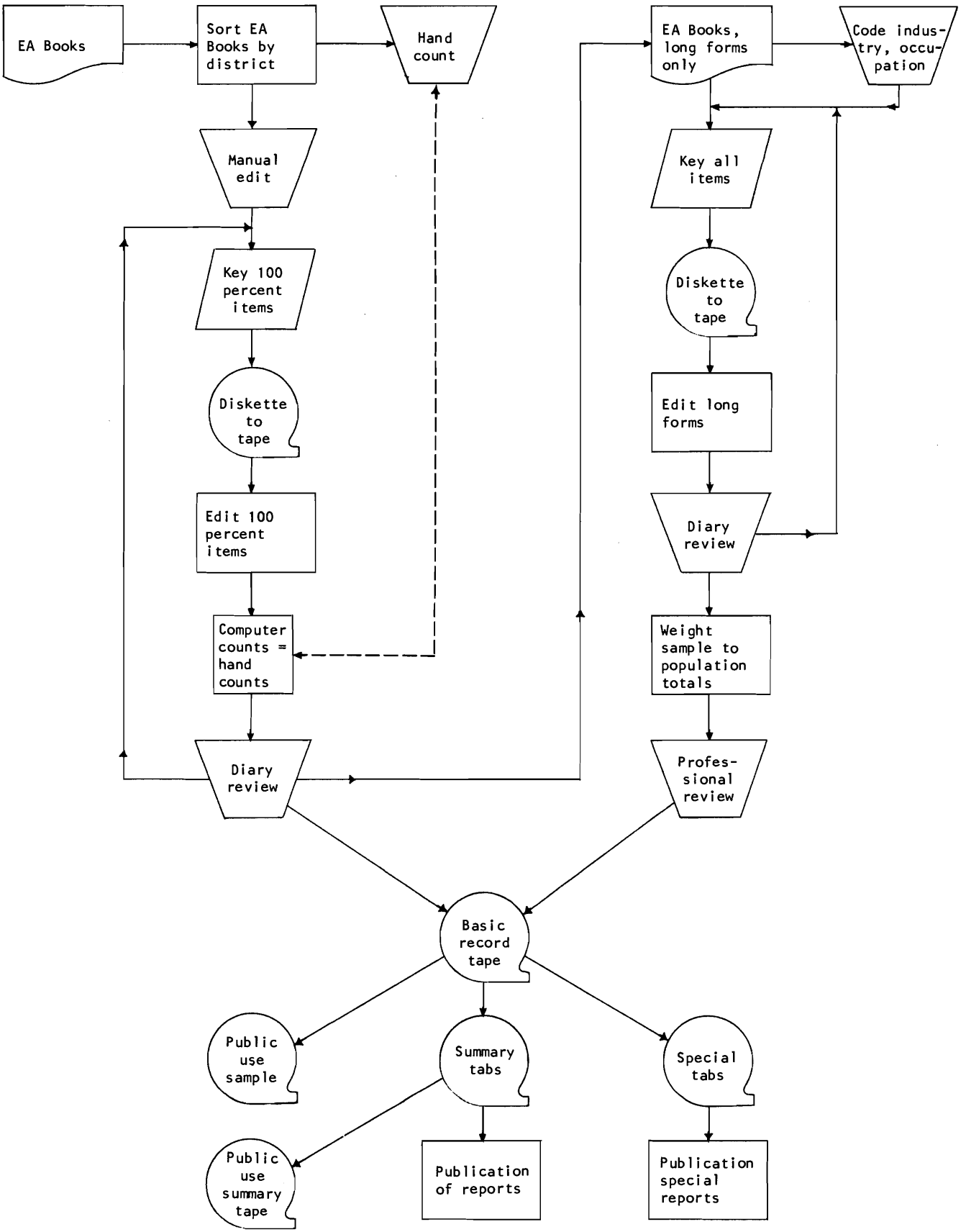


Figure F-3e. FLOW CHART OF PROCESSING OPERATIONS FOR ONE-PHASE CENSUS STYLE FORMS IN EA BOOK

SHORT AND LONG FORMS IN EA BOOK



5.2 *Processing forms in EA Book style*

The processing of short and long forms in the EA Book style is a little different from the processing of booklet style forms. Since the forms in the EA Book are bound together and will not be separated, the control of the forms will be simpler. There is less of a chance that forms can be lost. However, the operational control procedures will have to be designed to make sure that all continuation forms are processed.

6. EVALUATION PLANS

As stated in Part A of this case study, the knowledge and understanding of the magnitude of the possible error in a census is essential to any census program. The rationale and procedures used in the evaluation program planned for Popstan's census will be followed closely in proposing an evaluation plan for a one-phase census. The principal difference, of course, will be that Popstan will combine its evaluation program with the second phase data collection operation which is not being done in the one-phase design. In the one-phase design, the evaluation program will be a separate operation. The goals of this program will be to measure the completeness of the census enumeration by selected areas of the country and to obtain estimates of the characteristics of persons who were missed in the census on a national level.

6.1 *Methodology*

In the alternative one-phase census design proposed here, one out of every five households is asked a longer and more detailed set of questions. The Post Enumeration Survey (PES), however, will not differentiate between the short and long form households. Like the Popstan Phase II, it will be evaluating the coverage of the census but not the content. The estimates of coverage will come from a dual system estimation procedure based on a match between the 100 percent count and a sample survey conducted six months after the census.

The survey will be very similar in design to Popstan's Phase II program. The principal differences will be that the PES will not need as large a sample size as the Phase II program and the Phase II detailed characteristics questions will not be asked.

6.2 *Sample design*

On the matter of sample size for the Post Enumeration Survey there are three basic considerations which must be taken into account. One is the total budget available for the study. In Popstan there is a modest amount of funds available for assessing the quality of the census. A goal, therefore, will be to carry out the PES with the smallest sample size which is large enough to give meaningful results. It is hypothesized that the undercoverage for Popstan will be about 4 percent of the population. It is felt that if the 4 percent undercoverage rate can be measured with a standard error of 1 percentage point (that is, a relative error of 25 percent) then the results will be useful to Popstan for evaluating the quality of the census.

The second consideration is the degree of intraclass correlation within segments for missed persons or households. This phenomenon is a measure of the degree to which missed units of observation (persons or housing units) tend to be grouped together. When the intraclass correlation is high, the sample should consist of many clusters of small size; on the other hand, when the intraclass correlation is low, then fewer clusters can be used which are correspondingly larger. In some studies it has been shown that the correlation

for missed units is about .1 for clusters of around 25-30 housing units. In an optimum sampling scheme, therefore, this value suggests that the clusters should not contain more than about 20 housing units, which is also the same size as the average sector in the Phase II sample design.

The third consideration which has a large effect on the overall size of the sample is the number of separate geographic arreas for which fairly reliable coverage rates are desired. In Popstan there are ten such areas. These are the ten domains used in the Phase II design: the two largest cities and the urban and rural parts of the four regions. The sample size will be roughly the same in each of the ten areas. This is done in order to have approximately the same relative precision in all ten areas. For the country as a whole, the relative precision will be somewhat higher. For Popstan, it is assumed that the undercoverage is fairly uniform throughout the country, so that none of the ten areas will have an undercoverage rate which deviates very much from 4 percent. In countries where the undercoverage would vary by geographic area, the estimated sample size for each area would correspondingly vary. For example, in areas where the undercoverage is low, say one percent, a larger sample size is necessary to measure it with the same relative precision (that is, 25 percent relative error). On the other hand, in areas where the rate is high, such as 10 percent, a relatively small sample can give a reliable estimate.

Taking into account the three key considerations discussed above, the sample size for each of the 10 geographic strata or domains should be on the order of about 1,100 households, or 11,000 nationwide. Mathematically, this works out as follows:

Letting $P = .04$ (for a 4 percent undercoverage rate)

$\sigma(P) = .01$ (for the standard error)

and $f = 1 + \delta(\bar{n} - 1)$, where

δ = intraclass correlation = .1

$\bar{n}$ = average cluster size = 20 households

we then have

$$\sigma^2(P) = \frac{PQ}{n} f$$

$$= \frac{PQ}{n} \left[1 + \delta(\bar{n} - 1)\right].$$

Solving for n, the sample size, we get

$$n = \frac{(.04)(.96)}{(.01)^2} \left[1 + .1(19)\right]$$

$$\doteq 1100$$

With sector (cluster) sizes of 20, this results in a total of 55 sectors to be selected in each area. With this sample, coverage rates can be measured for each of the 10 geographic strata separately; for the nation as a whole, coverage rates could be produced for several broad age groups and by sex and by marital status.

The process used to select the sample for the PES is almost identical to that used for the Phase II sample. The differences result from the smaller sample size being used in the PES and from each domain having the same sample size. This means that the sector sampling rate for each domain will be different in order to allow for the different populations of the domains. However, the other procedures such as assigning district sampling codes and delineating sectors are identical to the Phase II operations.

6.3 *Questionnaires and other forms*

The questionnaire to be used in the one-phase census PES is similar to that used for

the Popstan Phase II survey. Figure F-3f is an illustration of the PES questionnaire. The matching form from Phase II of Popstan can be used without any changes. A copy of this form is shown in the exhibits for reference.

6.4 *Data collection*

The data collection operations for the one-phase census PES will be essentially the same as Popstan's Phase II program. It will be conducted from the Province Census Offices (PCO's) under the direction of the Field Census Branch of the NSO. The number of supervisors and enumerators required will be less since the PES sample size is considerably smaller than that for Phase II. With the reduced sample size, the locations of the sample segments will be much further apart overall than in Phase II. This will add to the average travel time per household. Since the NSO wants to close the field offices as soon as possible, the individual enumerator workload will be reduced from that of the Popstan Phase II plan. In many cases, the enumerators will be assigned two instead of three sectors. The crew leader and supervisory crew leader staff will be reduced from the Phase II levels also while the PCO staff will probably not change. In a few provinces where the workload is smallest, the province technician will take over the duties normally assigned to a supervisory crew leader. The number of enumerators supervised by each crew leader will have to be reduced because the enumerator assignments will be scattered over a larger area. Instead of the average ten enumerators supervised by a Phase II crew leader, only seven will be assigned in the one-phase PES. These considerations will result in the hiring of seven supervisory crew leaders, 40 crew leaders, and 275 enumerators.

The various operations to be conducted before enumeration such as the quick count, recruitment, and training can be carried out as planned for the Popstan Phase II program. The timing would remain unchanged while the number of personnel involved would reflect the smaller PES program.

6.5 *Data processing and matching*

The principal difference between the Popstan Phase II and PES programs relating to data processing is that the PES does not require an occupation and industry coding operation. Since these questions are asked as part of the one-phase census enumeration, the necessary coding is part of that operation. Another difference, which will also reduce the processing time, is that the workload is significantly smaller. Aside from these differences, the Phase II data processing plans are applicable for the PES program.

The matching operation as described in Part D, chapter 7 for Phase II can be used with no change for the PES match. Even though there are additional questions on the Phase II questionnaire, they are not very useful for matching purposes. Therefore, their absence in the PES will not adversely affect the operation.

6.6 *Estimation of coverage*

The procedures established for the Phase II estimation of coverage are applicable to the PES results. One important difference is that the PES will only provide coverage estimates for total and the urban and rural populations in each of the 10 domains. Coverage estimates by age and sex will be provided only for the country as a whole. This is because the sample size in each of the 10 domains is not large enough to provide estimates by characteristics with an acceptable reliability. In Popstan Phase II, the detailed estimates were possible only because of the larger sample size required for the additional population characteristics desired of each domain.

Figure F-3f. POST ENUMERATION SURVEY QUESTIONNAIRE

BOOK_____OF_____

FORM PH-6
(AUGUST 1979)

MINISTRY OF ECONOMY
NATIONAL STATISTICAL OFFICE

POST ENUMERATION SURVEY QUESTIONNAIRE
POPSTAN 1980 CENSUSES OF POPULATION AND HOUSING

CONFIDENTIALITY – This inquiry is required by law. The information is accorded confidential treatment and cannot be used for taxation, investigation, or regulation.

NOTICE TO FINDER This book is the property of the Popstan Government. If found, please return to the nearest post office.

POSTMASTER: Return to the National Statistical Office, Ministry of Economy, St. Regis.

A. Stratum	**D.** Barrio/village	**G.** Sample sector code
B. Province	**E.** Place	**H.** Sector letter
C. District	**F.** EA number	**I.** Serial number

H-1. Address or location

_____

_____

_____

_____

H-2. Type of living quarters
1 Conventional
2 Improvised HU *(makeshift shelter of waste materials, barn, cave, warehouse, etc.)*
3 Mobile HU *(tent, boat, wagon, etc.)*
4 Collective quarters
5 None

H-3. Number of housing units in this house (building)
1 1 unit, detached
2 1 unit, attached
3 2 to 4 units
4 5 to 9 units
5 10 units or more

H-4. Walls – main construction material of outer walls of this house (building)
1 Stone, cement, stucco, brick
2 Metal
3 Wood
4 Bamboo, leaves, reed, mud
5 Other material

H-5. How many rooms are in this housing unit?

_____ Rooms

H-6. Is this unit owned by someone living in it or is it rented?

OCCUPIED
1 Owned or being bought
2 Rented for cash
3 Rent free or other arrangement
4 VACANT

FOR VACANT UNITS

H-7. Vacancy status
1 For rent or sale
2 Not for rent or sale

H-8. Condition of vacant unit
1 Habitable for year-round use
2 Habitable for seasonal use
3 Not habitable

REMINDER *Be sure to get all the information needed for the back page, except when NO in 12.*

Principal respondent	Line No. of principal respondent	Date of interview

Enumerator's name

Notes:

OFFICE USE ONLY

1. MATCHING	Date	Name
a. Matching		
b. Verification		
c. Professional review		

Notes:

Figure F-3f. POST ENUMERATION SURVEY QUESTIONNAIRE--Continued

POST ENUMERATION SURVEY QUESTIONNAIRE –

LINE NUMBER (1)	**What is the name of the head of this household?** **Names of all other persons who usually live here? Be sure to include babies and elderly persons.** **Members of the household away on vacation? On business? In the hospital?** **Any persons who have no usual residence elsewhere?** *List in this order:* *Head* *Spouse of head* *Never married children of head/spouse (by age)* *Ever married children of head/spouse and their families (by age)* *Other relatives of head/spouse* *Nonrelatives* (2)	**What is . . .'s relationship to the head of this household?** *Interpret categories 3 to 7 to mean relationship to head or spouse.* 1 Head; 2 Spouse of head; 3 Son/daughter; 4 Spouse of son/daughter; 5 Grandchild or great-grandchild; 6 Parent; 7 Other relative; 8 Nonrelative (3)	**Sex** (4)	**Age as of last birthday** *If under 1 year, circle 00.* (5)	**Marital Status** 1 Married; 2 Divorced; 3 Separated; 4 Widowed; 5 Never married *If MALE or code 2, 3, 4, or 5, skip to 8.* (6)	**Married Females** *If husband lives in this HU, enter his line number.* *If husband lives elsewhere, circle 00.* (7)
01	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
02	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
03	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
04	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
05	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
06	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
07	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
08	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
09	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
10	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
11	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00
12	Family name / Given name	1 Head; 2 Spouse; 3 Son/dau; 4 SP of son/dau; 5 Grandchild (or great); 6 Parent; 7 Other relative; 8 Nonrelative	1 M; 2 F	00	1 Mar; 2 Div; 3 Sep; 4 Wid; 5 NM	00

Figure F-3f. POST ENUMERATION SURVEY QUESTIONNAIRE--Continued

Continued

Has . . . lived in this housing unit continuously for more than 1 year? (8)	COLUMN A	When did . . . move to this housing unit? (9)	FINAL MATCH STATUS
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	
1 Yes – *Enter "X" in column A.* → *Do not ask 9* 2 No	☐	1 On or before 1 July 1980 ← *Enter "X" in column A.* 2 After 1 July 1980 3 Born since 1 July 1980	

Enumerator Instructions

After listing the household and completing items 2 through 9 ASK the following questions:

(10) I have listed _____ persons.
Name each one.

Have I missed anyone who is living here now?

If YES and person is usual resident or has no usual residence elsewhere, add to item 2.

After listing the last person in the HU, draw a wavy line across columns 1 and 2 to indicate the end of the household listing.

If listing is continued on additional questionnaire, mark X in this circle. → ◯

(11) *Number of boxes marked in column A.*

1 None – *Skip to 13*

2 One or more

(12) You have told me that the following person(s) was/were living here on 1 July 1980.
Give names of persons who have a box marked in column A.

Was anyone else living in this housing unit on 1 July 1980 who is not living here now?

1 Yes – *Go to 17 and complete back page*

2 No – *END INTERVIEW*

(13) You have told me that everyone living in this housing unit now moved here after 1 July 1980. Can you tell me who was living here on 1 July 1980?

1 Yes – *Go to 17 and complete back page*

2 No – *Go to back page and complete 14*

Notes

Figure F-3f. POST ENUMERATION SURVEY QUESTIONNAIRE--Continued

POST ENUMERATION SURVEY QUESTIONNAIRE – Continued
OUT-MOVERS

14. Where can I get information about persons who were living in this housing unit on 1 July 1980?
Neighbors, apartment manager, post office employee, local official, etc. List most likely source first.

1. _____
(Name) *(Address)* *(Type of source)*

2. _____
(Name) *(Address)* *(Type of source)*

3. _____
(Name) *(Address)* *(Type of source)*

Enumerator Instructions: After completing the interview for this HU, investigate the sources listed above and complete 15 to 22.

15. Name of respondent for this page *(If other than principal respondent)*

Line number (16)	17–18. Names of (persons/other persons) who lived in this housing unit on 1 July 1980? *List in same order as shown in question 2 on page 2.* Family name (17)	Given name (18)	CHARACTERISTICS AS OF 1 JULY 1980: Relationship to head (19)	Sex (20)	Age (21)	Marital status (22)	Mover status (23)	Match status (24)
5–01			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–02			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–03			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–04			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–05			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–06			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–07			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–08			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–09			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–10			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–11			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	
5–12			1 Head 2 Spouse 3 Son/dau 4 Other relative 5 Nonrel	1 M 2 F		1 Married 2 Divorced 3 Separated 4 Widowed 5 Never married	4	

17a. If listing continued on additional form, mark (X) in this circle ⟶ ◯

Notes:

Chapter F-4. SAMPLE DESIGN FOR A ONE-PHASE CENSUS

1. *INTRODUCTION*

The sample design for a one-phase census is much different from the design for a two-phase census. Early arguments for a one-phase census were persuasive. The decision shifted back and forth between a one-phase or two-phase census (see chapter A-8, section 5.3) with much of the discussion centering around aspects of sampling associated with each method. The final decision was to use a two-phase census. This has been described in detail in Parts B and D. In this section we present a sample plan for a one-phase census as if a one-phase census had been chosen.

2. *DECISIONS AFFECTING SAMPLE DESIGN*

In Popstan, some data are required for small areas, particularly, counts of the number of inhabitants. These data are urgently needed for barrios, cities, and small areas within large cities where there has been a large in-migration over the past 10 years. Consequently a 100 percent census will be made to obtain counts of the population by basic characteristics such as age, sex, relationship to head, and marital status. Other data are needed only for larger areas such as provinces or metropolitan areas; the census questions for such data will therefore be asked on a sample basis.

When very detailed data are needed for small areas (smaller than provinces), this has a very important bearing not only on the overall sample size but also on the choice between a one-phase or two-phase census. One reason, for example, the two phase census was practical, as described in Parts B and D, is that the level of geographic and substantive detail is fairly broad so that only about 25,000 households were required for sampling and interviewing. It is more practical from an operational point of view to carry out a "small" sample operation of 25,000 households 6 months after the main census than to carry one out that late for several hundred thousand households. On the other hand, when the required tabulations from the sample are so detailed as to call for huge samples, amounting to as much as 10, 20, or 25 percent of the population, the only practical way to conduct the census is in a single phase operation in which both the 100 percent and sample items are collected simultaneously. In the one-phase sample design described in this chapter, it is assumed that data for small administrative areas are desired, necessitating a sample size of several hundred thousand and constituting a large fraction of Popstan's population.

Having decided to obtain some data from all persons and housing units (HU's), and other data from only a sample, we must decide on the methods of selecting the sample and obtaining the sample data. The procedures must be part of an economical total census plan, which also includes the collection of the 100 percent data.

2.1 *Sample collection versus sample tabulation*

Some countries have preferred to collect all the questionnaire items, both the basic and detailed population and housing items, on a 100 percent basis. The basic items are processed on a 100 percent basis and the detailed items on a sample basis. Such a plan offers certain advantages, namely:

(1) It introduces the flexibility of having all the data potentially available for later analysis to meet unforeseen needs.

(2) The enumerators can be trained to deal with only a single collection procedure.

(3) Sampling biases that may occur in the collection stage are avoided.

However, it is expensive to conduct the interviews and therefore wasteful to collect detailed data from everyone when only, for example, one-quarter of it may be processed. Moreover, it is difficult to envision data needs, no matter how detailed, where a 25 percent sample would not suffice. So even unforeseen needs could reasonably be met with a proper sample collection plan, at a great savings in collection costs.

2.2 *Choice of final unit of sampling (HU)*

One of the first decisions on the sample design is whether (a) to select a sample of persons from the complete census list of all persons, or (b) to select a sample of HU's from the census list of all HU's and then to include all persons in these HU's as the sample of persons. This decision depends on the kind of tabulations needed, and the possibility of designing simple forms for selecting the samples in either case. Since we wish to publish tables for persons and also for HU's, we must have a probability sample of both types of units. This could be done by either procedure. If a 20 percent probability sample of persons is selected, for example, this will include an expected 20 percent of all household heads. The HU's of these household heads would be a 20 percent probability sample of all HU's.

Following the first method we could select a sample of persons for sample population data, and then obtain sample HU data whenever the selected person was a household head. This would enable us to prepare the desired tables for population characteristics, and also those for HU characteristics. However, the tabulation plans also include some tables containing information about the persons living in the sample HU's: for example, the frequency distribution of HU's by number of persons per room. Such tables could not be obtained under the first procedure because we would not have data on <u>all</u> the persons living in each sample HU.

Therefore, the second procedure was chosen for Popstan--a sample of HU's with population questions asked of all persons in the selected HU's.

2.3 *Exception: A sample of persons in collective quarters*

A small part of the population of Popstan lives in collective quarters (CQ's) such as institutions, military barracks, college dormitories, rooming houses, etc. In a particular CQ the persons may be quite similar and therefore, an entire CQ would not be a good sampling unit. Furthermore, the CQ's are not included in housing unit tabulations. For these reasons it was decided to select a sample of persons within CQ's for sample population questions. The basic, 100 percent, population questions will be asked of all persons in CQ's and the sample questions will be asked of a sample of these persons. Homeless persons are also listed on the form for CQ's and sampled from the list of such persons.

2.4 *Sampling enumeration areas*

It was decided, largely for analytical reasons, to have a clustered sample of persons, to be obtained by including all persons in selected sample HU's. There are other reasons for considering a possible further clustering of the sample, to be accomplished by selecting groups of HU's rather than individual HU's. One way to do this would be to ask the sample questions in only a sample of EA's.

This would have some definite advantages:

(1) Only a limited number of enumerators would need to be trained in how to ask the additional questions.

(2) Control of the sample is easier if the same questions are asked at every household in the EA.

(3) The editing, coding, and tabulating operations are also simplified.

On the other hand, a sample of EA's will not permit detailed cross-tabulations for small areas such as districts or small urban areas. Because of the large population increase since 1970, the NSO decided that it should provide tabulations, using sample data, for small areas. The total number of EA's in Popstan, approximately 32,000, is not large enough to obtain small-area tabulations using only a sample of EA's.

The desire to make tabulations for small areas resulted in a decision to sample within each EA in the country. This requires a number of additional controls and a different kind of enumeration procedure; it also causes other complications which would not occur if the sample were restricted to selected EA's.

However, the NSO decided that this kind of sample was worth the added effort, supervision, and controls. One result of sampling within each EA is that it will be possible to use the EA's as building blocks for tabulation of sample data, and to provide tables for groups of EA's which do not follow barrio and district lines but which constitute a planning zone or other special-purpose area. These tabulations are not planned as part of the principal census tabulations but they can be made upon request; for example, on the request of a city planning board which may need data for special areas.

Another advantage of sampling within each EA is that the results could be used later in the design of follow-up or intercensal surveys. For example, suppose an internal migration survey was planned after the census. With a sample of persons from each EA it would be possible to stratify the EA's by proportion of recent migrants to make the migration survey sample more efficient. Such a procedure would not be possible if the census sample were restricted to selected EA's.

2.5 *Sampling and data collection plans*

Before proceeding much further with decisions about the sample size and design, a plan must be devised for the collection of the two types of data. As discussed in chapter 3, the 100 percent population and housing items will be obtained from all persons and all HU's, and the sample population and housing items will be obtained from some HU's in every EA. This section discusses several possible plans, and concludes with the one considered best for a one-phase census. Each of these plans begins with the preparation of a list of all HU's (and CQ's) in an EA. Thereafter the procedures differ as described.

2.51 Basic and supplementary questionnaires.--One possible procedure is to have a short basic questionnaire for every HU, with space for all the basic (100 percent) population items for all persons in the HU. Columns could be used for the few basic population items, lines for persons, and additional spaces for the few basic HU items. A supplementary questionnaire could be used at sample HU's. This would have more columns to record the sample population items, and more spaces for the sample HU items. Both forms would have identifying information to permit combining and cross-classifying individual HU data.

For sample HU's, the basic form and the supplementary form could both be filled out on a single visit. This would complicate the work of enumerators, who would be required to use different procedures at different HU's and

to coordinate more papers. To avoid this problem, a variation of this procedure would be to make two visits, using only the short (basic) form at all HU's on the first visit, and the supplementary form at sample HU's on the second visit.

There are obvious complications in both variations of this procedure. While the two-visit method would solve some problems, it would introduce other problems, especially in the identification of selected HU's, and in added time and costs.

2.52 Two or more questionnaires of equal length.--Another proposal which at first seems attractive is to divide the sample questions into two approximately equal groups and to put each group on a different questionnaire. The list of HU's would also be divided at random into two equal parts, and one half of the HU's would be enumerated on one questionnaire and the other half on the other questionnaire. Both questionnaires would contain the 100 percent items in addition to half of the sample items. This procedure has two advantages: (a) it reduces the burden of questions and spreads it more uniformly over all respondents, and (b) it is presumed that this procedure would give the enumerators less incentive to manipulate the sample and therefore to cause bias in the results.

However, there are several objections to this procedure:

(1) It results in a 50 percent sample for each set of sample questions, which may be more than is needed.

(2) It is doubtful that it would really save any time in the field operation, and the office costs might actually be increased because of the need to distinguish two different types of questionnaires.

(3) Dividing the sample questions into two mutually exclusive groups makes it impossible to provide cross-tabulations of the sample questions which are on one questionnaire with those on the other. It might be possible to design a pair of questionnaires which would be satisfactory for the immediate (minimum) census tabulations, as recommended by the United Nations and other international organizations, but some data might later be needed which would require cross-tabulating an item on one form against an item on the other form, and it could not be done, since the two items would not be available for the same units.

(4) The ability to control the sample by such a procedure has never been proven. Enumerators, with the best of intentions, may be tempted to use the questionnaire with information on occupation and industry at households where there are several persons employed, for example, and the questionnaire with inquiries about education at households where there are several persons in school. In both instances, their good intentions would nevertheless bias the census results.

A variation of this plan is to divide the sample into four parts and ask one-fourth of the sample questions at each household. With this plan, there may be less opportunity for enumerator selection bias, but we have no definite evidence of this. The disadvantages listed above would still exist, except that the sample would be a 25 percent sample. The opportunities for cross-tabulations would be greatly decreased and many important tabulations may not be possible, even with the most careful questionnaire design. Also, the use of four different questionnaires would make it more difficult to control, edit, code, and tabulate the data.

2.53 Decision to use short and long forms.--If Popstan were to do a one-phase census, it would use a short form for most of the households, with only a few basic questions (the 100 percent items); the remaining (sample) housing units would be enumerated at the same time on a "long" questionnaire, which includes the basic questions and also the supplementary or sample questions (see exhibits F-3-1 and F-3-2). The questionnaires would be bound

into an EA Book, in the proper order, so that the form to be used (short or long) would be determined by the order in which the enumerator visits the HU's. The effect would be to select a systematic sample of HU's for the long form. It is believed that this procedure would reduce the tendency of some enumerators to use the long form for a large family or one considered to be "typical." There may still be some biases in the sampling procedure, but it is expected that they can be held to a minimum. The effect of possible errors in sample selection can be reduced by means of a ratio estimating method in which the total number of persons enumerated on both the long and short forms will be used to adjust the sample estimates to the population for the whole country.

3. *SAMPLE SIZE*

The decision on the size of the sample is based on a number of factors. The most important is accuracy, followed closely by cost. Other considerations are staff availability and the timing of the census.

In every country the demand for greater accuracy is growing. Accuracy is usually measured in a somewhat backwards fashion by measuring the size of the error. Both sampling and nonsampling errors are present when a sample is used; however, the following discussion shall concentrate only on sampling error. When referring to "sampling error" (or to the precision of a sample result), the reference is being made to how closely the results from a sample can reproduce the results which would be obtained if a complete count were taken, using the same methods of measurement, that is, the same questionaire, interview procedures, types of enumerators, supervision, etc. Thus, the difference between a sample result and the result from a complete count taken under the same conditions is measured by what we will refer to as the precision (or reliability) of the sample result. The difference between the sample result and the true value (sometimes called the population parameter) is called the accuracy of the sample survey. It is the accuracy of a survey which is chiefly of interest, but it is the precision which can be measured in most instances. The object is to design a survey so that the combined effect of the two will be minimized.

Measures of sampling error, or precision, are termed "standard errors" and depend upon the estimator being used, the sample size, and other features of the sample design such as clustering and stratification. In the present discussion, we shall consider estimation of some population total or proportion estimated through data based on a simple random sample of population units. The desired estimate could be the total number of males aged 24-35 years in the population, or the proportion of all housing units in the population which have electricity. In this context, the problem of determining an acceptable sample size for a desired level of precision will be discussed.

3.1 *Important factors*

Often, an estimate of a characteristic of the total number of people in a province or city is important to the decision makers in the country. Perhaps a country would like to know the proportion of persons in the labor force who are unemployed. In this case, the size and design of the sample will be determined by the need to obtain information on this characteristic, allowing no more than a certain level of error.

Many of the estimates made from a sample population and housing census are based on attributes, sometimes called 0,1 variates, rather than magnitudes. Either one has the

characteristic (coded 1) or one does not (coded 0). As an example, the desired statistic may be an estimate of the total number of persons who have completed 8 years or more of schooling, or the number of housing units having inside pumped water. The sampling error of estimates of this kind depend largely on the percentage of all units which have the characteristics and on the size of the sample. Survey planners working together with a statistician familiar with sampling techniques can estimate the size of sample needed to obtain the desired level of reliability for the major characteristics to be estimated by the sample data. The following section discusses the problem of determination of sample size when the major characteristic being estimated is a population proportion and the sampling scheme is simple random sampling.

3.2 *Determination of sample size*

Before discussing the details of the problem of determining sample size, some needed notations should be defined. The notations which will be used to denote population parameters as well as sample values are shown in figure F-4a.

Under the assumption of simple random sampling, the purpose is to estimate P. An unbiased estimator of P under simple random sampling is p. The variance (that is, the

Figure F-4a. NOTATIONS

POPULATION		SAMPLE	
N =	Total number of units in the population	n =	Total number of units in the sample
N_1 =	Number of units in the population possessing the characteristic of interest	n_1 =	Number of units in the sample possessing the characteristic of interest
N_2 =	Number of units in the population not possessing the characteristic of interest	n_2 =	Number of units in the sample not possessing the characteristic of interest
=	$N - N_1$	=	$n - n_1$
P =	Proportion of units in the population possessing the characteristic of interest	p =	Proportion of units in the sample possessing the characteristic of interest
=	N_1/N	=	n_1/n
Q =	Proportion of units in the population not possessing the characteristic of interest	q =	Proportion of units in the sample not possessing the characteristic of interest
=	$1 - P$	=	$1 - p$

square of the standard error) of this estimator is denoted by $V(p)$ and has the following form:

(4.1)
$$V(p) = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right)$$
$$\doteq \frac{PQ}{n}\left(\frac{N-n}{N}\right), \text{ for large } N.$$

The fraction (n/N) is sometimes referred to as the sampling fraction and is denoted by $f = n/N$. Thus,

(4.2)
$$V(p) \doteq (1-f)\frac{PQ}{n}.$$

Equivalently, the standard error $(S.E.)$ of the estimator p is

(4.3)
$$S.E.(p) = \sqrt{V(p)} = \sqrt{(1-f)\frac{PQ}{n}}.$$

In practice, after the data are obtained for the n sample units, the standard error needs to be estimated; or equivalently, the variance needs to be estimated. An estimator of $V(p)$ will be denoted by $v(p)$ and is given for this situation by

(4.4)
$$v(p) = (1-f)\frac{pq}{n}.$$

Since the estimates based on a sample are subject to sampling variability, interval estimates are constructed which take this variability into consideration. These intervals are termed "confidence intervals" and examples would be 95 percent confidence intervals, 90 percent confidence intervals, etc. The level of confidence, expressed as a percent, is a measure of belief (or confidence) that the interval actually contains the parameter being estimated. The interpretation of this confidence level has its setting in the idea of repeated sampling. Basically, a 95 percent confidence interval implies that if all possible samples from a population were independently drawn, and if 95 percent confidence interval estimates were constructed for each sample, then 95 percent of the intervals would contain the parameter being estimated and 5 percent would not. Obviously, one wishes to have a high level of confidence in estimates. Thus, confidence levels are generally high such as 90 percent or 95 percent as opposed to 10 percent or 15 percent. In practice, only one confidence interval is constructed for any particular estimate since only one sample was selected--that is, not all possible samples are selected. Additionally, in practice, if one has a high level of confidence in an interval estimate, one generally assumes that the parameter is indeed in that interval.

A 95 percent confidence interval estimate of P would be

$$p \pm 2\sqrt{(1-f)\frac{PQ}{n}},$$

whereas, a 68 percent confidence interval would be

$$p \pm 1\sqrt{(1-f)\frac{PQ}{n}}.$$

In general a $(1-\alpha) \cdot 100$ percent confidence interval for P would be

(4.5)
$$p \pm Z_{\alpha/2}\sqrt{(1-f)\frac{PQ}{n}},$$

where $Z_{\alpha/2}$ is a number that results in the area under a standard normal curve between $Z_{\alpha/2}$ and $Z_{1-\alpha/2}$ being equal to $1-\alpha$.

In practice, two assurances are desired:

(1) A high level of confidence in the confidence interval.

(2) A confidence interval narrow enough to have a sufficiently reliable estimate.

To see how all of this affects the determination of the sample size, suppose there is a need to estimate some population proportion, P, to within $\pm E$ (for example, to within $\pm$ 1 percentage point) with some fixed level of confidence. The equation to be considered is

(4.6)
$$E = Z_{\alpha/2} \sqrt{(1-f)\frac{PQ}{n}},$$
where $f = n/N$. Solving the above equation for n yields

(4.7)
$$n = \frac{Z^2_{\alpha/2} PQ}{E^2\left(1 + \frac{Z^2_{\alpha/2} PQ}{NE^2}\right)}.$$

This formula becomes simplified if n/N is small (say, $n/N \leq .05$):

(4.8)
$$n \doteq \frac{Z^2_{\alpha/2} PQ}{E^2}.$$

This last formula is frequently used in practice. If $n/N > .05$, then this formula is conservative in the sense that it overestimates to some degree the actual sample size needed. It should be noted, however, that in order to use either of the above two formulas, some idea of the value of P is required. If one has absolutely no idea of the value of P, then setting $P = Q = .5$ results in a conservative estimate for n. The following table gives the required sample size (assuming $n/N \leq .05$) needed to estimate P to within $\pm$ 1 percentage point with stated confidence.

Estimated value of P	Confidence level		
	68 percent ($Z_{\alpha/2} = 1$)	90 percent ($Z_{\alpha/2} = 1.65$)	95 percent ($Z_{\alpha/2} = 2$)
.05 or .95	$n =$ 475	$n \doteq$ 1,294	$n =$ 1,900
.10 or .90	900	2,451	3,600
.15 or .85	1,275	3,472	5,100
.20 or .80	1,600	4,356	6,400
.25 or .75	1,875	5,105	7,500
.30 or .70	2,100	5,718	8,400
.35 or .65	2,275	6,194	9,100
.40 or .60	2,400	6,534	9,600
.45 or .55	2,475	6,739	9,900
.50	2,500	6,807	10,000

Entries in the above table are the required sample sizes assuming $f = n/N$ is small, say $f \leq .05$. If f is not small, an adjustment must be made to obtain the proper sample size, n', still satisfying the reliability requirements. This adjustment is given by

(4.9)
$$n' = \frac{n}{1 + \frac{n}{N}},$$

and results in a sample size smaller than that shown in the table.

As can be seen in the above table, having some information about P (not restricted to $P = .5$), results in a smaller sample size while the reliability requirements are still maintained. However, if it is thought that P is $.01 \leq P \leq .10$, then E (formula 4.6) may be made smaller than if it were thought that P is $.30 \leq P \leq .70$. This thinking is the result of consideration of relative variation, or the relative standard error. In this case, the relative standard error is defined as the ratio of the standard error to the parameter being estimated.

The above remarks on determination of sample size can also be applied to estimating N_1, the total number of units in the population possessing the characteristic of interest. Letting $\hat{N}_1 = Np$, the variance of $\hat{N}_1$ is

(4.10)
$$V(\hat{N}_1) = N^2 V(p)$$
$$\doteq N^2(1-f)\frac{PQ}{n}, \text{ for large } N.$$

Thus, a $(1-\alpha)$ 100 percent confidence interval for N_1 would be

(4.11)
$$Np \pm Z_{\alpha/2} \; N\sqrt{(1-f)\frac{PQ}{n}}.$$

In practice $V(\hat{N}_1)$ would be estimated from the sample data by

(4.12)
$$v(\hat{N}_1) = N \sqrt{(1-f)\frac{pq}{n}}.$$

In order to estimate N_1 to within, say, $\pm D$ with specified confidence, the required sample size would be obtained by setting

(4.13)
$$D = Z_{\alpha/2} \; N\sqrt{(1-f)\frac{PQ}{n}}.$$

Solving for n, results in

(4.14)
$$n = \frac{Z^2_{\alpha/2} \; N^2 PQ}{D^2\left(1 + \frac{Z^2_{\alpha/2} \; NPQ}{D^2}\right)}.$$

As before, if $f \leq .05$, the following can be used

(4.15)
$$n \doteq \frac{Z^2_{\alpha/2} \; N^2 PQ}{D^2}.$$

If the above formula is used first and then it is found that $f > .05$, an adjusted sample size can be obtained as described earlier.

3.3 *Examples*

As an example, suppose it is wished to estimate the proportion, P, of rural housing units in the Samchok Region which have income below a certain specified level. If there are an estimated 479,000 housing units in this area, and if it is desirable to estimate P to within ± .5 percentage points with 95 percent confidence, then P should be $P \leq .40$ for this region. To determine the required sample size, initially consider

$$n = \frac{Z^2_{\alpha/2} \; PQ}{E^2}$$

$$= \frac{(2^2)(.40)(.60)}{(.005)^2}$$

Note: For $P \leq .40$, PQ is a maximum when $P = .4$, $Q = .6$. This gives a conservative estimate of n.

$= 38,400$ housing units. Thus,

$$f = \frac{38,400}{479,000} \doteq .08.$$

The above formula assumes that f is negligible $(f \leq .05)$. Since $f > .05$, an adjusted (and smaller) sample size can be obtained that will satisfy the reliability requirements by computing.

$$n' = \frac{n}{1 + \frac{n}{N}}$$

$$= \frac{38,400}{1 + \frac{38,400}{479,000}}$$

$= 35,550$ housing units.

Therefore, a sample of 35,550 housing units is required to estimate P. P equals the proportion of rural housing units in Samchok Region having income below a certain specified level, to within ± .5 percentage points with 95 percent confidence.

As an additional example, suppose that an estimate of the total number of housing units which contain 7 or more permanent residents is needed. In addition, it is desirable to estimate this figure to within ± 2,500 housing units with 90 percent confidence. If the proportion of the total housing units having 7 or more permanent residents is unknown, then the following equation should be used to estimate the required size of sample:

$$n = \frac{Z^2_{\alpha/2} \; N^2 PQ}{D^2}$$

$$= \frac{(1.65)^2 (479,000)^2 (.5)(.5)}{(2,500)^2}$$

$= 24,987$ housing units.

Thus, approximately 25,000 housing units $\left(f = \frac{25,000}{479,000} \doteq .05\right)$ must be selected in order to satisfy the desired precision requirements. Since $f \doteq .05$, adjustment of this sample size is not required.

It should be noted that there are many characteristics being estimated through data obtained by the long form. In practice, a few major characteristics need to be identified for the purpose of sample size determination. Sample sizes could be worked out separately for each of these characteristics and for each geographic area or domain for

which tabulations are needed. Then, budget permitting, the largest of these sample sizes in a particular domain would be selected as the sample size to be used in practice. Selection of the largest n would allow all the major characteristics to be estimated with the desired individual precisions.

In Popstan, for the one-phase census, it was ultimately decided that a systematic sample of one-fifth of the households would be designated to receive the long form. The 20 percent sample was seen as being large enough to satisfy demands for the most detailed tabulations, even in small domains.

3.4 *Stratified, not simple random sample*

All of the foregoing discussion on the size of sample was based on the assumption that the sample was a simple random sample. But a systematic sample will be used for Popstan and that implies stratification. The introduction of stratification results in a departure from the simple random model upon which the previous formulas are based. By a systematic sample, it is meant that every k^{th} case will be selected as a sampling unit. Because of the implied stratification introduced, the systematic sample has the effect of making the sampling error smaller than it would be for a simple random sample of the same size. Thus, the estimates based on the formula for a standard error in a simple random sample are conservative.

3.5 *Clustered sample*

When dealing with population characteristics (versus housing characteristics), a new complication arises. Although the decision to sample people could have been made, it is ordinarily easier to sample housing units. The sampling unit is the HU and data are collected for all persons in the sample HU. This results in a cluster of persons. The clustering of persons in HU's will have a different effect on error estimates for different characteristics. Clustering increases sampling errors the most for characteristics which are similar or identical for HU members, and it increases the errors the least for characteristics which tend to be different for the HU members. HU clustering will actually decrease sampling errors for characteristics that have more internal variability in the average HU than for characteristics of a group of the same number of persons selected by simple random sampling.

Since clustering affects certain items differently, it may be useful to make an adjustment to the calculated standard errors separately for each item, at least for those items for which clustering is expected to vary from the average. This can be done by applying a multiplying factor to the standard error that was calculated using the simple random sample formula. Figure F-4b shows the kind of values that might be obtained for some population characteristics. This table is included here for illustrative purposes only, and is not based upon an actual calculation. The factors should be used only as an indication of the factors which might be found for a country; actual factors vary from country to country.

Figure F-4b. FACTOR TO BE APPLIED TO STANDARD ERRORS

Characteristic	Factor
Age	0.8
Province of birth	1.2
Language spoken in home	1.4
Duration of residence	1.8
School enrollment	0.8
Years of school completed	1.0
Marital status	1.0
Employment status	1.1
Occupation	1.0
Industry	1.0
Hours worked	1.1

(etc.)

Sometimes an experienced statistician, familiar with the country and its sample design, can make a rough estimate of the factors to be used. If it is desirable to check the reliability of such factors, or to compute adjustment factors for a country not having any previous experience in this field, some research is needed. Since the adjustment factor is a prediction of how a precise error estimate differs from that of the simple random sample formula, it can be checked by computing the errors (for a few items) by a precise method and also by the simple random sampling formula, using the same sample data and comparing the results. Such a comparison should be made for a number of HU and population characteristics, and for areas of various sizes.

4. ERROR ESTIMATES BASED ON ANALYSIS OF DATA

Rough estimates of sampling error were discussed earlier (section 3). That the reliability of the rough estimating formulas requires checking by more precise methods was also mentioned. In fact, the adjustment factors used in the rough methods would probably be based on some analysis of a more precise kind.

4.1 *Need for more exact error estimates*

Since the 20 percent enumeration sample for Popstan is a probability sample, it is possible to compute precise estimates of the sampling errors from the sample data themselves. Computations of sampling error should be made, particularly for attributes which are very variable; or highly correlated within a household, such as nativity or language spoken in the home; and also for some magnitude items such as income. The computation of sampling errors should be considered an important part of the census and should have adequate budgeting.

4.2 *Random group method*

Variances can be computed by a relatively easy method which does not require obtaining totals and sums of squares for every element (household or person) in the population. This is the random group method; the sampled HU's within an area are assigned to groups at random and these groups are treated as if they were separate samples for estimating the area total. A variance is then computed between the set of estimated totals and for the combined estimate. These variances can then be added for a number of areas; hence, obtaining an estimate of the variance for a larger area or for the whole country. The variances are finally converted to standard errors of the estimates. The specific plan for Popstan is described below. Figure F-4c (Worksheet: Computation of Province Area Variances for the Enumeration Sample) illustrates the required computations.

4.3 *Assignment of random group codes within provinces*

There are 15 provinces and two metropolitan areas in Popstan. Within each of these 17 areas (obtained by separating the metropolitan areas from the remainders of the provinces which contain them), each sample HU and the residents, and each sample person living in collective quarters[1] (CQ's), will be assigned at random to one of a number of "random groups." The number of random groups should be large enough to give reasonably good estimates of variance but small enough to make the resulting subsamples reasonably large. A number such as 10 or 20 is convenient to manipulate. It was decided to use 20 random groups for this purpose in Popstan. This gives 19 degrees of freedom for the

[1] *The assignment to random group in CQ's is made for individual persons, rather than for all persons in one CQ.*

Figure F-4c. WORKSHEET: COMPUTATION OF PROVINCE AREA VARIANCES FOR THE ENUMERATION SAMPLE, FORM PH-409

(The worksheet illustrated here provides for computing variances of aggregates for two characteristics and for computing the ratio of the aggregate of the second to that of the first characteristic. The results in columns 3 and 5 of line E below the table are the variances corresponding to equation 4.18. The result in column 6 is the covariance of equation 4.23. Computations outside the double line (those in column 6 and on lines H to K) are made only if it is wished to estimate the ratio of the second to the first characteristic)

FORM PH-409
POPSTAN

WORKSHEET

COMPUTATION OF PROVINCE AREA VARIANCES FOR THE ENUMERATION SAMPLE

(a) Sheet 2 of 34

(b) Province ARTESIA Code 03 (c) ☐ Urban ☒ Rural

(d) First characteristic (y) Children 5 to 11 years

(e) Second characteristic (x) Children 5 to 11 attending school

(f) Prepared by E. Anowar Date 3 February, 1981

(g) Verified by M. Basher Date 3 March, 1981

Random group	First characteristic (y)		Second characteristic (x)		Covariance (column 2 X column 4)
	Weighted sample total: y_j	y_j^2	Weighted sample total: x_j	x_j^2	
(1)	(2)	(3)	(4)	(5)	(6)
1	3,214	10,329,796	2,402	5,769,604	7,720,028
2	3,532	12,475,024	2,752	7,573,504	9,720,064
3	4,046	16,370,116	3,016	9,096,256	12,202,736
4	4,183	17,497,489	3,125	9,765,625	13,071,875
5	3,275	10,725,625	2,714	7,365,796	8,888,350
6	4,195	17,598,025	3,087	9,529,569	12,949,965
7	3,420	11,696,400	2,809	7,890,481	9,606,780
14	3,552	12,616,704	2,905	8,439,025	10,318,560
15	4,518	20,412,324	3,109	9,665,881	14,046,462
16	3,264	10,653,696	2,780	7,728,400	9,073,920
17	3,733	13,935,289	2,872	8,248,384	10,721,176
18	4,524	20,466,576	3,140	9,859,600	14,205,360
19	4,042	16,337,764	2,983	8,898,289	12,057,286
20	4,415	19,492,225	3,316	10,995,856	14,640,140
A. Total	y" = 77,542	Σy_j^2 = 305,092,146	x" = 58,385	Σx_j^2 = 171,464,505	$\Sigma y_j x_j$ = 228,293,882
B.		$20\ \Sigma y_j^2$ = 6,101,842,920		$20\ \Sigma x_j^2$ = 3,429,290,100	$20\ \Sigma y_j x_j$ = 4,565,877,640
C.		$(y")^2$ = 6,012,761,764		$(x")^2$ = 3,408,808,225	(y"x") = 4,527,289,670
D.		B - C = 89,081,156		B - C = 20,481,875	B - C = 38,587,970
E.	$s^2(y")$ = .0421 D →	3,750,317	$s^2(x")$ = .0421 D →	862,287	Cov (x", y") = .0421 D = 1,624,554
F.	s (y") = $\sqrt{E}$ →	1,937	s(x") = $\sqrt{E}$ →	929	
G.	v(y") = $\frac{s(y")}{y"}$ →	= .02497 = 2.5 %	v(x") = $\frac{s(x")}{x"}$ →	= .0159 = 1.6 %	

H. $r = \frac{x"}{y"}$ = .7529 ; $r^2s^2(y")$ = 2,126,055 ; 2r cov(x", y") = 2,446,253

I. $s^2(r) = \frac{s^2(x") + r^2s^2(y") - 2r\ cov(x", y")}{(y")^2} = \frac{542,089}{6,012,761,764} = .0000902$

J. s(r) = .0095 ; $v(r) = \frac{s(r)}{r}$ = .0126 = 1.3% (relative error as percent of estimate)

K. 100r = 75.3 (%); 100 s(r) = 0.95 (percentage points)

Item d and e--There is no significance in the order of listing of the two characteristics unless it is necessary to compute the ratio of one to the other. In that case, put numerator characteristic as "second" and denominator charateristic "first."
Column 1--Random group codes are explained in the text.
Columns 2 and 4--For attributes, their weighted sample totals are sums of the integral weights computed in the ratio adjustment procedure; for magnitudes, the weighted sample totals are sums of the products obtained by multiplying the integral weight by the magnitude for each individual unit (person or HU).
Columns 3, 5, and 6--These need not be filled in line by line: only the total sums of squares (columns 3 and 5), and of cross products (column 6) are needed.
Lines A to G--Perform the indicated computations. The factor .0421 on line E is the product (1-f) ÷ 19, obtained by using f=.20 for the sampling fraction.
Lines H, I, J, and K--See note above this figure and text.

estimate of variance for each area. The resulting subsamples will contain $\frac{1}{20}$ of the original 20 percent enumeration sample, which will be an expected minimum of 230 HU's and 1,350 persons in the smallest province area. The 20 groups are designated 1, 2, ... , 20. In order to associate persons with HU's, all of the persons in a HU have the same random group assignment as the head of the HU.

It is usually convenient to assign the codes after the data records have been separated into district urban and rural parts for the ratio estimates. At that time, it is no longer necessary to keep the urban and rural parts of individual districts next to each other in the file.

For the purpose of forming random groups within province areas (that is, within the urban and rural parts of provinces separately), the records for each province must be rearranged by first separating all urban district parts from all rural district parts.

The sample was selected within each EA separately; consequently, sorting of EA's by city size, central city, or any other characteristics associated with EA's may be introduced before the random group numbers are assigned. Sample persons within CQ's should be in a separate part of the file, since these persons were separately sampled.

Once the desired geographic order of EA's within districts has been established, the sample records for each province area of Popstan should be put in the following order:

Districts in numerical order

EA's in the desired geographical order[2]
  HU serial number within EA
  CQ line number within EA

Remark: Additional stratification (for example by size of HU) should not be used since it was not part of the sample design; its use would introduce a spurious reduction in the estimate of the sampling errors, and make the sampling estimates seem more reliable than they are.

The assignment of random group numbers can be made by starting with a new random number from 1 to 20 in each province area, continuing until the number "20" is reached, and then assigning "1" to the next HU or CQ record listed. For example, if random number 07 is selected as the starting number for the urban part of Artesia Province, the first record in the urban part of the first district would belong to group "7," the second to group "8," and the third to group "9," the fourth to group "10," etc. The 21st record would also be in group "7," and the pattern would repeat thereafter. The numbering is continued throughout all districts in the province.

4.4 *Estimates of aggregates for random groups*

Let x_j be the total for a particular characteristic in the jth random group for the urban or rural part of a particular province or metropolitan area, hereafter called a "province area." There are 34 province areas. For a 0,1 variate (an attribute),[3] this total is the sum of the weights for the records containing this characteristic (for example, the sum of the weights for children aged 5 to 11 attending school).

The estimate for the province area is the sum of the 20 random group totals:

(4.16)
$$x'' = \sum_{j=1}^{20} x_j .$$

[2] *The desired geographic order will frequently be the same as the serial number order of the EA's. (Serial numbers of EA's are usually assigned with publication requirements in mind.)*

[3] *For other variates (such as income, or years of school completed) the total is the weighted sum of the individual reports, using the same weights as of 0,1 variates.*

Twenty separate estimates for the characteristic could be made, by multiplying the total for each random group by 20:

$$x''_j = 20x_j . \tag{4.17}$$

Each x''_j would give an estimate of the total for the province area based on $\frac{1}{20}$ of the records for that particular area. These 20 estimates will vary; using them an estimate of the sampling error of the estimate x'' in equation (4.16) can be calculated. Actually, a shorter method based on the values of x_j instead of those of x''_j is used.

4.5 *Variance of estimates of aggregation*

The formula for the computation of the variance of a simple random sample by the random group method is given in standard textbooks.[4] The resulting estimate of the variance will have fewer degrees of freedom, and be less precise than if each record were used individually in computing the variance. However, this formula is obviously much easier to use than the alternative formula which uses the individual records.

The estimate of the variance for the province area estimate x'' is given by

$$s^2(x'') = (1-f) \times \left[\frac{20\sum_j^{20} x_j^2 - (x'')^2}{19} \right] \tag{4.18}$$

where $f(=.20)$ is the sampling fraction, and $1-f=.80$. (This computation is performed in lines A to E at the bottom of the worksheet illustrated in figure F-4c.)

In order to obtain the variance of the estimated characteristic for a combined group of province areas, a subscript "h" is added to indicate the particular province area, and sum for the group of province areas. For example, to obtain results for the entire country, all 34 areas are added:

$$s^2(x'') = \sum_{h-1}^{34} s^2\left(x''_h\right). \tag{4.19}$$

For the combined urban and rural parts of a single province, these two areas are added. For subareas of provinces (for example, a specific village or barrio), the variances can also be calculated using these same random group formulas. This is done by simply defining the characteristic x_j as the jth random group total for the characteristic in the subarea rather than in the urban or rural part of the province.

The formulas for the random group method are based on the assumption that there is a simple random sample within each province area. Actually, it is a systematic sample. For some characteristics, the systematic selection of the sample may result in an estimate which has a smaller error than that which would have come from a simple random sample. Therefore, the use of these formulas may lead to a slight overestimate of the true variance.

4.6 *Other measures of variability*

Several common measures of variability can be derived from the variance; the choice depends upon how one wishes to talk about the reliability of the estimate. The two following measures are more appropriate than the variance for reporting errors because they are easier to understand. Also, they can be interpreted in terms of confidence intervals. (See figure F-4c, lines F and G.)

4.61 Estimated standard error.--The estimated standard error of an estimate is the square root of the estimated variance; it is frequently indicated by the letter s, with a subscript or parenthetical expression to

---

[4] *See, for example, Sample Survey Methods and Theory, Volume I: Methods and Applications, by M. H. Hansen, W. N. Horwitz, and W. G. Madow, New York, Wiley and Sons, 1953, p. 440.*

identify the kind of estimate. The following equation shows the use of the subscript.

(4.20) $$s(x'') = \sqrt{s^2(x'')}\,.$$

The 68 percent confidence interval is obtained by taking $x'' \pm s(x'')$.

4.62 Estimated relative standard error.--The estimated relative standard error of an estimate (also called the estimated "coefficient of variation of the estimate") is the estimated standard error divided by the estimate itself. It is sometimes indicated by the symbol *CV*. The letter v followed by a parenthetical expression is used in the following equation.

(4.21) $$v(x'') = \frac{s(x'')}{x''}.$$

The 68 percent confidence interval (expressed as a percentage error) is obtained by taking $v(x'') \times 100$.

4.7 *Standard errors of differences*

For differences between sample estimates and other sample estimates or known totals, two situations are distinguished:

(1) For the difference between the sample figure and one based on a complete count (for example, the comparison of a 1980 sample estimate of educational attainment with the 1970 complete count for the same statistic) the standard error is identical with the standard error of the sample figure alone (the 1980 estimate).

(2) For a difference between two sample figures, the standard error is approximately the square root of the sum of the squares of the standard errors of the two estimates considered separately. This formula is quite accurate for the difference between estimates of the same characteristics in two different areas, and for the difference between separate and uncorrelated characteristics in the same area. If, however, there is high positive correlation between the two characteristics, the formula will overestimate the true standard error.

4.8 *Estimates of percentages and ratios, and their variances*

Frequently there is a special interest in a percentage or ratio where both the numerator and denominator are random variables; for example, the proportion of children 5 to 11 who are attending school. The variance formula will reflect the variability in both characteristics, and the correlation between them. (See figure F-4c, column 6 and lines H to K.)

For an estimate of a ratio, $r = \frac{x''}{y''}$, the estimated variance is given approximately by

(4.22) $$s^2(r) = \frac{s^2(x'') + r^2 s^2(y'') - 2r\,cov(x''y'')}{(y'')^2}$$

where

x'', y'' are estimated totals for the area (province, group of provinces, or whole country).

$s^2(y'')$ is obtained from equation (4.18) by substituting y and y'' for x and x'' respectively.

The estimate of the covariance within a province area is

(4.23) $$cov(x'',y'') = (1-f) \times \left[\frac{20\sum_j^{20} x_j y_j - x''y''}{19}\right].$$

Using another subscript h and summing as in equation (4.19) will give an estimate of the covariance for a larger area. The standard error of r is the square root of the variance; the corresponding percentage, $100r$, has a standard error which is 100 times that for r itself.

Remark: The standard error, coefficient of variation (or relative standard error), and the variance are all measures of the variability of the sample estimate. They are usually computed to three significant figures, at most, and may be good estimates to no more than two significant figures.

4.9 *Combining province area variances for larger areas*

While the variances and covariances of aggregates $[s^2(x''), s^2(y''),$ and $cov(x'', y'')]$ may be added to obtain corresponding results for combined areas, this is not true of the variance of a ratio $[s^2(r)]$. The variance of a ratio for a combination of province areas must be recomputed using the results obtained from simple addition of the aggregates $(x''$ and $y'')$, and of the variances and covariances of the aggregates. Figure F-4d (Worksheet: Combining Variances for the Enumeration Sample) illustrates a form for combining variances for both aggregates and ratios, and for computing related measures of sampling error.

4.10 *Obtaining variances as a by-product of census tabulations*

The precise variance computations described above can be carried out for only a limited number of the most important characteristics for which sample data are to be published; usually it is not feasible to compute variances for more than about 200 of the important characteristics, even with computers. The number is limited by the size of the computer, and the time required to obtain the 20 random group totals, instead of a single total for each tabulation area (province, municipality, village, etc.).

After records are edited, the records are sorted into the order required for the various tabulation runs, and the census tables are produced. If the sort includes the random group code as the lowest level, data for computation of variances can be obtained as a by-product of the tabulation. Following are two possible methods of doing this.

The first method uses <u>random group totals</u>. During the production of each table the items for which variances are required can be summed to the random group level, and the random group total written on an auxiliary tape (with the proper identification of the province, table, random group, etc.). This method requires tabulating these selected items in the regular tally matrix, and using another smaller matrix for the variance summaries for the selected characteristics. Variances can then be computed from the auxiliary tape using a program written for the purpose.

The second method uses <u>cumulative random group totals</u>. This procedure, which does not require the use of additional space for tallying the random group totals, tabulates in the regular tally matrix and writes the cumulative totals in the tally matrix on an auxiliary tape at the end of each random group; the writing may be of the whole matrix, or of a few of the cells selected from the tally matrix.

This procedure will thus result in a tape containing cumulative totals. The first output written on the tape will be for random group "1"; the second will be the sum of groups "1" and "2"; and the third will be the sum of groups "1," "2," and "3," etc. The first step in the computation of the variances is to subtract successive totals in order to "uncumulate" the data and to obtain the individual random group totals. After this, the variance computation is the same as described above.

The advantages of producing variances as a by-product of tabulation are:

(1) The variance tabulations are produced at approximately the same time as the sample estimates for the census.

(2) The cost of producing sampling errors is usually less if they are developed at the same time as the tabulation for estimates.

(3) The totals obtained in the variance tabulations will agree exactly with the census figures which are to be published.

Figure F-4d. WORKSHEET: COMBINING VARIANCES FOR THE ENUMERATION SAMPLE, FORM PH-410

(The worksheet has 17 listing lines. Only two are needed for obtaining individual province results, as in the illustration; all 17 lines would be used to obtain national urban or rural results, and two lines for combining the latter to obtain national results for urban and rural areas combined)

FORM PH-410
POPSTAN

WORKSHEET

COMBINING VARIANCES
FOR THE ENUMERATION SAMPLE

Sheet 1 of 20

(a) Combined area Artesia Province (urban and rural)

(b) First characteristic (y) Children 5 to 11

(c) Second characteristic (x) Children 5 to 11 attending school

(d) Prepared by Alona Mozur — Date 16 March, 1981

(e) Verified by Silva Costa — Date 26 March, 1981

Province areas to be combined (1)	First characteristic (y): y" (2)	First characteristic (y): $s^2(y")$ (3)	Second characteristic (x): x" (4)	Second characteristic (x): $s^2(x")$ (5)	Covariance cov(x",y") (6)
03 U	45,142	2,202,930	35,346	533,725	725,724
03 R	77,542	3,750,317	58,385	862,287	1,624,554
A. Total	$y_c"$ = 122,684	$s^2(y_c")$ = 5,953,247	$x_c"$ = 93,731	$s^2(x_c")$ = 1,396,012	Cov $(x_c" y_c")$ = 2,350,278
B.	$(y_c")^2$ = 15,051,363,856	$s(y_c")$ = 2,440		$s(x_c")$ = 1,182	
C.	$v(y_c") = \frac{s(y_c")}{y_c"}$ →	= .0198 = 2.0 %	$v(x_c") = \frac{s(x_c")}{x_c"}$ →	= .01261 = 1.3 %	

D. $r_c = \frac{x_c"}{y_c"}$ = .7640 , $r_c^2 s^2(y_c")$ = 3,474,910 , $2r_c \text{ cov}(x_c" y_c")$ = 3,591,225

E. $s^2(r_c) = \frac{s^2(x_c") + r_c^2 s^2(y_c") - 2r_c \text{ cov}(x_c" y_c")}{(y_c")^2}$ = 1,279,697 / 15,051,363,856 = .000085

F. $s(r_c)$ = .0092 , $v(r_c) = \frac{s(r_c)}{r_c}$ = .0120 = 1.2 % (relative error as percent of estimate)

G. $100r_c$ = 76.4 (%) , $100s(r_c)$ = 0.9 (percentage points).

INSTRUCTIONS

Enter data on listing lines from Worksheet for Computation of Province Area Variance (Form Ph-409).

Perform indicated computations on lines A to G. Note that subscript c refers to combined area. Omit entries and computations outside the double lines unless the ratio of the second characteristic to the first is required.

If, however, the random group sort is not included as a part of the standard census processing, the computation of sampling errors will require other methods. These methods may involve additional computer programming and processing. Some shortcut methods may be considered, but processing after tabulation is not recommended, since the variance tabulations are less likely to be produced if they are dependent on a separate operation. This is a very important consideration. After the principal tables needed for publication of census totals and estimates have been prepared, there may not be enough time, money, and technical manpower to make additional tabulations to obtain estimates of the variances.

Chapter F-5. DATA PROCESSING CONSIDERATIONS

1. INTRODUCTION

The following chapter explores several alternatives to the methods used in the Popstan processing chapters in Parts B and C. Also included are discussions of software packages for the editing and tabulating of census data.

2. ALTERNATIVE METHODS OF DATA ENTRY

As discussed in Part A, there are many data entry devices and techniques now available. These techniques can be classified into three basic categories: keying, automatic reading, and source data sensing.

2.1 *Planning*

Ordinarily, a statistical office does not have the luxury of designing a data entry system solely for processing population censuses. Costly data processing equipment must be justified by its use in other operations of the statistical office. Therefore, if new data entry devices are to be acquired for a census, the choice of the method must be coordinated with data entry plans for other censuses, surveys, payroll, and other administrative purposes.

Data entry devices suitable for a census are not always compatible with the needs of surveys or payroll. A method that may be best for a population census may not be practical for an agricultural census. A population census is characterized by massive amounts of data that must be converted to machine readable form in as economical a way as possible consistent with established standards of accuracy.

2.2 *General recommendations*

The following are general principles to be kept in mind when choosing a more advanced data entry system.

(1) The faster data can be converted to machine-readable form, the better. Time is money.

(2) The less data is handled by either man or machine, the better. Unnecessary handling of data may result in data deterioration.

(3) It is better to edit data mechanically nearer the source of entry. Sophisticated imputations for missing information and consistency checks which compare related pieces of information should be performed at a later stage prior to tabulation. But simpler forms of edits may be carried out with considerable cost and time saving with modern data entry devices. The simpler edits are those that ensure that the keyer has entered all the data available from the source document. A check for valid codes, a check to determine whether all records for a household or whether all records within an EA have been keyed are appropriate at this point. The latter is called "zero balance editing" because when the number of records for a household or EA are compared to control numbers there should be a zero balance or remainder.

(4) Errors cause many seen and unseen costs. Machines that introduce fewer errors during data conversion and during data transfer are better. For example, devices that rely on sight verification are not a good choice because they introduce a chance for human error.

(5) Finally, and obviously, the system that costs less, yet incorporates the above principles, is preferred.

In designing a data entry system, consideration also must be given to present equipment, human resources available, budget, and

the degree of error that can be tolerated. The principles cited above must be balanced with each other as each data entry method has its own strengths and weaknesses.

2.3 *Methods of data entry*

The various modes of data entry are presented below in the interest of showing a complete picture of what types of equipment are presently available to statistical offices. The characteristics of a census preclude the use of some methods. Other methods have not been perfected. Although some of these methods are not practical for processing censuses now, they may prove useful in the future.

Entry Category	Device
Keying	Keypunch Verifier Key to paper tape Buffered key punch Key to magnetic tape Key to disk Key to diskette Key to cassette
Automatic reader	Optical character reader (OCR) Optical mark reader (OMR) Bar code reader Magnetic ink character reader (MICR) Magnetic stripe reader
Source data sensing	Scales Counting Voice recognition

Population censuses are characterized by large volumes of data. Depending on the size of the population, millions of records may be involved. To handle this mamouth task, two basic methods have proven to be the most practical to convert data to machine readable form, at this time; either some type of key entry device or some type of mark reader is used.

Optical character readers (OCR) may be a practical method in future censuses. It is hoped that hand-printed answers on the census questionnaire can be entered directly into the computer. No country has been able to accomplish this as yet. The primary reason this has been unsuccessful is the difficulty of training interviewers to make numeric characters or letters that meet the tolerance of the OCR scanners for size and shape. When handwritten characters are malformed, the OCR scanner cannot read them. This results in an unacceptable reject rate.

3. *OPTICAL MARK READERS (OMR's)*

Optical mark reading (OMR) devices have been used for several decades in some countries to process the censuses of population and housing. The amount of success in using the equipment has varied by country. Recent improvements have made the use of OMR devices a much more viable alternative to keying devices. Therefore, during the present round of censuses and in the future, many more countries will be using various models of these devices.

3.1 *Advantages*

There are several obvious advantages that OMR equipment has over keying equipment. However, there are some theoretical advantages that do not always appear in practice.

3.11 Speed.--Ordinarily, optical mark readers are much faster than key entry devices. With OMR, speed is measured in pages, not characters. One characteristic of OMR's is that the reading speed is slowed only slightly when the number of characters per page is doubled, or even tripled. With keying equipment, 200 characters to a page will consume twice as much entry time as 100 characters to a page.

3.12 Cost.--Due to economies of size, a statistical installation should have a minimum number of documents to process in one day, approximately 5,000. Given that level, OMR equipment would be much less expensive than key entry equipment needed to process the same number of documents.

3.13 Less personnel and space.--An OMR requires one person to operate the device. In countries where personnel available for training as data entry operators are scarce, OMR's have a strong argument in favor of their use. In addition, the government often trains key operators, only to lose these people to commercial establishments which pay higher salaries than the government. Since only a few operators will be needed, it is possible to pay them a higher salary.

The need for fewer devices also saves space. Most OMR's are not larger than one keying station. Therefore, the need to rent or acquire large amounts of additional space for the data entry operation may not be necessary. Thus, the use of OMR's may be less costly than the use of keying equipment over a long period of time.

3.14 Verification unnecessary.--The use of OMR's eliminates the verification operation needed for keying operations by eliminating keying errors. Transcription errors made during the keying operation are the most common type of error introduced to the data. By eliminating this type of error, the result may be that the computer editing operation will be less costly.

3.2 *Disadvantages*

Consideration must also be given to the disadvantages of OMR equipment when making a decision on data entry equipment. The disadvantages should be weighed against the advantages before a final decision is made.

3.21 Error rate.--The most important disadvantage of OMR equipment is a relatively high error rate. The problem lies not with the OMR equipment itself, but with several aspects related to the method. The scanning machine demands high quality printing. If marks are not calibrated exactly, the machines will misread. Atmospheric conditions disturb the reading since paper which has absorbed moisture will expand causing misreadings.

Similar problems arise from the way forms are treated in the field. Mishandling of forms in the field are inevitable and such circumstances contribute to error rates. Rainspots, coffee or tea stains, stray marks, wrinkles, folds, and tears will cause misreadings. Careful control of the data collection operation could mitigate this problem. However, a transcription operation may have to be planned as part of processing. Thus, transcription errors may be introduced into the data.

Some models of the newer devices are more tolerant of mishandled forms. Internal programming can discriminate between stray marks by interpreting differences in shading. When selecting devices, this feature should be investigated.

3.22 Services.--All machines with moving parts break down sooner or later. The servicing of the equipment is a very important consideration in the decision to buy or rent OMR equipment. If servicing is not available in the country, long delays in processing the census can be expected. The same problem also arises with other data entry equipment. However, since only one or two OMR devices are used compared with 20 to 100 key entry devices the results may be a complete halt of the operation.

3.23 Questionnaires.--Optical readers vary in their ability to read questionnaires of different size. Many are restricted to documents no greater in size than 8½ inches by 11 inches (21¼ by 27½ centimeters). This restriction has an effect on the design and content of the questionnaires and the processing procedures.

In addition, more space on the questionnaire is required for each question. Answers must be pre-coded and the answer codes must be

spaced wide enough apart so that one mark can be distinguished from another. The Popstan Phase I questionnaire, already in precoded format, is shown in exhibit F-5-1 in OMR format. The OMR form is 3½ times as long as the form required for keying.

Until recently, most OMR equipment has required that the questionnaires be printed on extremely high quality paper. The newer models do not have as high a requirement. However, even the lower quality paper required may not be available in most countries. This could lead to an increase in costs. Not only is the paper a problem, but the printing of the questionnaires must be especially precise and special inks must be used. Many countries do not have the printing capabilities needed and the printing has had to be done outside the country; thus, creating opportunities for delay and costly mishaps. The correction of printing galley proofs and institution of quality control procedures are made difficult by distance. Further complications can be caused by a difference in language between the printer and the statistical agency.

3.24 Processing.--If booklet type questionnaires are needed to collect all the information, another problem may arise. Most optical mark readers can accept single page sheets, not booklets. To solve this problem the booklets must be cut on the bound edge and in this way single sheets can be fed into the optical reader. Careful attention must be given to the now detached sheets so that they remain in proper order. Sheets for a household must be kept together by restapling, stamping the serial number on each sheet, or other procedures.

3.3 Options

The planning and conduct of the census must take into consideration the advantages and disadvantages of OMR equipment. This is especially true when making the decisions on the data collection procedures, the content and format of the questionnaires, and the data processing procedures.

The one main advantage of using OMR equipment is the speed at which the data can be processed. One country has planned a two-phase census using this advantage. The first phase has been designed to collect only the most basic data on a 100-percent basis. A household form that can accommodate data for up to 5 persons will be used; continuation forms for larger households are planned. In this way, the basic information can be processed and published within a very short time. The second phase of the data collection will use a booklet form very much like the Popstan Phase II form. Keying equipment, which is already available in the country, will be used to process this form. Since the Phase II data are collected on a sample basis, not as much equipment, personnel, or time will be needed to process the data.

Another alternative is to use multiple questionnaires. Several countries have had success using this method. A separate questionnaire is used for each person in the household; the housing questionnaire is also on a separate piece of paper. The household roster part of the questionnaire can be included on the housing questionnaire. Controlling these loose sheets of paper is extremely difficult. Each sheet will have to have the housing unit serial number entered on it correctly in order to assemble all the information for computer editing and tabulation. This method can be quite feasible if the housing and population censuses are taken at different times.

OMR equipment has been used for one-phase censuses. Some countries have used booklet type forms in which most of the items were precoded. Items that could not be precoded, such as occupation and industry, were collected

by the traditional write-in method. Special spaces were left on the questionnaire for coding these items; the coding was done by an office manual coding operation. Some countries have solved the problem of not being able to use booklet type questionnaires on commercially available equipment by designing and building their own machines.

Another option, which should be considered last, if at all, is to use the services of several international private corporations. The companies will help design and print the questionnaires according to the specifications of their own equipment and will process the questionnaires. This may be less costly than renting the equipment. However, problems of confidentiality may be introduced. Also, the country loses effective control of the processing operations.

4. *DATA ENTRY EDITS*

Chapter A-9 discusses the types of edits that have been traditionally performed on data for censuses and surveys: enumerator edit, field edit, preliminary office edit, manual central office edit, and computer edit. Now, due to advances in technology, data entry edit must be added to the list.

In general, it is best to make corrections as close to the source of error as possible. However, editing at the data entry stage should only be viewed as improving the quality of the operation itself. Just as manual editing is not a substitute for computer editing, neither is data entry editing.

Since the early days of card punching, attempts have been made to identify and correct errors made during the data entry process itself; these attempts have taken the form of verification. As computers have become more powerful and more sophisticated, so has data entry equipment. Machines available today are capable of performing various tests or edits on the data as they are being entered; therefore, errors made during the data collection, manual editing-coding, and data entry processes can be detected and corrected prior to the computer processing. At this time, however, unless the data entry equipment is connected to a computer, the equipment can only indicate (flag) entries which may be in error; all correction of data must be done manually or later by the computer.

Human judgment is essential to <u>designing</u> data edits, but edit decisions should NOT be made by key operators when entering data. This leads to arbitrary decisions; it makes production of edit statistics on errors and corrections impossible. However, equipment which assists the keyer to work accurately is a valuable tool that increases productivity and accuracy.

4.1 *Types of data entry equipment*

As mentioned earlier, there are several different types of data entry equipment, ranging from card punching to voice transmission. Not all of them are well suited to performing editing tasks concurrently with their major function, data entry, and most types are strictly limited in their editing functions and capabilities.

4.11 <u>Keypunch equipment</u>.--Until recently, the most widely-used type of data entry equipment was the keypunch machine. The capabilities of these machines are basically confined to the data entry function. Some keypunch models do not immediately release the keyed card, but store the data until the keying for one card is finished. In this manner, self-detected errors by the keypunch operator can be corrected before the card is actually punched. These two features of keypunch machines, verification and data storage, give them some very limited editing capabilities,

but generally only that of correcting the errors which have occurred in the keying operation itself.

4.12 Key-to-tape and key-to-disk equipment.--The kinds of data entry equipment that have the best features for editing, at the present time, are the key entry devices, such as key-to-disk and key-to-tape machines. The manufacturers of these machines have gone quite far in the development of editing functions within their systems. The key-to-disk machines are usually better than the key-to-tape, for they have the capability of intermittent storage, which allows more flexibility in editing.

4.13 Optical scanning equipment.--Although optical scanning equipment, in the form of optical character readers and optical mark readers, has been in existence for several years, there still remains a very wide range of editing capabilities among the various manufacturers of such equipment. Several of the manufacturers of optical scanning equipment provide a minimum of correction capability within their equipment, while a few others have some features which allow very desirable editing functions to be performed. These kinds of machines have the potential to perform quite extensive detection, correction, and modification, but up to now the thrust of their development has been concentrated on refining the technological aspects. This is understandable, for the problems with optical recognition, such as paper costs, forms quality, and recognition of handwritten characters, have required much strenuous effort to solve.

4.14 Other equipment.--Other methods of data conversion, such as voice entry and the digitizer machine, are still too new to have developed many subsidiary capabilities. These techniques are just emerging from preliminary development stages and have still to be perfected and accepted as effective data entry devices. They contain no editing capabilities as yet, but probably, in the near future, can be enhanced with numerous capabilities facilitating editing.

4.2 *Types of data entry editing*

Editing as data are keyed is best used in a population census to perform validation checks. These should be done in hierarchical order; that is, on the character, field, record, and batch levels in that order. When a character is first keyed, a character check may be performed; when the entire field has been keyed, an edit check on the field can be performed. This process can continue until the entire record is keyed; then a record edit may be performed, and when a batch or group of records is completed, a batch edit can be performed.

4.21 Character and field checks.--A large amount and variety of character and field checks can be done on some key-to-disk data entry equipment. These checks are especially important in controlling the quality of the keying.

(1) Numeric-only checks verify that only the digits 0 through 9 and the plus or minus signs are permitted. Alphabetic-only checks prevent the entrance of digits 0 through 9 into the field. Check digit verification is used to ensure that a field containing numeric data is keyed correctly. To do this, a digit is added to the end of the original data field. This digit was computed to be arithmetically related to the rest of the field. When the data are keyed, this final digit can be checked to ensure that the entire field is correct. Key entry devices exist that can generate or validate check digits using up to 15 different algorithms.

(2) Field boundary checks prevent the keying of too many characters in a field, hence overflowing into the next field. Field use checks specify whether a field is all upper-case, all lower-case, or a combination of the two.

(3) Range checks are used to specify upper and lower limits for a number entered in a particular field, then check to validate

that the number is within that range.

(4) Mandatory entry checks ensure that at least one character is entered in the field. Mandatory completion checks ensure that each and every character in the field is keyed.

(5) Sign checks verify that all keys are signed either positively or negatively.

(6) Justification checks verify that the data is shifted to its proper location within the field.

(7) Fill checks verify that unused character positions within a field are correctly completed with zeros, spaces, dashes, etc.

4.22 Record checks.--Record edits can be utilized to make several different types of comparisons after the entire record has been keyed. Several different kinds of functions can then be performed.

(1) Data fields can be compared against values stored in a table which contains all valid or invalid entries.

(2) Ascendancy or descendancy checks can ensure that a number entered in a field is either greater or lesser than another number, which could be from another record or from another field in the same record--such as age and years lived in district.

(3) Duplicate checks can verify if fields in a current record exactly conform to fields in a prior record. Fields within the record, or the record itself, may contain the capability to be flagged for future reference.

(4) Record counter fields denote the sequencing of records within a batch. Data insertion capabilities allow fields to be added to records.

4.23 Batch edits.--Batch edits occur at the termination of a group of records and are used to test relationships between records within the batch or group. Batch totals or balances automatically verify sums of identical fields for several records. Deletion or insertion of whole records can be accomplished at batch edit time. Some key entry devices contain an editing language, similar in scope to many programming languages, which give them the capability of performing almost all of the types of edits done at batch time as would be accomplished during a later computer edit.

5. *COMPUTER FACILITIES*

The details on the selection and acquisition of a computer system are covered in Chapter A-9, Data Processing Considerations. This section is concerned with problems and alternative plans of action when available equipment must be used even though larger, more reliable machinery would facilitate the processing.

5.1 *Small computer systems*

In processing the census, both the capacity and configuration[1] of the computer place limitations on the programs that can be used. Editing packages using the hot deck and other sophisticated techniques require a relatively large computer system. As in Popstan, the capacity and configuration can be enhanced to accommodate one or more of the software packages available; however, this is not always possible due to the costs involved. If enhancement of the computer is not feasible, there are several alternatives that can be considered.

One alternative is to select a package program or write a custom program which indicates (flags) errors but does not impute or allocate values. Corrections will be made manually. It may be argued that it would be better to do all editing manually. However, it would be less costly in terms of time and

[1] *Capacity refers to (a) the amount of space available in the computer for storage and manipulation of data and (b) the speed at which the data can be processed. Configuration refers to the hardware components (tape drives, on-line-disk units, other input/output devices, etc.) that are connected to each other; they must operate as one system.*

personnel and more accurate and thorough to have the computer indicate where corrections are needed. If the data are manually edited, there is no guarantee that the data are error free or that they have been edited consistently such as is assured by a comprehensive, thoroughly tested, computer editing program.

Another solution is to use a more powerful program package and pass the data through several times, each time doing a separate editing step. For example, one pass could be for range checks, a second pass for editing the basic characteristics (relationship, sex, age, and marital status), a third pass for editing educational characteristics, and so forth. This procedure will only work if the available machine is only marginally smaller than is required to edit the data in one pass.

This type of procedure would require a stricter and more elaborate operational control system. Records reflecting the status of each control or work unit as it goes through the computer processing will have to be maintained to ensure that each step is completed.

Using the second alternative, the processing will take a longer time, not only because of the multiple passes planned, but also because the work unit may be rejected during several of the passes. Therefore, the work unit may have to be manually reviewed and reprocessed more than once, that is, each time it is rejected. Given the control problems and the additional computer time required, this alternative is still more conducive to consistency in editing and correcting the data than manual editing.

5.2 *Large computer systems*

In some countries, the problems may be caused by computer systems which are capable of processing multiple jobs at one time. A major problem in these instances is control. One of the most common problems is that data tapes may be misfiled and erased by accident.

When using this type of computer system, priority scheduling should be instituted. The census should have a high priority. Unfortunately, often the machine scheduling is bypassed. If this is a chronic problem, blocks of computer time should be set aside for processing the census data alone. In this way, the data can be passed through the computer at a quicker rate and the control problems will be mitigated. In some instances, it may be feasible to reserve the computer to process the census data during a second or third shift. If the computer is not used on a three-shift basis, it may be worthwhile to hire and train the required personnel to run the computer for a third shift.

5.3 *Power failures*

Interruptions in the computer processing resulting from power failures or fluctuations are a serious problem in many countries; these interruptions can result in the loss of any output stored in the computer. If a new computer is to be acquired, serious consideration should be given to acquiring equipment that does not need special environment and a large supply of electrical power. These computers are easily installed, an ordinary electrical outlet is all that is required, and need less physical space. Added advantages are that a service staff and a system programming support staff are not required.

If the census is to be processed on an already installed computer system, it may be wise to invest in a gasoline or diesel driven generator. In this way, a constant power source is assured.

If neither of the above alternatives is feasible, then procedures should be established to circumvent or mitigate the problem.

One solution would be to schedule production runs during off-peak hours. That is, at times when less power is being used in general and power failures are less likely to occur, such as at night. Another solution is to keep the individual jobs small and to write the output files on off-line media, preferably tapes, as they are produced. If this is done, the input can be loaded to start at the point when the power failure occurred and the output of the previous processing will not be lost.

6. *SOFTWARE DEVELOPMENT*

As discussed in Part A, some custom programs always will be needed for processing the census; in fact, many countries prefer to produce all the software required rather than to acquire generalized packages. The production of software should be very carefully planned and should start as early as possible.

All programs should be written by the time the experimental or pilot census is to be processed. It is essential that the programs be tested, debugged, and operational by the time the first census questionnaires are received from the field. This will allow the programmers to concentrate on day-to-day problems that arise. The editing programs should be started at least 1 to 1½ years before the experimental census.

Generally, it is best to write the programs using a standard language such as COBOL which can be used on almost all computer systems; this is especially true when equipment changes are planned. In addition, if equipment failures occur during processing and an outside computer must be used, the program will be much more portable; there will not be the problem of rewriting the programs into another language, which will take time and delay the processing of the data.

If it is decided to write custom programs to prepare tabulations, rather than using a tabulation package, the tables should be carefully organized into groups for programming. To minimize the number of times the entire data file must be read to produce the tabulations, as many tables as possible should be prepared by each program. It is also desirable to place similar tables in the same program to simplify programming and any necessary modifications later. Priority should be placed on grouping together tables using the same boxhead.

7. *COMPUTER EDITING*

Traditionally, editing by the computer has been performed on one record at a time and one variable at a time. As each record was entered, a series of tests were performed on the variables sequentially. Generally, the edits were only concerned with consistency between the variables within each record. A major criticism of this method of editing was that often good data was changed to make it consistent with earlier changes in the record.

With the advent of larger and faster machines and more sophisticated techniques, editing programs have been written which now check consistency of variables in one record with variables in another related record, such as between household members. Also, more tests can be made before an entry is changed. CONCOR, which is discussed in section 8, is one such program.

In recent years, another method of editing has been used; the Fellegi-Holt method.[2] Unfortunately, the programs which have been written to employ this method can only be used on extremely large computer systems. However, in the future, advances in the field of computer science may make it possible to write more efficient programs using this technique.

---

[2] *I. P. Fellegi and D. Holt, "A Systematic Approach to Automatic Edit and Imputation," Journal of the American Statistical Association, Vol. 71, No. 353 (March 1976), Applications Section.*

The purpose of the technique is to make the minimum number of changes in each record. Consistency checks between related variables within a record and between related variables in other household records are used to find the variable (variables) that has the highest inconsistency rate before any changes are made. For each data record, tests are made of all logical relationships between all characteristics that have valid codes; blanks and non-valid entries are ignored. The percentage of inconsistency for each characteristic is calculated. Then, the characteristic that has the highest percentage of inconsistency is changed. For example, if tests of a person's record result in age being inconsistent 35 percent of the time, marital status 5 percent of the time, and relationship 10 percent of the time, the value for age will be changed first. After each change, tests are run again to ensure that the changes have not resulted in additional inconsistencies.

8. *SOFTWARE PACKAGES*

Several general editing and tabulating software packages have been developed over the last few years by the International Statistical Programs Center of the U.S. Bureau of the Census. The tabulation package has been used in many countries. The editing package, as of this time, is in the process of being installed in various countries.

8.1 *CENTS and COCENTS*

CENTS and COCENTS were developed to tabulate large census files. CENTS is written in ALC, the IBM assembly language, and COCENTS is written in COBOL. Although the programs are basically the same, CENTS is slightly more powerful and faster.

The program can be run on a wide range of small-to-medium-sized general purpose computers. Generalized software packages which are parameter driven, offer flexibility in table formatting to the user. The programs have been successfully installed and used in many countries around the world.

Some of the characteristics that make these programs one of the most widely-used packages are listed below:

(1) The parameters are easy to prepare and use. Little or no programming experience is needed; subject matter specialists are able to prepare the parameters after a minimum of instruction.

(2) Hierarchical files can be processed.

(3) Very little computer memory (space) is required and many tabulations can be produced with one pass of the data file.

(4) Execution is extremely quick. Little computer time is used, thus lowering the processing cost.

(5) Tabulations can be produced in small runs if necessary, and later consolidated for publication. This is especially useful if power failures are a problem.

(6) Tabulations are produced in print-ready form.

(7) Tabulations comparing historical census files with current data can be produced.

(8) Both right- and left-hand stubs can be printed in the same table; or the right- or left-hand stubs can be used separately.

(9) Upper and lower case printing can be obtained in the final tables. Footnotes can also be produced.

(10) Input files do not have to be in a special format; punch cards, tapes, or disk files can be used.

(11) Fractional and integer weighting can be used for sample data.

(12) Measures such as means, medians, averages, and percent distributions can be calculated and printed in the tables.

(13) Custom programs can be inserted in all of the separate parts of the software for special purposes.

(14) Comments can be inserted with ease.

The programs with reference manuals can be acquired for the cost of copying. Various conversions of the programs have already been made and, therefore, customized versions can usually be prepared for the user on request.

8.2 CONCOR

The CONCOR edit system was originally developed by the Latin American Demographic Center (CELADE), the system has been enhanced and modified by ISPC. CONCOR is written in COBOL and can be used on medium-to-large computer systems.

8.21 Computer requirements for installation.--The CONCOR program requires the following:

(1) A minimum of 128K (128,000) bytes or characters of primary storage.

(2) A minimum of 4 million bytes or characters of on-line disk storage.

(3) A COBOL compiler.

8.22 Major editing capabilities.--CONCOR has several features that make it very attractive to statistical offices with limited programming capability but large files of data from a census or survey. The principal features of CONCOR are:

(1) Structural Checks: These checks are designed to ascertain whether all the records that should be present for a particular questionnaire are in fact supplied.

(2) Range Checks: These checks are designed to find out whether a variable has a value that is inside or outside the limits specified for that variable. These checks, which are often called "data validation," or valid value checks, are concerned with one variable at a time.

(3) Consistency Checks: In these checks, two or more variables are compared and the consistency of their values checked. CONCOR provides the capability to locate these variables in the same record, or in different records, or to accept them as computed values. This new interrecord checking capability represents a significant improvement over previous versions of CONCOR.

(4) Automatic Corrections: When censuses or large surveys are being processed, the user will find it essential to have the computer make some or all of the editing corrections automatically. Routines are included in the system to handle imputations based on both the "cold-deck" and the "hot-deck" methods.

(5) "Error-Free" Output File: CONCOR can automatically create an output file containing the data records from an input file to which automatic corrections have been made. This output file will be identical in format to the input data file, and will provide the user with the ability to enter the "edited" file in CONCOR and inspect the corrected data for possible inconsistencies introduced during the correction process. This is an invaluable tool for testing the user's edit specification statements in the CONCOR program before initiating production processing of the final data file.

(6) Auxiliary Output File: CONCOR provides the ability for the user to specify the content and structure of additional output data files. This allows the user to create derivative files containing subsamples, recoded data, or original variables transferred into a new format.

8.23 Input data file constraints.--To be edited using CONCOR, the input data file must satisfy the following restrictions:

(1) All input records in the file must be of equal length (for IBM System 370 OS computers, the range of permissible record lengths is 20 to 32,760 bytes).

(2) All records must have a questionnaire identification number or code which is a string of fields of fixed format and fixed length, and these fields must be in the same position in every record. This condition may be satisfied by a collection of fields giving geographic coding information and a household number which is unique within the lowest geographic level.

(3) Each record must have a record type code in a fixed position and of a fixed length and in the same position for all record types.

(4) All records for each questionnaire must be physically contiguous on the input file.

(5) All data records for a particular geographic control area must be contiguous on the input file if the user wishes to have CONCOR compile statistics separately for each unique control area.

(6) CONCOR allows up to 50 different record types to be defined with any record type having a maximum of 999 occurrences in any one questionnaire. Also, a maximum of 745 unique user-identifiers may be defined.

8.24 Edit diary.--CONCOR is capable of producing three types of tallies on the edits which have been performed on the data. One report (see figure F-5a) produces summary statistics on edits to individual variables from the questionnaire. A second report (see figure F-5b) shows summary statistics by type of edit, be it range check, consistency check, or an allocation. A third type of report shows errors by indivdual questionnaires (see figure F-5c). In addition, for each edit run a summary report is produced (figure F-5d).

The first three types of reports can be generated for every control area (usually a geographic area) or for a grand total of all questionnaires edited if no control parameters are included. All or any one of these reports can be suppressed. The third type of report, the description of every edit made for each questionnaire, would produce volumes of information. If a census was being processed, much of the information would be unneeded; it would certainly be too voluminous to handle effectively. However, this part of the diary should be used during the pretesting stage to ensure that the program is working correctly.

Figure F-5a. EDIT STATISTICS REPORT--COUNT OF EDITS BY VARIABLE

```
VERSION 2.2                                   C O N C O R                                                    PAGE    34
                                          EDIT STATISTICS REPORT                                     RUN DATE   03/26/81
                                            BY USER-IDENTIFIER

CONTROL AREA: 1:3:1:0:1

*************************************************
*   P08B-CITIZENSHIP                            *
*************************************************

ERROR TYPE: RANGE
-------------------------------------------------------------------------------------------------------------------------
SOURCE                                                                                NOMATCH       NUMBER       NUMBER      PER
  LINE   ERROR                                                                          VALUE        CASES        CASES     CENT
NUMBER  NUMBER  MESSAGE                                                              ASSIGNED       TESTED       FAILED   FAILED
-------------------------------------------------------------------------------------------------------------------------
  1093  SYSTEM  RANGE P08B VALUE 1,2,9 NOMATCH 9                                          YES            8            0     0.00

                      TOTAL RANGE COMMANDS FOR THIS IDENTIFIER = 1
                                  TOTAL NUMBER OF CASES TESTED = 8
                                  TOTAL NUMBER OF CASES FAILED = 0
                                    PERCENTAGE OF CASES FAILED = 0.00


ERROR TYPE: ASSERT
-------------------------------------------------------------------------------------------------------------------------
SOURCE                                                                                                NUMBER       NUMBER      PER
  LINE   ERROR                                                                                         CASES        CASES     CENT
NUMBER  NUMBER  MESSAGE                                                                               TESTED       FAILED   FAILED
-------------------------------------------------------------------------------------------------------------------------
  1096  SYSTEM  ASSERT P08B = 1                . MUST BE POPSTAN CITIZEN                                580          127    21.90

                     TOTAL ASSERT COMMANDS FOR THIS IDENTIFIER = 1
                                  TOTAL NUMBER OF CASES TESTED = 580
                                  TOTAL NUMBER OF CASES FAILED = 127
                                    PERCENTAGE OF CASES FAILED = 21.90


ERROR TYPE: ALLOCATE
-------------------------------------------------------------------------------------------------------------------------
SOURCE
  LINE                                                                                                  NUMBER OF
NUMBER          SOURCE PROGRAM TEXT                                                                   ALLOCATIONS
-------------------------------------------------------------------------------------------------------------------------
  1098          ALLOCATE P08B = 1                                                                             127

                  TOTAL NUMBER OF ALLOCATIONS FOR THIS IDENTIFIER = 127
```

Figure F-5b. EDIT STATISTICS REPORT--COUNT OF EDITS BY TYPE OF EDITS

VERSION 2.2 C O N C O R PAGE 69

(C O N S I S T E N C Y A N D C O R R E C T I O N) RUN DATE 03/26/81

E D I T A N D I M P U T A T I O N S Y S T E M

EDIT STATISTICS REPORT

BY EDIT SPECIFICATION COMMAND

CONTROL AREA: 1:3:1:0:1

****
* U S E R - S P E C I F I E D M E S S A G E S *
****

SOURCE LINE NUMBER	ERROR NUMBER	MESSAGE	NUMBER CASES TESTED	NUMBER CASES FAILED	PER CENT FAILED
26	1	STRUCTURE ERROR - INVALID RECORD TYPE EXISTS	124	1	0.81
45	10	STRUCTURE ERROR - NO HOUSING RECORD	7	1	14.29
64	20	STRUCTURE ERROR - MORE THAN 1 HOUSING RECORD	7	6	85.71
91	25	STRUCTURE ERROR - EXCESS HOUSING RECORDS	0	0	0.00
107	30	STRUCTURE ERROR - WRONG NUMBER OF POPULATION RECORDS	124	9	7.26
120	40	STRUCTURE ERROR - NO PERSONS IN OCCUPIED UNIT	106	1	0.94
125	50	STRUCTURE ERROR - PERSONS PRESENT IN VACANT UNIT	18	17	94.44
188	60	GROUP QUARTERS NOT SHOWN IN TYPE OF UNIT	2	0	0.00
209	70	RELATIONSHIP CODE SHOULD BE GROUP QUARTERS	16	16	100.00
228	100	STRUCTURE ERROR - NO HEAD OF HOUSEHOLD	119	11	9.24
384	110	STRUCTURE ERROR - EXCESS HEAD RECORD	445	9	2.02
507	120	STRUCTURE ERROR - EXCESS SPOUSE RECORDS	340	1	0.29

TOTAL NUMBER OF USER-DEFINED MESSAGES = 12
TOTAL NUMBER OF CASES TESTED = 1,308
TOTAL NUMBER OF CASES FAILED = 72
PERCENTAGE OF CASES FAILED = 5.50

****
* S Y S T E M - G E N E R A T E D M E S S A G E S *
****

SOURCE LINE NUMBER	MESSAGE	NUMBER CASES TESTED	NUMBER CASES FAILED	PER CENT FAILED
116	ALLOCATE H15 (1) = N03	9	NA	NA
197	ALLOCATE H01 (1) = 98	0	NA	NA
214	ALLOCATE P02 (N05) = 8	16	NA	NA
280	ALLOCATE P02 (N22) = 0	10	NA	NA
388	ALLOCATE P02 (N05) = 5	9	NA	NA
511	ALLOCATE P02 (N05) = 5	1	NA	NA
529	RANGE H01 VALUE 0 TO 9 NOMATCH 0	123	0	0.00

Figure F-5c. EDIT STATISTICS REPORT--COUNT OF EDITS BY QUESTIONNAIRE

```
VERSION 2.2                                   C O N C O R                                                    PAGE    28
                                          EDIT STATISTICS REPORT                                   RUN DATE   03/26/81
                                            BY QUESTIONNAIRE

CONTROL AREA:  1:3:1:0:1

QUESTIONNAIRE IDENTIFICATION:  0103:001:00:001:041
-------------------------------------------------------------------------------------------------------------------------
SOURCE
 LINE   MESSAGE
NUMBER  NUMBER   TEXT                                                                  IDENTIFIER(S) WITH CURRENT VALUE(S)
-------------------------------------------------------------------------------------------------------------------------
    64      20   STRUCTURE ERROR - MORE THAN 1 HOUSING RECORD
   107      30   STRUCTURE ERROR - WRONG NUMBER OF POPULATION RECORDS                  N02 = 0  N03 = 2
   125      50   STRUCTURE ERROR - PERSONS PRESENT IN VACANT UNIT                      N01 = 6  N03 = 2
   995  SYSTEM   ASSERT P02 (2) = 1 IMPLIES P05 = 0 OR P05 = 3   . IF SPOUSE           P02 (2) = 1  P05 (1) = 1  P05 (1) = 1
  1236  SYSTEM   ASSERT P11 < 23 IMPLIES P12C <> 1   . IF LESS THAN 3 YEARS SEC        P11 (1) = 15  P12C (1) = 1
  1018  SYSTEM   ASSERT P05 = P05 (1)          . MARITAL STATUS MUST BE SAME AS HEAD   P05 (2) = 1  P05 (1) = 0
  1096  SYSTEM   ASSERT P08B = 1               . MUST BE POPSTAN CITIZEN               P08B (2) = 2
  1296  SYSTEM   RANGE P18 VALUE 0 TO 19,99 NOMATCH 0                                  P18 (2) = 60
  1301  SYSTEM   RANGE P19A, P19B, P19C, P19D VALUE 0 TO 19,99 NOMATCH 0               P19A (2) = 30
  1301  SYSTEM   RANGE P19A, P19B, P19C, P19D VALUE 0 TO 19,99 NOMATCH 0               P19D (2) = 30

             TOTAL ERROR FREQUENCY (FAILED) FOR THIS QUESTIONNAIRE =      10


QUESTIONNAIRE IDENTIFICATION:  0103:001:00:001:042
-------------------------------------------------------------------------------------------------------------------------
SOURCE
 LINE   MESSAGE
NUMBER  NUMBER   TEXT                                                                  IDENTIFIER(S) WITH CURRENT VALUE(S)
-------------------------------------------------------------------------------------------------------------------------
   577  SYSTEM   RANGE H10 VALUE 1 TO 6, 9 NOMATCH 9                                   H10 (1) = BLANK
   578  SYSTEM   RANGE H11 VALUE 0 TO 9 NOMATCH 0                                       H11 (1) = BLANK
  1262  SYSTEM   RANGE P13 VALUE 1, 2, 9 NOMATCH 9                                     P13 (1) = 0

             TOTAL ERROR FREQUENCY (FAILED) FOR THIS QUESTIONNAIRE =       3


QUESTIONNAIRE IDENTIFICATION:  0103:001:00:001:043
-------------------------------------------------------------------------------------------------------------------------
SOURCE
 LINE   MESSAGE
NUMBER  NUMBER   TEXT                                                                  IDENTIFIER(S) WITH CURRENT VALUE(S)
-------------------------------------------------------------------------------------------------------------------------
   995  SYSTEM   ASSERT P02 (2) = 1 IMPLIES P05 = 0 OR P05 = 3   . IF SPOUSE           P02 (2) = 1  P05 (1) = 1  P05 (1) = 1
  1236  SYSTEM   ASSERT P11 < 23 IMPLIES P12C <> 1   . IF LESS THAN 3 YEARS SEC        P11 (1) = 14  P12C (1) = 1
  1248  SYSTEM   ASSERT P11 < 32 IMPLIES P12D <> 1   . IF LESS THAN 2 YEARS UNIV       P11 (1) = 14  P12D (1) = 1
  1262  SYSTEM   RANGE P13 VALUE 1, 2, 9 NOMATCH 9                                     P13 (1) = 0
```

Figure F-5d. EDIT STATISTICS REPORT--CONTROL LEVEL SUMMARY

```
VERSION 2.2                                  C O N C O R                                        PAGE    74
                          ( C O N S I S T E N C Y   A N D   C O R R E C T I O N )            RUN DATE  03/26/81
                            E D I T   A N D   I M P U T A T I O N   S Y S T E M
                                          EDIT STATISTICS REPORT
                                           CONTROL LEVEL SUMMARY

CONTROL AREA: 1:3:1:0:1

            INPUT STATISTICS
                  NUMBER OF QUESTIONNAIRES PROCESSED                        124
                          NUMBER OF RECORDS PROCESSED
                                                  TOTAL                     719
                                                  BY TYPES
                                                    HOUSI                   129
                                                    POPUL                   588
                                                    INVALID                   2

            OUTPUT STATISTICS

                  VIA OUTPUT COMMAND

                      NUMBER OF QUESTIONNAIRES OUTPUT
                          USING 'QUESTIONNAIRE' KEYWORD ONLY                  0

                          NUMBER OF RECORDS OUTPUT
                                                  TOTAL                     711
                                                  BY TYPES
                                                    HOUSI                   123
                                                    POPUL                   588

                                        ERROR TOLERANCE SPECIFIED: 0.80

                                 TOTAL NUMBER OF NON-BLANK CASES: 20,114
                                   TOTAL NUMBER OF CASES CHANGED: 2,203
                                     PERCENTAGE OF CASES CHANGED: 10.95

****THIS CONTROL LEVEL IS NOT WITHIN THE SPECIFIED TOLERANCE     ****
```

Chapter F-6. ALTERNATIVE EDIT SPECIFICATIONS

1. *INTRODUCTION*

Part C of the Popstan case study showed examples of edit specifications in some detail. The edits presented in this chapter are similar in most respects in that they are designed for the Phase I questionnaire and not for the forms illustrated in Part F but differ in the programming requirements. The edits presented here require the capability of searching all person records within a household for information which would assist in allocating an answer to an item that had been left blank or was in error. The CONCOR edit package described in the previous chapter, has this capability.

2. *EDITING CONSIDERATIONS*

Missing responses and inconsistent responses make it difficult to present census data that will be easily understood by the users. The process of "editing" makes the data more convenient for the analysts and, at the same time, promotes wider use of the data by lending a degree of reliability to the statistics.

As discussed in chapter A-9 of the case study, some users prefer to tabulate missing and inconsistent responses as a "not reported" category; others prefer that these cases be distributed proportionately among the reported consistent entries. Still others recommend rules for imputing "likely" answers for the missing and inconsistent responses; the use of computers makes it feasible and efficient to impute responses based on other information on the questionnaire or on reported information for a person or housing unit with similar characteristics.

Editing can be simple or complex. When the percentage of missing or inconsistent responses is low (less than 1 or 2 percent), any reasonable edit rules are not likely to affect the use of the data. When the percentage is high (5 to 10 percent, or more), there is the danger that imputation will distort the census results, whether the procedure is simple or complex.

It can be argued that, since the computer can look at many characteristics, the editing process should take advantage of this feature. Thus, edit procedures involving many related characteristics may result in imputing more reasonable responses than a simple edit could produce. On the other hand, the computer creates the possibility of producing poor census data. Edit specifications should be set up jointly by experienced subject matter specialists and data processing specialists. Variables that are examined in the tests for consistency should be selected carefully to determine the degree of correlation between characteristics. Census publications should describe the edit and imputation specifications and then give the percentage of responses that were changed or imputed. Thus, analysts will be in a better position to judge the quality of the data; for example, a high percentage would be a warning to use the data with caution.

For the population data, it would be desirable to check the characteristics of all the people in a housing unit at the same time. The editing package described in Part C of this case study was not able to do that. It could save data from previous persons in a household, and such data could be referenced

but not changed. The edit package described in chapter 5 places all the records for one housing unit in a work area in the computer and edits the population records for the whole household at the same time. This capability greatly improves the ability to correct the data and to make more logical allocations for missing data.

3. *POPULATION EDITS*

Edits discussed in this chapter are for the items on the population records for a household that are affected by the expanded capabilities. The housing edits gain little advantage from this enhanced capability and therefore, edits for housing items will be no different than those presented in Part C, chapter 6, section 8, and exhibit C-6-3, which presents the edits in flow chart style.

The capability of examining data from other records within the household is called an interrecord check. The edit for each data item takes into consideration the changes that have been made to the items edited previously. The population items will be edited in the sequence found on the questionnaire except for age which is edited more effectively after marital status.

3.1 *Use of hot deck techniques*

The edit program will use the hot deck method to impute missing values. This technique takes advantage of the fact that data are sorted by EA's. Technicians assume that this increases the probability of obtaining a correct answer because imputed values are taken from records of people living in the same small geographic area who are normally similar to each other in background.

Before using the hot deck techniques, however, an attempt will be made to use related items which aid in assigning a value that is more likely to be correct than a hot deck value. For instance, if marital status is missing, rather than hot deck, the edit program will determine whether the person has a spouse in the household. If so, the program will assign MARRIED, a more probable value than if the program had used a hot deck. But when no such evidence is present, the program must rely on a hot deck value.

The probability of obtaining a correct hot deck value usually can be increased by making the hot deck array more detailed. As an example, the program could hot deck marital status using relationship alone. But the percentage of widowed or divorced men and women increases with age. Therefore, it makes sense to hot deck marital status by age and relationship. The program will take the value for marital status from the last preceding valid record, which has been stored in the hot deck, that has the same relationship and age as that of the person whose record is being edited.

This, however, can create a new problem. The questionnaire items are edited in a fixed sequence with age edited after marital status. Sometimes both marital status and age are missing from a record. Thus, it is impossible to take the value for marital status from the last preceding record that has the same age and relationship values because the program cannot determine into which age category this record falls. It is possible to establish an item in the hot deck array for "not reported" items. This would allow the program to assign a value for marital status using the marital status category from the last preceding record that has the same relationship category and "not reported" age. Two facts, however, argue against doing this. One is that these "not reported" cases in the same combination are so few that it would be difficult to update the hot deck array for the missing item. Secondly, there is no

relationship between a missing age and the proper relationship entry.

The solution to the problem creates more work for the programmer, but results in a statistically cleaner product. The program first will test to determine whether the items to be used have valid codes. If the record for the person being edited does not have a valid code for the item, the item will not be used in the hot deck for this record. A simpler hot deck array will be created. To continue the earlier example, if the program must hot deck marital status because the value is missing, the hot deck array will ordinarily be two dimensional; age and relationship will form the two items involved. If after testing, the program finds no valid code for age, it will hot deck marital status by relationship alone. Because relationship is edited before marital status, no invalid codes will be found for that item. The program will be written using these same principles for all hot deck procedures.

3.2 *Preliminary edits*

Before editing item by item, the computer program must check for valid records and duplicate line numbers. It must also determine whether the records being edited are for inmates.

3.21 Check for valid records.--Key operators can make a mistake in keying a record, and on occasion they will forget to delete this fragmentary and erroneous information. One function of the pre-edits will be to examine the file for fragmentary records which should have been deleted. The most common case will be a record which contains geographical codes but no population or housing information. Such a record will be deleted.

3.22 Duplicate line numbers.--These can appear either through rekeying a record or part of a record, or by miskeying the line number of another record. As each record is read into the edit program, the record will be compared with the other records of the household. If all the characteristics are the same as in another record, the first record will be saved and the second record deleted. If a record has less than three characteristics that are non-blank, or that are different from another record, the record will be deleted.

3.23 Inmate.--People who live in places such as orphanages, nursing homes, chronic disease wards in hospitals, and penitentiaries are considered INMATES. These places have one thing in common: the residents are under custody or care. If this HU has only one person, the inmate code will be checked to see if it is marked NO. If it is NO, the program will proceed to the next edit. Otherwise it will be assigned NO. It has been decided that to be classified as an institution with inmates, there must be more than one person.

If the HU has more than two records in which relationship is either HEAD, SPOUSE, SON/DAUGHTER, or OTHER RELATIVE (thus providing some indication that this is a family home), the inmate code will be examined. If it is NO, everything is consistent and the program will go on to the next edit. If it is not NO, it will be assigned NO.

If the indicators for a family home are not present (relationship are all NONRELATIVE or blank), the inmate code will be checked to see if it is YES. If it is, the program will assign INMATE to the relationship item in all person records for the HU. If the inmate code is not YES, it will assigned YES.

3.3 *Relationship*

Relationship is one of the most important characteristics since it will be used in many of the other edits for the household. The first part of this edit makes sure that there is one and only one head of the household. It will accept the first person in the household who is coded HEAD. Subsequent records coded HEAD will be assigned a blank for relationship. If there are no records coded HEAD, the first person will be given that designation.

Following a check for the head of the household, a series of checks will be made to determine whether the relationship of the people in the household is consistent with one another. First a check will be made for females who have their husband in the household. If the husband is head, the woman's relationship will be changed to SPOUSE if it is not SPOUSE already. If the woman's husband is a son or other relative of the head, she will be assigned OTHER RELATIVE. If her husband is no relation to the head, she too will be assigned NONRELATIVE of head if the entry for relationship is not already NONRELATIVE.

If a man or woman's mother is head or spouse of head, the person will be given a value of SON/DAUGHTER in relationship. A check will be made for a man whose wife is in the household. If the man's wife is daughter or other relative of head, the program will assign OTHER RELATIVE to his relationship. If a woman is the mother of the head or spouse of head (by matching her line number to "line number of mother" in the head's or spouse of head's record), she will be assigned OTHER RELATIVE in relationship.

When these consistency checks have been made, if there is a valid code for relationship for the record, the edit will proceed to the next edit. If there still remains a nonvalid code, the program will hot deck relationship by age and sex.

3.4 *Sex*

The code for P-3, sex, will be changed even though a valid code exists (1 or 2) in two instances where contradictory evidence seems strong. When sex is male and P-6, line number of husband, is greater than 00 (indicating that this is a woman who has a husband) and at the same time P-14, children ever born, is greater than 00 (indicating that this is a woman who had children), sex will be changed to FEMALE based on this relatively strong evidence. On the other hand when P-3, sex, is coded "2," female, yet upon checking P-6, line number of husband, of the other members of the household, the program finds this person's line number (thus indicating that this person is the husband of a female in the household), and at the same time the entry for P-14, children ever born, is blank, the program assumes that a mistake has been made and changes the code to MALE.

If the entry for P-3, sex, is blank or nonvalid, the program will use the entry for relationship to head and implied sex of spouse to determine the correct code. If the relationship to head is the HEAD, the program checks to see whether the spouse is present (by checking for another person in household whose relationship is SPOUSE OF HEAD). By determining the sex of the spouse, the opposite sex is assigned to the head.

If the relationship to head is SPOUSE and the sex of the head is given, the program assigns to this person the sex opposite that of the head's sex.

There are numerous clues in the questionnaire that could indicate whether or not

a respondent is female. If the person's sex has not yet been determined by the program and any of these female indicators are present, then the record for this person will be assigned FEMALE sex. For example, if the person being edited has children ever born, children living in this household, children living elsewhere, children dead, children born alive in the last 12 months, or this person is the mother of someone else in the household (this person's line number equals the line number of mother, P-8, of another person in the household) or if this person has a husband in the household or elsewhere (P-6 equals 00 or equals the line number of a male in the household), then sex will be assigned a value of 2, FEMALE.

Likewise, if this person is the husband of someone else in the household (this person's line number equals the line number of husband of a female in the household), the entry for sex will be assigned MALE or a value of "1." If there are insufficient data to impute sex from the above items, an entry will be allocated from a hot deck which has been set up to store codes for sex by age and relationship.

3.5 *Marital status*

Although marital status will be tabulated only for persons 12 years or older, it will be edited for everyone and age will not be edited until after marital status.

A few consistency checks will be made on the data for marital status. If marital status is MARRIED, the program assumes the code is valid. If the code is DIVORCED, SEPARATED, WIDOWED, or NEVER MARRIED, however, the records for the other people in the household will be searched to see if any of them are this person's spouse. If so, marital status for this person will be changed to MARRIED. In particular, if the person being edited is male and he has a wife in the household (his line number matches the line number of husband of female), then his marital status will be changed to MARRIED. If the person being edited is female and she has a husband in the household (her line number of husband matches the line number of a male), her marital status will be changed to MARRIED.

If there is no entry for marital status, but the entry for relationship to head equals HEAD, the program checks to see whether the head's spouse is present (by checking relationship to head for other member of the household). If spouse of head is present, the program assumes marital status for the head is MARRIED. If the spouse of head is not present, and this person is male with children present, the program hot decks marital status by age with children present. If no children are present, the program hot decks marital status by age with no children present. A man who is head of the household, but whose wife is not in the household is most likely to be divorced or separated, perhaps widowed, or even married with his wife temporarily living with her mother outside the houseing unit. To handle this situation, the hot deck procedure is most appropriate.

If marital status is blank, but relationship is HEAD, the head's spouse is not present, sex is not male, but P-6, line number of husband equals 00 (meaning husband lives elsewhere), then the program will assign MARRIED. If line number of husband is not 00, the program will hot deck marital status by children ever born.

If the relationship equals SPOUSE, the program will assign MARRIED to the record. If relationship is neither HEAD nor SPOUSE, and this person is male and this person's line number equals the line number of husband of

a married female in the household, the program will assign MARRIED. This person is presumed on this evidence to be the husband of the female. If there is no wife in the house for this person, the program will hot deck by age and relationship.

Finally, if sex is not MALE, but there is evidence (line of husband is 00 or equals a line of a male in house) that the woman has a husband, the program assigns MARRIED to marital status. Otherwise, the program will hot deck by age, relationship, and children ever born.

3.6 Age

The edit for age is long, but not difficult. There are many parts, but each part makes use of a clear relationship between one population characteristic and another.

The edit begins with a check for valid code, a value of 00 to 99. If the code is valid, a quick check will be made to see whether this person's age is consistent with his/her mother's age (if the person's mother is found in the household) and with this person's children's age (if this person is a woman and has children in the household). If the ages are inconsistent, this person's age will be blanked and the program will transfer back to the beginning of the age edit.

The edit will then determine whether this record is that of a married male (sex equals male, and marital status equals married). If so, a search will be made among the other records of the household for his wife (a woman has his line number as line number of husband). If no wife is found, the program goes to the next part of the edit. If his wife is found, the wife's age will be checked for validity (she must be at least 12 years old). The program now uses a special hot deck value. Based on the fact that the difference between ages of a husband and wife varies much less than the ages themselves, a hot deck in the program will store the difference in age (from previous records) of a husband and wife. This value is added to the age of the wife of this person to form a computed age. Just to be sure that this computed age is consistent with other characteristics, it will be compared with years in district and highest grade completed. Such a check is advisable because a computed age could give a person less years in age than the number of years the person has lived in the district.

When comparison with the age of his wife is not useful in determining the age of a man, the program will check relationship. If relationship is HEAD, the program will check the other records of the household (if there are any) for any with relationship equal to SON/DAUGHTER. Using an age difference, hot deck similar to the technique described above for husband and wife, the program will check the son or daughter's age and compute an age for this person. The computed age will then proceed through a consistency check to avoid obvious error by comparing the computed age with years in district and highest grade of school completed.

If the record being edited is that of a married woman, the hot deck technique using differences between the age of husband and wife will be used. Consistency checks will be made between the computed age and the age of her children, her years in the district, and her highest grade completed.

In the case of a female who is not a married woman with her husband present in the household, a search will be made for her eldest child (if she has any children) in the household. If any of her children are found, the hot deck technique of age differences

will be used to compute an age. This computed age will then be checked for consistency with her years in district and highest grade completed.

When a man or woman does not fall into one of the categories described above, the program will search for the person's mother in the household. If the person's mother is found, an age will be computed using the difference in age hot deck. The difference in age between child and mother varies much more than that between husband and wife and this is the reason that this edit is tried only after the husband/wife age difference technique is used. Numerous consistency checks are made between the computed age resulting from mother's age minus the hot deck value. The computed age will be checked against the educational characteristics, years in district, and highest grade completed; and if less than 12 years old, marital status, children, and economic activity. The last check presumes that a person is at least 12 years old if he/she has ever married, has children, or has an economic activity of any kind.

If a person whose age is missing has not fallen into any of the previous categories, and this person's relationship equals SON/DAUGHTER OF HEAD, a computed age will be made using the head's age and a hot deck value. This computed age then must pass a number of consistency checks before actually being assigned as the person's age.

If a person is neither a married male nor a married female with spouse present in the household, nor is the person a female with children, nor does the person have his/her mother in the household, nor is son or daughter of the head, then he/she is funneled into one of three groups based on other information in the person's record. If a person is head, has ever been married, has children, or has an economic activity, there is good reason to believe the person is equal to or greater than 12 years old. Based on that information, the value received from the hot deck should not be below 12 for this group. Similarly, if a person attends school, has any schooling, or can read and write, but is not head, has never been married, has no children, and has no economic activity, then this person will be placed in a group less than 12 years old but greater than or equal to 5. The hot deck value can then only be within that range. A similar group will be found for those less than 5 years old. Although not perfect, this technique limits the range of values that the hot deck can take.

3.7 *Line number of husband*

Line number of husband is relevant only for married females; therefore, for persons other than married females, entries will be eliminated. If this person is head, and has her husband present in the household, the program will check to determine whether P-6, line number of husband, actually equals the line number of the person who has SPOUSE for relationship. If it does not, the program will make line number of husband, the spouse's line number.

If this woman's husband is not present in the household, a check will be made to determine whether the line number of husband, P-6, equals 00, meaning the husband is living elsewhere. If it does, all is consistent and the program proceeds to the next edit. If it does not, the program assumes marital status is correct, that is, the woman is truly married and since no husband was found in the household, the program will assign 00 to line number of husband, assuming the husband lives elsewhere.

If this woman's relationship to head is SPOUSE, then the program will check to see

whether P-6, line number of husband equals the line number of the head. If it does, the program proceeds to the next edit. If it does not, the program will assign the line number of the head to P-6.

When a married females's relationship is neither HEAD nor SPOUSE, the line number of husband will be checked to see if it is 00. If it is, the program will proceed to the next edit. If it is not, then (since the woman is married and her husband is not living elsewhere) he should be living in the household, and thus his line number should be this woman's line number of husband. If it is not, the program assumes that her marital status is correct, that she truly is married, and assigns 00 to her line number of husband.

3.8 *Mother living and mother's line number*

The edits for these two items are shown separately in the flow charts in the exhibit. They are, however, interrelated. For persons who report other than YES for mother living, mother's line number will be checked to see whether it is 00 or whether it equals the line number of a female greater than or equal to 12 years old. If either of these cases are true, the program assumes the person has a mother and assigns YES to P-7. If neither of these cases are true, the program checks the entry for P-7. If it is NO or DOES NOT KNOW, the program skips to the edit for P-9. Otherwise it assigns DOES NOT KNOW as the entry.

Line number of mother is edited only for those persons who have a YES answer in P-7, indicating that they have a mother. If line number of mother equals 00, this is a valid answer and no other consistency check can be made. The program proceeds to the next edit. If the entry is not 00, a check will be made to see whether the line number of mother, P-8, equals the line number of someone in the household who really could be this person's mother, that is, a female who is 12 years or older. If so, the program proceeds to the next edit, but if not, the program assigns a value of 00, making the assumption that the person truly has a mother, but that she lives elsewhere.

3.9 *Migration*

The entries for country of birth and years lived in the district can be checked for consistency since there are strong relationships between the two items. There are also relationships between different members of the household and assumptions can be made from other family members as to whether or not this person has migrated.

3.91 Country of birth.--The entry for country of birth will only be altered if it is out of range (not 1 through 4). If the entry for years in district is ALWAYS, the code for Popstan will be assigned country of birth. If the entry is other than ALWAYS, information from other people in the household will be looked at for clues to this person's country of birth.

If the country of birth is blank or nonvalid and the years in district is other than ALWAYS, a search is made for the person's mother. If the mother is found in the household, the mother's years in district is examined. If mother's years in district is ALWAYS, this person's country of birth is assigned POPSTAN. If mother did not always live in district, but this person's age is less than or equal to the number of years that mother has lived in the district, the program assigns POPSTAN to country of birth. If this person's age is greater than the number of years mother has lived in the district, and mother's country of birth is valid, this person's country of birth is assigned the same country of birth as mother's.

If this person's mother is not in the household, but this person is a son or daughter of the head, then several checks can be made using information from the head's record. If the head's years in district is ALWAYS, the program will assign POPSTAN as country of birth to this person's record. If the head's years in district is not ALWAYS, but this person's age is less than or equal to the number of years that the head has spent in this district, the program will assign POPSTAN as this person's country of birth. If this person's age is more than the number of years the head has spent in this district, the program will assign the head's country of birth if it has a valid code of 1 through 4.

Quite different imputations can be made depending on whether or not a person is 12 years old or over. If a person is less than 12 years of age, country of birth will be hot decked from the first previous record for a child under 12 years old by age and sex. If the person is 12 or older and is female, the program will check to see if she has a husband in the household. If the woman has a husband and he has a valid country of birth code, the program will assign his country of birth code to her record. If the husband does not have a valid country or birth code, his years in district will be looked at. If his years in district is coded ALWAYS, this woman's country of birth will be assigned POPSTAN. If the husband's years in district is not ALWAYS, then the woman's country of birth will be hot decked by age and sex.

When a woman over 12 years old does not have a husband in the household, she may be the mother of a child in the household. In this case, the program will search for her eldest child. If the child cannot be found, the program hot decks country of birth by age and sex. If the child has a valid country of birth code and this person's (the child's mother's) years in district are greater than the child's age, the program will hot deck country of birth by age and sex. But if this person's years in district are less than or equal to the child's age, the program will use the child's country of birth to assign to this person.

When the person whose record the program is editing is not female, but male, the program can try to find the man's wife, or if he is the head of the family, the program will try to find the man's children. First the program attempts to find the man's wife. If found, and his years in the district are less than or equal to his wife's years in the district, the wife's country of birth is assigned to the man's record. If the man's years in district are greater than his wife's, country of birth will be imputed by age and sex using a hot deck. When the man is the head of the family and he has a son or daughter present in the household, and the man has been in the district for the same amount of time or less than the child's age, then the program will assign the same country of birth as his child's. If time in district is greater than his child's age, the program will hot deck by age.

3.92 Years lived in district.--Like country of birth, this question is important in compiling statistics on the mobility of the population. In some instances, a subgroup of the population may be far more mobile than the nation as a whole. The edit will take into consideration the person's place of birth and the responses for other members of the households.

If the years a person has lived in the district is ALWAYS, but the country of birth is not POPSTAN, the person's age will be assigned to years in district. It is assumed that although born outside of Popstan the person moved into the district when he/she

was less than 1 year of age. The next part of the edit checks for a valid entry (00 to 97) in years lived in district. Since the number of years a person has lived in the district cannot be greater than the person's age, age will be assigned to the years in district if this situation exists.

When the category does not have a valid code, the program will do an interrecord check by searching for the person's mother in the household. If found, the mother can provide information helpful in assigning missing values. If the person's mother has always lived in the district and her country of birth is POPSTAN (as it should be), the program will assign ALWAYS to this person's years in district category. If mother's country of birth is not POPSTAN even though her years in district is ALWAYS (something is wrong with mother's categories), the program will ignore mother's country of birth and assign age to years in district. If mother's years in district is not ALWAYS, but is a valid code (00 through 97) and the person's age is less than the number of years the mother has lived in the district, the edit will go back and check mother's country of birth. If mother's country of birth is POPSTAN, the program will assign ALWAYS to this persons years in district. If mother's country of birth is not POPSTAN, the program will assign this person's age to years in district. However, if a person's age is equal to or greater than mother's years in district, the program will assign mother's years in district to this person's years in district.

When the person's record being edited is a son or daughter of the head, it is sometimes useful to check the head's record for possible information to aid in assigning values for missing data. When the head was born in Popstan and always lived in this district, the program will assign ALWAYS to this person's years in district. When the head always lived in the district, but was not born in Popstan, this person's age will be assigned to district. When the head's entry for years in district is not ALWAYS, but is a valid code, this information can be used if it is consistent with the age of the record of the person being edited. If the person's age is equal to or greater than the number of years in the district of the head, the program will use the head's years in district as the years in district of the son or daughter (this person). If this person's age is less than the head's years in district, the program will assign a value depending on the country of birth of the head, ALWAYS if the head was born in POPSTAN; if it was not Popstan, the program assigns the person's age to years in district.

When all of the above efforts fail to produce a good value, the program will assign NOT REPORTED to years in district for this person.

3.10 Education and literacy

The edits for education and literacy apply only to people who are 5 years or older. Interrecord checks provide very little assistance in providing evidence for a person's educational level. Therefore, edits for school attendance, highest grade completed, and ability to read and write (literacy) will not be presented here since they will be no different than those provided in Part C. (Refer to chapter C-6, section 6.7, and edit P-11 found in exhibit C-6-2 in the appendix for Part C.)

3.11 Fertility

The questions on fertility, items P-14 through P-17, are asked of all females 12 years or older. A woman now past childbearing age will still have children ever born. Item P-18 is asked only of females 12 to 49

years old. Although some women can bear children past the age of 49, the event is so rare it is considered insignificant.

The edits for questions on fertility can make little use of interrecord checks. For this reason, no alternative edits will be shown here for fertility items. The reader is referred to chapter C-6, section 6.8 and exhibit C-6-2, edits P-14 to P-18, which may be found in the appendix of Part C.

3.12 Economic characteristics

There is relatively no relationship between the working status of one member of a household and another. For this reason, examining the records of other people in the household would only make edits for economic characteristics longer and more complicated at little or no gain in the quality of the data. The edits presented in Pact C, chapter C-6, section 6.9 would not change since no interrecord checks would be made. The reader is referred to the above section in Part C and the accompanying flow charts which are found in exhibit C-6-2 in the appendix.

4. SERIAL NUMBER

After the program has edited each item in all the records for a housing unit, a new serial number will be assigned to the housing record and each population record. This precaution is taken to avoid difficulties which may occur when the records are sorted before being tabulated. Key operators occasionally enter the wrong serial number through a transcription error resulting in two sets of records, actually distinct, which have the same serial number.

APPENDIX

Exhibit F-3-1. EA BOOK SHORT-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN

A. BLOCK NUMBER	B. SERIAL NUMBER	C. INMATES 1 ☐ Yes 2 ☐ No	FORM PH-2 (JULY 1979)

POPSTAN 1980 CENSUSES OF POPULATION AND HOUSING
SHORT FORM

MINISTRY OF ECONOMY
NATIONAL STATISTICAL OFFICE

CONFIDENTIAL - This inquiry is required by law. The information is accorded confidential treatment and cannot be used for taxation, investigation, or regulation.

ALL PERSONS

LINE NUMBER (E)	WHAT IS THE NAME OF THE HEAD OF THIS HOUSEHOLD? NAMES OF ALL OTHER PERSONS WHO USUALLY LIVE HERE? BE SURE TO INCLUDE BABIES AND ELDERLY PERSONS. MEMBERS OF THE HOUSEHOLD AWAY ON VACATION? ON BUSINESS? IN THE HOSPITAL? ANY PERSONS WHO HAVE NO USUAL RESIDENCE ELSEWHERE? *List in this order: Head, Spouse of head, Never married children of head or spouse (by age), Ever married children of head or spouse and their families (by age), Other relatives of head or spouse, Nonrelatives of head or spouse* P-1	WHAT IS ...'S RELATIONSHIP TO THE HEAD OF THIS HOUSEHOLD? 1-Head 2-Spouse of head 3-Son or daughter of head/spouse 4-Other relative 5-Non-relative P-2	SEX 1-Male 2-Female P-3	AGE AS OF LAST BIRTHDAY *If under 1 year, circle 00.* P-4	MARITAL STATUS 1-Married 2-Divorced 3-Separated 4-Widowed 5-Never married *If MALE or code 2, 3, 4, or 5, skip to P-7.* P-5	*MARRIED FEMALES* *If husband lives in this HU, enter his line number.* *If husband lives elsewhere, circle 00.* P-6	IS ...'S MOTHER LIVING? 1-Yes 2-No 3-Does not know *If NO or DOES NOT KNOW, end interview for this person.* P-7	*If mother lives in this HU, enter her line number.* *If mother lives elsewhere, circle 00.* P-8
01		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00
02		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00
03		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00
04		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00
05		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00

ALL PERSONS

LINE NUMBER (E)	WHAT IS THE NAME OF THE HEAD OF THIS HOUSEHOLD? NAMES OF ALL OTHER PERSONS WHO USUALLY LIVE HERE? BE SURE TO INCLUDE BABIES AND ELDERLY PERSONS. MEMBERS OF THE HOUSEHOLD AWAY ON VACATION? ON BUSINESS? IN THE HOSPITAL? ANY PERSONS WHO HAVE NO USUAL RESIDENCE ELSEWHERE? *List in this order: Head, Spouse of head, Never married children of head or spouse (by age), Ever married children of head or spouse and their families (by age), Other relatives of head or spouse, Nonrelatives of head or spouse* P-1	WHAT IS ...'S RELATIONSHIP TO THE HEAD OF THIS HOUSEHOLD? 1-Head 2-Spouse of head 3-Son or daughter of head/spouse 4-Other relative 5-Non-relative P-2	SEX 1-Male 2-Female P-3	AGE AS OF LAST BIRTHDAY *If under 1 year, circle 00.* P-4	MARITAL STATUS 1-Married 2-Divorced 3-Separated 4-Widowed 5-Never married *If MALE or code 2, 3, 4, or 5, skip to P-7.* P-5	*MARRIED FEMALES* *If husband lives in this HU, enter his line number.* *If husband lives elsewhere, circle 00.* P-6	IS ...'S MOTHER LIVING? 1-Yes 2-No 3-Does not know *If NO or DOES NOT KNOW, end interview for this person.* P-7	*If mother lives in this HU, enter her line number.* *If mother lives elsewhere, circle 00.* P-8
06		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00
07		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00
08		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00
09		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00
10		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00

P-1a. If household continued, mark X in this circle. ◯

P-1b. TOTAL PERSONS--I have listed _____ persons. Have I missed anyone? *If YES and person is usual resident or has no usual residence elsewhere, add to P-1.*

P-1c. If only nonresidents are living here, mark X in this circle. ◯ *Complete items H-1 to H-10 below.*

NSO USE ONLY | P-1d. Number of lines cancelled. | P-1e. Population count after edit.

ALL HOUSING UNITS

H-1. TYPE OF LIVING QUARTERS
1 - Conventional HU
2 - Improvised HU (makeshift shelter of waste materials, barn, cave, warehouse, etc.)
3 - Mobile HU (tent, boat, wagon, etc.)
4 - Collective quarters } *End interview.*
5 - None (for homeless) }

H-2. NUMBER OF HOUSING UNITS IN THIS HOUSE (BUILDING)
1 - 1 unit, detached
2 - 1 unit, attached
3 - 2 to 4 units
4 - 5 to 9 units
5 - 10 units or more

H-3. WALLS - Main construction material of outer walls of this house (building)
1 - Stone, cement, stucco, brick
2 - Metal
3 - Wood
4 - Bamboo, leaves, reed, mud
5 - Other material

H-4. ROOF - Main construction material of roof on this house (building)
1 - Stone, cement, brick, tile, slate
2 - Metal
3 - Wood
4 - Bamboo, leaves, reed, mud
5 - Other material or none

H-5. FLOORS - Main construction material of floors in this housing unit
1 - Stone, cement, lime
2 - Tile, slate, marble
3 - Wood
4 - Leaves, reed, mud
5 - Other material

H-6. HOW MANY ROOMS ARE IN THIS UNIT?
_____ rooms

H-7. IS THIS UNIT OWNED BY SOMEONE LIVING IN IT OR IS IT RENTED?
OCCUPIED:
1 - Owned or being bought - *End interview.*
2 - Rented for cash
3 - Rent free or other arrangement - *End interview.*
4 - VACANT - *Skip to H-9.*

H-8. WHAT IS THE AMOUNT OF RENT PAID MONTHLY?
Round to the nearest goldar.
_____ goldars
End interview.

VACANT UNITS

H-9. VACANCY STATUS
1 - For rent or sale
2 - Not for rent or sale

H-10. CONDITION OF VACANT UNIT
1 - Habitable for year-round use
2 - Habitable for seasonal use
3 - Not habitable

PRINCIPAL RESPONDENT

Line number of principal respondent	Date of interview

REMARKS

Exhibit F-3-2. EA BOOK LONG-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN

A. BLOCK NUMBER | B. SERIAL NUMBER | C. INMATES 1 ☐ Yes 2 ☐ No | FORM PH-3 (JULY 1979)

POPSTAN 1980 CENSUSES OF POPULATION AND HOUSING

LONG FORM

MINISTRY OF ECONOMY
NATIONAL STATISTICAL OFFICE

CONFIDENTIAL - This inquiry is required by law. The information is accorded confidential treatment and cannot be used for taxation, investigation, or regulation.

LINE NUMBER (E)	ALL PERSONS: WHAT IS THE NAME OF THE HEAD OF THIS HOUSEHOLD? NAMES OF ALL OTHER PERSONS WHO USUALLY LIVE HERE? BE SURE TO INCLUDE BABIES AND ELDERLY PERSONS. MEMBERS OF THE HOUSEHOLD AWAY ON VACATION? ON BUSINESS? IN THE HOSPITAL? ANY PERSONS WHO HAVE NO USUAL RESIDENCE ELSEWHERE? *List in this order: Head; Spouse of head; Never married children of head or spouse (by age); Ever married children of head or spouse and their families (by age); Other relatives of head or spouse; Nonrelatives of head or spouse* P-1	WHAT IS ...'S REALTIONSHIP TO THE HEAD OF THIS HOUSEHOLD? 1-Head 2-Spouse of head 3-Son or daughter of head/spouse 4-Other relative 5-Non-relative P-2	SEX 1-Male 2-Female P-3	AGE AS OF LAST BIRTHDAY *If under 1 year, circle 00.* P-4	MARITAL STATUS 1-Married 2-Divorced 3-Separated 4-Widowed 5-Never married *If MALE or code 2, 3, 4, or 5, skip to P-7.* P-5	*MARRIED FEMALES* *If husband lives in this HU, enter his line number.* *If husband lives elsewhere circle 00.* P-6	IS ...'S MOTHER LIVING? 1-Yes 2-No 3-Does not know *If NO or DOES NOT KNOW, skip to P-9.* P-7	*If mother lives in this HU, enter her line number.* *If mother lives elsewhere circle 00.* P-8	IN WHAT COUNTRY WAS ... BORN? 1-Popstan 2-Endar 3-Victoria 4-Other P-9	HOW MANY YEARS HAS ... LIVED CONTINUOUSLY IN THIS DISTRICT? *If under 1 year, circle 00.* *If always, circle 98.* P-10	PERSONS 5 YEARS AND OLDER: DOES ... ATTEND SCHOOL? 1-Yes 2-No *Omit if 35 years or older.* P-11	WHAT IS THE HIGHEST GRADE OR YEAR THAT ... HAS COMPLETED? None (N) 00; Primary (P) 01 02 03 04 05; Middle (M) 06 07 08; High school (H) 09 10 11 12; University (U) 13 14 15 16 17+ *If 04 or more, skip to P-14.* P-12	CAN ... READ AND WRITE? 1-Yes 2-No P-13	FEMALES 12 AND OLDER: HOW MANY CHILDREN HAVE BEEN BORN ALIVE TO ...? *If NONE, circle 00 and skip to P-16.* P-14	HOW MANY OF THESE CHILDREN ARE STILL LIVING? *If NONE circle 00.* P-15	PERSONS 12 YEARS AND OLDER: DURING JUNE WHAT WAS ... DOING MOST OF THE TIME-- 1-Working *Skip to P-20.* 2-Own housework 3-Going to school 4-Something else P-16	During June -- DID ... DO ANY WORK AT ALL NOT INCLUDING OWN HOUSEWORK? 1-Yes *Skip to P-20.* 2-No P-17	WAS ... LOOKING FOR WORK? 1-Yes *End interview for this person.* 2-No P-18	WHY WAS ... NOT LOOKING FOR WORK? 1-Had job *Skip to P-21.* 2-Job not available 3-Did not want job 4-Unable to work *If 2, 3, or 4 end interview for this person.* P-19	HOW MANY HOURS PER WEEK DID ... USUALLY WORK AT ALL JOBS, NOT INCLUDING OWN HOUSEWORK? P-20	WHAT WAS ...'S PRINCIPAL JOB? *Describe the specific kind of work that ... was doing such as: making clay pots, servant, driving a truck, midwife, farm helper, etc.* P-21	IN WHAT KIND OF BUSINESS OR INDUSTRY WAS THIS? *Describe the specific type of business or industry such as: farm, bus company, coal mine, automobile repair shop, Armed Forces, etc.* P-22	WAS ...-- 1-Employer (ER) 2-Employee for pay (EE) 3-Self-employed or own account worker 4-Unpaid family worker 5-Armed Forces (AF) P-23
01		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00; P 01 02 03 04 05; M 06 07 08; H 09 10 11 12; U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF
02		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00; P 01 02 03 04 05; M 06 07 08; H 09 10 11 12; U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF
03		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00; P 01 02 03 04 05; M 06 07 08; H 09 10 11 12; U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF
04		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00; P 01 02 03 04 05; M 06 07 08; H 09 10 11 12; U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF
05		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00; P 01 02 03 04 05; M 06 07 08; H 09 10 11 12; U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF
06		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00; P 01 02 03 04 05; M 06 07 08; H 09 10 11 12; U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF
07		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00; P 01 02 03 04 05; M 06 07 08; H 09 10 11 12; U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF
08		1- Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00; P 01 02 03 04 05; M 06 07 08; H 09 10 11 12; U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF

P-1a. If household continued on back, mark X in this circle. ⟶ ◯

P-1b. TOTAL PERSONS--I have listed _____ persons. Have I missed anyone? *If YES and person is usual resident or has no usual residence elsewhere, add to P-1.*

P-1c. If only nonresidents are living here, mark X in this circle. ⟶ ◯ *Complete items H-1 to H-10 on back.*

NSO USE ONLY | P-1d. Number of lines cancelled | P-1e. Population count after edit

Exhibit F-3-2. EA BOOK LONG-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN--Continued

	ALL PERSONS										PERSONS 5 YEARS AND OLDER			FEMALES 12 AND OLDER		PERSONS 12 YEARS AND OLDER								
																	During June --							
LINE NUMBER (E)	WHAT IS THE NAME OF THE HEAD OF THIS HOUSEHOLD? NAMES OF ALL OTHER PERSONS WHO USUALLY LIVE HERE? BE SURE TO INCLUDE BABIES AND ELDERLY PERSONS. MEMBERS OF THE HOUSEHOLD AWAY ON VACATION? ON BUSINESS? IN THE HOSPITAL? ANY PERSONS WHO HAVE NO USUAL RESIDENCE ELSEWHERE? *List in this order:* *Head* *Spouse of head* *Never married children of head or spouse (by age)* *Ever married children of head or spouse and their families (by age)* *Other relatives of head or spouse* *Nonrelatives of head or spouse*	WHAT IS ...'S RELATIONSHIP TO THE HEAD OF THIS HOUSEHOLD? 1-Head 2-Spouse of head 3-Son or daughter of head/spouse 4-Other relative 5-Non-relative	SEX 1-Male 2-Female	AGE AS OF LAST BIRTHDAY *If under 1 year, circle 00.*	MARITAL STATUS 1-Married 2-Divorced 3-Separated 4-Widowed 5-Never married *If MALE or code 2, 3, 4, or 5, skip to P-7.*	*MARRIED FEMALES* *If husband lives in this HU, enter his line number.* *If husband lives elsewhere circle 00.*	IS ...'S MOTHER LIVING? 1-Yes 2-No 3-Does not know *If NO or DOES NOT KNOW, skip to P-9.*	*If mother lives in this HU, enter her line number.* *If mother lives elsewhere circle 00.*	IN WHAT COUNTRY WAS ... BORN? 1-Popstan 2-Endar 3-Victgria 4-Other	HOW MANY YEARS HAS ... LIVED CONTINUOUSLY IN THIS DISTRICT? *If under 1 year, circle 00.* *If always, circle 98.*	DOES ... ATTEND SCHOOL? 1-Yes 2-No *Omit if 35 years or older.*	WHAT IS THE HIGHEST GRADE OR YEAR THAT ... HAS COMPLETED? None (N) 00 Primary (P) 01 02 03 04 05 Middle (M) 06 07 08 High school (H) 09 10 11 12 University (U) 13 14 15 16 17+ *If 04 or more, skip to P-14.*	CAN ... READ AND WRITE? 1-Yes 2-No	HOW MANY CHILDREN HAVE BEEN BORN ALIVE TO ...? *If NONE, circle 00 and skip to P-16.*	HOW MANY OF THESE CHILDREN ARE STILL LIVING? *If NONE circle 00.*	DURING JUNE WHAT WAS ... DOING MOST OF THE TIME-- 1-Working *Skip to P-20.* 2-Own housework 3-Going to school 4-Something else	DID ... DO ANY WORK AT ALL NOT INCLUDING OWN HOUSEWORK? 1-Yes *Skip to P-20.* 2-No	WAS ... LOOKING FOR WORK? 1-Yes *End interview for this person.* 2-No	WHY WAS ... NOT LOOKING FOR WORK? 1-Had job *Skip to P-21.* 2-Job not available 3-Did not want job 4-Unable to work *If 2, 3, or 4 end interview for this person.*	HOW MANY HOURS PER WEEK DID ... USUALLY WORK AT ALL JOBS, NOT INCLUDING OWN HOUSEWORK?	WHAT WAS ...'S PRINCIPAL JOB? *Describe the specific kind of work that ... was doing such as: making clay pots, servant, driving a truck, midwife, farm helper, etc.*	IN WHAT KIND OF BUSINESS OR INDUSTRY WAS THIS? *Describe the specific type of business or industry such as: farm, bus company, coal mine, automobile repair shop, Armed Forces, etc.*	WAS ...-- 1-Employer (ER) 2-Employee for pay (EE) 3-Self-employed or own account worker 4-Unpaid family worker 5-Armed Forces (AF)
	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12	P-13	P-14	P-15	P-16	P-17	P-18	P-19	P-20	P-21	P-22	P-23
09		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00 P 01 02 03 04 05 M 06 07 08 H 09 10 11 12 U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF
10		1-Head 2-Spouse 3-Son/Dau 4-Oth rel 5-Nonrel	1 - M 2 - F	00	1-Mar 2-Div 3-Sep 4-Wid 5-NM	00	1-Yes 2-No 3-Does not know	00	1-Pop 2-End 3-Vic 4-Other	00 98	1-Yes 2-No	N 00 P 01 02 03 04 05 M 06 07 08 H 09 10 11 12 U 13 14 15 16 17+	1-Yes 2-No	00	00	1-Work 2-Own hswk 3-School 4-Something else	1-Yes 2-No	1-Yes 2-No	1-Job 2-No job available 3-Not want 4-Unable				1-ER 2-EE 3-Self 4-Unpaid 5-AF

P-1e. If household continued, mark X in this circle. ◯

ALL HOUSING UNITS

H-1. TYPE OF LIVING QUARTERS
1 - Conventional HU
2 - Improvised HU (makeshift shelter of waste materials, barn, cave, warehouse, etc.)
3 - Mobile HU (tent, boat, wagon, etc.)
4 - Collective quarters
5 - None (for homeless)
Skip to H-19.

H-2. NUMBER OF HOUSING UNITS IN THIS HOUSE (BUILDING)
1 - 1 unit, detached
2 - 1 unit, attached
3 - 2 to 4 units
4 - 5 to 9 units
5 - 10 units or more

H-3. WALLS - Main construction material of outer walls of this house (building)
1 - Stone, cement, stucco, brick
2 - Metal
3 - Wood
4 - Bamboo, leaves, reed, mud
5 - Other material

H-4. ROOF - Main construction material of roof on this house (building)
1 - Stone, cement, brick, tile, slate
2 - Metal
3 - Wood
4 - Bamboo, leaves, reed, mud
5 - Other material or none

H-5. FLOORS - Main construction material of floors in this housing unit
1 - Stone, cement, lime
2 - Tile, slate, marble
3 - Wood
4 - Leaves, reed, mud
5 - Other material

H-6. HOW MANY ROOMS ARE IN THIS UNIT?
_____ rooms

H-7. IS THIS UNIT OWNED BY SOMEONE LIVING IN IT OR IS IT RENTED?
OCCUPIED:
1 - Owned or being bought - *Skip to H-11.*
2 - Rented for cash
3 - Rent free or other arrangement - *Skip to H-11.*
4 - VACANT - *Skip to H-9.*

H-8. WHAT IS THE AMOUNT OF RENT PAID MONTHLY?
Round to the nearest goldar.
_____ goldars
Skip to H-11.

VACANT UNITS

H-9. VACANCY STATUS
1 - For rent or sale
2 - Not for rent or sale

H-10. CONDITION OF VACANT UNIT
1 - Habitable for year-round use
2 - Habitable for seasonal use
3 - Not habitable

H-11. WAS THIS UNIT CONSTRUCTED IN THE LAST 12 MONTHS?
1 - Yes
2 - No

H-12. IS THERE PIPED RUNNING WATER FOR THIS UNIT?
1 - Piped water inside building (structure)
2 - Piped water outside building (structure) within 100 meters
3 - Piped water outside building (structure) beyond 100 meters
4 - No piped water

H-13. WHAT IS THE SOURCE OF DRINKING WATER FOR THIS UNIT?
1 - Pipe or pump inside building
2 - Pipe or pump outside building
3 - Bottled or canned water
4 - Closed well or closed spring
5 - Open well or spring
6 - River, lake, other

H-14. WHAT TYPE OF TOILET FACILITIES ARE AVAILABLE FOR THIS UNIT?
1 - Flush toilet for exclusive use of the occupants of this HU
2 - Flush toilet shared with occupants of another HU
3 - Out-house or out-building
4 - Covered pit
5 - Open pit, ditch, pail
6 - Other

H-15. WHAT TYPE OF BATHING FACILITIES ARE AVAILABLE FOR THIS UNIT?
1 - Bathtub or shower with piped water for exclusive use of the occupants of this HU
2 - Bathtub or shower with piped water shared with occupants of another HU
3 - Hand basin with piped water
4 - Portable tub or basin
5 - Other

H-16. IS THERE A TELEVISION SET IN THIS UNIT?
1 - Yes
2 - No

H-17. WHAT KIND OF LIGHTING DOES THIS UNIT HAVE?
1 - Electric
2 - Gas
3 - Kerosene
4 - Other or none

H-18. WHAT KIND OF FUEL IS USED MOST IN THIS UNIT FOR COOKING?
1 - Electricty
2 - Gas
3 - Kerosene or oil
4 - Coal or charcoal
5 - Wood
6 - Other or none

H-19. ARE ANY ARTICLES PRODUCED IN THIS UNIT FOR SALE OR BARTER - NOT INCLUDING AGRICULTURAL PRODUCTS?
1 - Yes
2 - No

PRINCIPAL RESPONDENT

Line number of principal respondent	Date of interview

REMARKS

Exhibit F-3-3. BOOKLET SHORT-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN

FORM PH-4
(August 1979)

MINISTRY OF ECONOMY
NATIONAL STATISTICAL OFFICE

BOOK ____ OF ____

POPSTAN 1980 CENSUSES OF POPULATION AND HOUSING
SHORT FORM

CONFIDENTIALITY - This inquiry is required by law. The information is accorded confidential treatment and cannot be used for taxation, investigation, or regulation.

NOTICE TO FINDER — This book is the property of the Popstan Government. If found, please return to the nearest post office.

POSTMASTER: Return to the National Statistical Office, Ministry of Economy, St. Regis.

A. PROVINCE	CODE	B. DISTRICT	CODE
C. BARRIO/VILLAGE	CODE	D. PLACE	CODE
E. WARD/TRACT CODE	F. BLOCK NO.	G. CL AREA NO.	H. TYPE OF PLACE
I. METRO AREA	J. EA NO.	K. SERIAL NO.	L. INMATES 1 ☐ Yes 2 ☐ No

M. ADDRESS OR LOCATION

N. NOTES:

PRINCIPAL RESPONDENT

LINE NUMBER OF PRINCIPAL RESPONDENT

DATE OF INTERVIEW

ENUMERATOR'S NAME

OFFICE USE ONLY

I. GENERAL EDITING AND CODING	Date	Name
a. Editing and coding.....		
b. Verification.....		
c. Control counts		
(1) Cancelled lines.....		
(2) Uncancelled lines.....		

Exhibit F-3-3. BOOKLET SHORT-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN--Continued

LINE NUMBER (1)	WHAT IS THE NAME OF THE HEAD OF THIS HOUSEHOLD? NAMES OF ALL OTHER PERSONS WHO USUALLY LIVE HERE? BE SURE TO INCLUDE BABIES AND ELDERLY PERSONS. MEMBERS OF THE HOUSEHOLD AWAY ON VACATION? ON BUSINESS? IN THE HOSPITAL? ANY PERSONS WHO HAVE NO USUAL RESIDENCE ELSEWHERE? *List in this order:* *Head* *Spouse of head* *Never married children of head/spouse (by age)* *Ever married children of head/spouse and their families (by age)* *Other relatives of head/spouse* *Nonrelatives* (2)	WHAT IS . . .'S RELATIONSHIP TO THE HEAD OF THIS HOUSEHOLD? *Interpret categories 3 to 7 to mean relationship to head or spouse.* 1 Head 2 Spouse of head 3 Son/daughter 4 Spouse of son/daughter 5 Grandchild or great-grandchild 6 Parent 7 Other relative 8 Nonrelative (3)	SEX (4)	AGE AS OF LAST BIRTHDAY *If under 1 year, circle 00.* (5)
01	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00
02	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00
03	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00
04	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00
05	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00
06	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00
07	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00
08	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00
09	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00
10	Family name Given name	1 Head 5 Grandchild (or great) 2 Spouse 6 Parent 3 Son/dau 7 Other relative 4 SP of son/dau 8 Nonrelative	1 M 2 F	00

ENUMERATOR INSTRUCTIONS	*After listing the household and completing items 2 through 3, ASK the following question:*	(13) I have listed _____ persons. *Name each one.* Have I missed anyone? *If YES and person is usual resident or has no usual residence elsewhere, add to item 2.*

Exhibit F-3-3. BOOKLET SHORT-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN--Continued

MARITAL STATUS	MARRIED FEMALES	IS . . .'S MOTHER LIVING?		IS . . .'S FATHER LIVING?		LINE NUMBER
1 Married 4 Widowed 2 Divorced 5 Never married 3 Separated *If MALE or code 2, 3, 4, or 5, skip to 8.*	*If husband lives in this HU, enter his line number.* *If husband lives elsewhere, circle 00.*	1 Yes 2 No 3 Does not know *If NO or DOES NOT KNOW, skip to 10.*	*If mother lives in this HU, enter her line number; if lives elsewhere, circle 00.*	1 Yes 2 No 3 Does not know *If NO or DOES NOT KNOW, end interview for this person.*	*If father lives in this HU, enter his line number; if lives elsewhere; circle 00.*	
(6)	(7)	(8)	(9)	(10)	(11)	(12)
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	01
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	02
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	03
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	04
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	05
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	06
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	07
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	08
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	09
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	10

(14) *If listing is continued on additional questionnaire, mark X in this circle.* ○

(15) *If only nonresidents are living here, mark X in this circle.* ○
Complete items H-1 to H-10 on back.

Exhibit F-3-3. BOOKLET SHORT-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN--Continued

ALL HOUSING UNITS

H-1. TYPE OF LIVING QUARTERS

1 - Conventional HU

2 - Improvised HU (makeshift shelter of waste materials, barn, cave, warehouse, etc.)

3 - Mobile HU (tent, boat, wagon, etc.)

4 - Collective quarters
5 - None (for homeless) } *End interview.*

H-2. NUMBER OF HOUSING UNITS IN THIS HOUSE (BUILDING)

1 - 1 unit, detached

2 - 1 unit, attached

3 - 2 to 4 units

4 - 5 to 9 units

5 - 10 units or more

H-3. WALLS - Main construction material of outer walls of this house (building)

1 - Stone, cement, stucco, brick

2 - Metal

3 - Wood

4 - Bamboo, leaves, reed, mud

5 - Other material

H-4. ROOF - Main construction material of roof on this house (building)

1 - Stone, cement, brick, tile, slate

2 - Metal

3 - Wood

4 - Bamboo, leaves, reed, mud

5 - Other material or none

H-5. FLOORS - Main construction material of floors in this housing unit

1 - Stone, cement, lime

2 - Tile, slate, marble

3 - Wood

4 - Leaves, reed, mud

5 - Other material

H-6. HOW MANY ROOMS ARE IN THIS UNIT?

_____ rooms

H-7. IS THIS UNIT OWNED BY SOMEONE LIVING IN IT OR IS IT RENTED?

OCCUPIED:

1 - Owned or being bought - *End interview.*

2 - Rented for cash

3 - Rent free or other arrangement - *End interview.*

4 - VACANT - *Skip to H-9*

H-8. WHAT IS THE AMOUNT OF RENT PAID MONTHLY?

Round to the nearest goldar.

_____goldars

End interview.

VACANT UNITS

H-9. VACANCY STATUS

1 - For rent or sale

2 - Not for rent or sale

H-10. CONDITION OF VACANT UNIT

1 - Habitable for year-round use

2 - Habitable for seasonal use

3 - Not habitable

Notes

Exhibit F-3-4. BOOKLET LONG-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN

FORM PH-5
(August 1979)

MINISTRY OF ECONOMY
NATIONAL STATISTICAL OFFICE

BOOK ____ OF ____

CONFIDENTIALITY - This inquiry is required by law. The information is accorded confidential treatment and cannot be used for taxation, investigation, or regulation.

POPSTAN 1980 CENSUSES OF POPULATION AND HOUSING
LONG FORM

NOTICE TO FINDER — This book is the property of the Popstan Government. If found, please return to the nearest post office.

POSTMASTER: Return to the National Statistical Office, Ministry of Economy, St. Regis.

A. PROVINCE	CODE	B. DISTRICT	CODE
C. BARRIO/VILLAGE	CODE	D. PLACE	CODE

E. WARD/TRACT CODE	F. BLOCK NO.	G. CL AREA NO.	H. TYPE OF PLACE
I. METRO AREA	J. EA NO.	K. SERIAL NO.	L. INMATES 1 ☐ Yes 2 ☐ No

M. ADDRESS OR LOCATION

N. NOTES

PRINCIPAL RESPONDENT

LINE NUMBER OF PRINCIPAL RESPONDENT

DATE OF INTERVIEW

ENUMERATOR'S NAME

OFFICE USE ONLY

	Date	Name
1. GENERAL EDITING AND CODING		
a. Editing and coding.....		
b. Verification.....		
c. Control counts		
(1) Cancelled lines.....		
(2) Uncancelled lines.....		
2. OCCUPATION AND INDUSTRY CODING	Date	Name
a. Coding.....		
b. Verification.....		

BEST AVAILABLE COPY

Exhibit F-3-4. BOOKLET LONG-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN--Continued

LINE NUMBER (1)	WHAT IS THE NAME OF THE HEAD OF THIS HOUSEHOLD? NAMES OF ALL OTHER PERSONS WHO USUALLY LIVE HERE? BE SURE TO INCLUDE BABIES AND ELDERLY PERSONS. MEMBERS OF THE HOUSEHOLD AWAY ON VACATION? ON BUSINESS? IN THE HOSPITAL? ANY PERSONS WHO HAVE NO USUAL RESIDENCE ELSEWHERE? *List in this order:* *Head* *Spouse of head* *Never married children of head/spouse (by age)* *Ever married children of head/spouse and their families (by age)* *Other relatives of head/spouse* *Nonrelatives* (2)	WHAT IS . . .'S RELATIONSHIP TO THE HEAD OF THIS HOUSEHOLD? *Interpret categories 3 to 7 to mean relationship to head or spouse.* 1 Head 2 Spouse of head 3 Son/daughter 4 Spouse of son/daughter 5 Grandchild or great-grandchild 6 Parent 7 Other relative 8 Nonrelative (3)	SEX (4)	AGE AS OF LAST BIRTHDAY *If under 1 year, circle 00.* (5)
01	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00
02	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00
03	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00
04	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00
05	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00
06	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00
07	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00
08	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00
09	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00
10	Family name Given name	1 Head 2 Spouse 3 Son/dau 4 SP of son/dau 5 Grandchild (or great) 6 Parent 7 Other relative 8 Nonrelative	1 M 2 F	00

ENUMERATOR INSTRUCTIONS	*After listing the household and completing items 2 through 3, ASK the following question:*	(13) I have listed _____ persons. *Name each one.* Have I missed anyone? *If YES and person is usual resident or has no usual residence elsewhere, add to item 2.*

Exhibit F-3-4. BOOKLET LONG-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN--Continued

MARITAL STATUS	MARRIED FEMALES	IS . . .'S MOTHER LIVING?		IS . . .'S FATHER LIVING?		LINE NUMBER
1 Married 4 Widowed 2 Divorced 5 Never married 3 Separated *If MALE or code 2, 3, 4, or 5, skip to 8.*	*If husband lives in this HU, enter his line number.* *If husband lives elsewhere, circle 00.*	1 Yes 2 No 3 Does not know *If NO or DOES NOT KNOW, skip to 10.*	*If mother lives in this HU, enter her line number; if lives elsewhere, circle 00.*	1 Yes 2 No 3 Does not know *If NO or DOES NOT KNOW, end interview for this person.*	*If father lives in this HU, enter his line number; if lives elsewhere; circle 00.*	
(6)	(7)	(8)	(9)	(10)	(11)	(12)
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	01
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	02
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	03
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	04
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	05
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	06
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	07
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	08
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	09
1 Mar 4 Wid 2 Div 5 NM 3 Sep	00	1 Yes 2 No 3 Does not know	00	1 Yes 2 No 3 Does not know	00	10

(14) *If listing is continued on additional questionnaire, mark X in this circle.* ○

(15) *If only nonresidents are living here, mark X in this circle.* ○
Complete items H-1 to H-10 on back.

Exhibit F-3-4. BOOKLET LONG-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN--Continued

a. Line No.	b. Last name	c. First name			
FOR NSO USE	d. Relationship	e. Sex	f. Age	g. Marital status	h. Line No. of spouse
	MOTHER		FATHER		
	i. Living	j. Line No.	k. Living	l. Line No.	

16. IN WHAT PROVINCE WAS . . . BORN?

Province

03	Artesia	18	Hali	33	Matanga
06	Copal	21	Kerac	36	Patan
09	Dari	24	Lacuna	39	Rift
12	Eris	27	Laya	42	Terra
15	Girda	30	Lira	45	Tumar

Foreign country

46	Endar	48	Congo	50	Roco	52	Other
47	Victoria	49	Bolu	51	Malay		

17. HOW LONG HAS . . . LIVED CONTINUOUSLY IN . . . *(name)* DISTRICT?

1 Always; born here - *Skip to 19*
2 Less than 1 year
3 1 up to 2 years
4 2 up to 5 years
5 5 years or more

18. IN WHAT PROVINCE DID . . . LIVE BEFORE MOVING TO THIS DISTRICT?

Province

03	Artesia	18	Hali	33	Matanga
06	Copal	21	Kerac	36	Patan
09	Dari	24	Lacuna	39	Rift
12	Eris	27	Laya	42	Terra
15	Girda	30	Lira	45	Tumar

Foreign country

46	Endar	48	Congo	50	Roco	52	Other
47	Victoria	49	Bolu	51	Malay		

19. WHAT LANGUAGE DOES . . . SPEAK MOST OFTEN AT HOME? *For children under 2, indicate the language most likely to learn.*

1	Popstani	3	Womba	5	Sino	7	Victorian
2	Betti	4	Ronda	6	Endari	8	Other

If Popstani and person is under 12 years, END INTERVIEW for this person. If Popstani and person is 12 or older, skip to 21.

20. DOES . . . SPEAK POPSTANI?

1 Yes
2 No

If person is under 12 years, END INTERVIEW for this person.

ALL PERSONS 12 YEARS OR OLDER

21. HAS . . . EVER ATTENDED SCHOOL?

1 Yes
2 No - *If female, skip to 24; if male skip to 43.*

22. WHAT IS THE HIGHEST GRADE OR YEAR THAT . . . HAS ATTENDED?

Primary	01	02	03	04	05
Middle	06	07	08		
High school	09	10	11	12	
University	13	14	15	16	17+

23. DID . . . COMPLETE THIS GRADE OR YEAR?

1 Now attending this grade/year
2 Completed this grade/year
3 Did not complete this grade/year

If MALE, skip to 43.

ALL FEMALES 12 YEARS OR OLDER

24. HOW MANY CHILDREN HAVE BEEN BORN ALIVE TO . . .?

00 None - *Skip to 39*

_____ Children

25. HOW MANY SONS HAVE BEEN BORN ALIVE TO . . .?

00 None - *Skip to 29*

_____ Sons

OF THESE SONS, HOW MANY -

26. ARE LIVING IN THIS HOUSING UNIT? _____
27. ARE LIVING ELSEWHERE? _____
28. HAVE DIED? _____

29. HOW MANY DAUGHTERS HAVE BEEN BORN ALIVE TO . . .?

00 None - *Skip to 33*

_____ Daughters

OF THESE DAUGHTERS, HOW MANY -

30. ARE LIVING IN THIS HOUSING UNIT? _____
31. ARE LIVING ELSEWHERE? _____
32. HAVE DIED? _____

33. IN WHAT YEAR DID . . .'S LAST LIVE BIRTH OCCUR?

_____ Year

34. IN WHAT MONTH DID . . .'S LAST LIVE BIRTH OCCUR?

01	January	05	May	09	September
02	February	06	June	10	October
03	March	07	July	11	November
04	April	08	August	12	December

35. WAS THIS CHILD A BOY OR GIRL?

1 Boy
2 Girl

36. IS THIS CHILD STILL LIVING?

1 Yes - *Skip to 39*
2 No

37. IN WHAT YEAR DID HE/SHE DIE?

_____ Year

38. IN WHAT MONTH DID HE/SHE DIE?

01	January	05	May	09	September
02	February	06	June	10	October
03	March	07	July	11	November
04	April	08	August	12	December

39. MARITAL STATUS

1 Married, divorced, separated, widowed
2 Never married - *Skip to 43*

40. HAS . . . BEEN MARRIED MORE THAN ONCE?

1 Yes
2 No - *Skip to 42*

41. DID . . .'S FIRST MARRIAGE END BECAUSE OF THE DEATH OF HER HUSBAND?

1 Yes
2 No

42. DATE OF (FIRST) MARRIAGE

_____ Month and year

BEST AVAILABLE COPY

Exhibit F-3-4. BOOKLET LONG-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POSPTAN--Continued

ALL PERSONS 12 YEARS OR OLDER

43. DURING JUNE WHAT WAS . . . DOING MOST OF THE TIME?

1 Working - *Skip to 51*
2 Own housework
3 Going to school
4 On vacation
5 Other - *Specify* _____

44. DID . . . DO ANY WORK AT ALL DURING JUNE NOT INCLUDING OWN HOUSEWORK?

1 Yes - *Skip to 51*
2 No

45. EVEN THOUGH . . . DID NOT WORK DURING JUNE, DID HE/SHE HAVE A JOB OR BUSINESS?

1 Yes
2 No - *Skip to 47*

46. WHY WAS . . . ABSENT FROM WORK DURING JUNE?

1 Own illness
2 Bad weather
3 Vacation
4 Labor dispute
5 Personal reasons
} *Skip to 51*
6 Seasonal factors
7 Has new job - to start in 30 days
8 Temporarily laid off
9 Other - *Specify* _____

47. WAS . . . LOOKING FOR WORK DURING JUNE?

1 Yes
2 No

48. WOULD . . . HAVE TAKEN A JOB IN JUNE?

1 Yes - *Skip to 50*
2 No

49. WHAT WAS . . .'S REASON FOR NOT (LOOKING FOR/TAKING) A JOB IN JUNE?

1 Found job
2 Illness
3 Family responsabilities
4 Attending school
5 On vacation
6 Waiting to go back to old job
7 Does not want job
8 Believes job not available
9 Unable to work

50. WHEN DID . . . LAST WORK--EVEN FOR A FEW DAYS?

1 Within the last 6 months
January 1980 through June 1980
2 6 months to 1 year ago
July 1979 through December 1979
3 1 year to 4 years ago
July 1976 through June 1979
4 5 years ago or more
June 1976 or earlier
} *Skip to 55*
5 Never worked - *Skip to 61*

51. WHAT WAS . . .'S PRINCIPAL JOB DURING JUNE?

Describe the specific kind of work that . . . was doing such as: making clay pots, teaching primary school, servant, driving a truck, midwife, farm helper, fruit picker, etc.

_____

52. IN WHAT KIND OF BUSINESS OR INDUSTRY WAS THIS?

Describe the specific type of business or industry such as: farm, automobile repair shop, primary school, bus company, coal mine, wholesale grain company, private home (servant), Armed Forces, etc.

_____

53. FOR THIS JOB WAS . . . --

1 Employee of private company or business for pay or wages
2 Government employee - National or local
3 Employer with paid employees
4 Self-employed or own-account worker
5 Unpaid worker in family farm or business
6 Armed Forces

54. IN ADDITION TO THE JOB DESCRIBED ABOVE DID . . . ALSO WORK AT OR HAVE ANY OTHER JOB IN JUNE?

1 Yes
2 No - *Skip to 58*

55. WHAT KIND OF WORK DID . . . DO IN THIS (SECOND/LAST) JOB?

Describe the specific kind of work that . . . was doing such as: making clay pots, teaching primary school, servant, driving a truck, midwife, farm helper, fruit picker, etc.

_____

56. IN WHAT KIND OF BUSINESS OR INDUSTRY WAS THIS?

Describe the specific type of business or industry such as: making clay pots, teaching primary school, servant, company, coal mine, wholesale grain company, private home (servant), Armed Forces, etc.

_____

57. FOR THIS (SECOND/LAST) JOB WAS . . . --

1 Employee of private company or business for pay or wages
2 Government employee - National or local
3 Employer with paid employees
4 Self-employed or own-account worker
5 Unpaid worker in family farm or business
6 Armed Forces

58. *If NO in 44, skip to 59.*
DURING JUNE HOW MANY HOURS DID . . . USUALLY WORK PER WEEK AT ALL JOBS NOT INCLUDING OWN HOUSEWORK?

_____ Hours

59. *If CODE 3 or 4 in 50, skip to 61*
DURING 1979 HOW MANY WEEKS DID . . . DO ANY WORK? *Include paid holidays, leave, and military service.*

_____ Weeks

60. DURING THESE _____ WEEKS WORKED IN 1979, HOW MANY HOURS DID . . . USUALLY WORK PER WEEK?

_____ Hours

61. FOR THE ENTIRE YEAR OF 1979 WHAT WAS . . .'S INCOME? *Round to the nearest goldar.*

00000 None

_____ Goldars

Exhibit F-3-4. BOOKLET LONG-FORM QUESTIONNAIRE FOR 1980 CENSUS OF POPSTAN--Continued

ALL HOUSING UNITS

H-1. TYPE OF LIVING QUARTERS
- 1 - Conventional HU
- 2 - Improvised HU (makeshift shelter of waste materials, barn, cave, warehouse, etc.)
- 3 - Mobile HU (tent, boat, wagon, etc.)
- 4 - Collective quarters } *Skip to H-19.*
- 5 - None (for homeless) }

H-2. NUMBER OF HOUSING UNITS IN THIS HOUSE (BUILDING)
- 1 - 1 unit, detached
- 2 - 1 unit, attached
- 3 - 2 to 4 units
- 4 - 5 to 9 units
- 5 - 10 units or more

H-3. WALLS - Main construction material of outer walls of this house (building)
- 1 - Stone, cement, stucco, brick
- 2 - Metal
- 3 - Wood
- 4 - Bamboo, leaves, reed, mud
- 5 - Other material

H-4. ROOF - Main construction material of roof on this house (building)
- 1 - Stone, cement, brick, tile, slate
- 2 - Metal
- 3 - Wood
- 4 - Bamboo, leaves, reed, mud
- 5 - Other material or none

H-5. FLOORS - Main construction material of floors in this housing unit
- 1 - Stone, cement, lime
- 2 - Tile, slate, marble
- 3 - Wood
- 4 - Leaves, reed, mud
- 5 - Other material

H-6. HOW MANY ROOMS ARE IN THIS UNIT?

_____ rooms

H-7. IS THIS UNIT OWNED BY SOMEONE LIVING IN IT OR IS IT RENTED?

OCCUPIED:
- 1 - Owned or being bought - *Skip to H-11.*
- 2 - Rented for cash
- 3 - Rent free or other arrangement - *Skip to H-11.*
- 4 - VACANT - *Skip to H-9.*

H-8. WHAT IS THE AMOUNT OF RENT PAID MONTHLY?

Round to the nearest goldar.

_____ goldars

Skip to H-11.

VACANT UNITS

H-9. VACANCY STATUS
- 1 - For rent or sale
- 2 - Not for rent or sale

H-10. CONDITION OF VACANT UNIT
- 1 - Habitable for year-round use
- 2 - Habitable for seasonal use
- 3 - Not habitable

OCCUPIED UNITS

H-11. WAS THIS UNIT CONSTRUCTED IN THE LAST 12 MONTHS?
- 1 - Yes
- 2 - No

H-12. IS THERE PIPED RUNNING WATER FOR THIS UNIT?
- 1 - Piped water inside building (structure)
- 2 - Piped water outside building (structure) within 100 meters
- 3 - Piped water outside building (structure) beyond 100 meters
- 4 - No piped water

H-13. WHAT IS THE SOURCE OF DRINKING WATER FOR THIS UNIT?
- 1 - Pipe or pump inside building
- 2 - Pipe or pump outside building
- 3 - Bottled or canned water
- 4 - Closed well or closed spring
- 5 - Open well or spring
- 6 - River, lake, other

H-14. WHAT TYPE OF TOILET FACILITIES ARE AVAILABLE FOR THIS UNIT?
- 1 - Flush toilet for exclusive use of the occupants of this HU
- 2 - Flush toilet shared with occupants of another HU
- 3 - Out-house or out-building
- 4 - Covered pit
- 5 - Open pit, ditch, pail
- 6 - Other

H-15. WHAT TYPE OF BATHING FACILITIES ARE AVAILABLE FOR THIS UNIT?
- 1 - Bathtub or shower with piped water for exclusive use of the occupants of this HU
- 2 - Bathtub or shower with piped water shared with occupants of another HU
- 3 - Hand basin with piped water
- 4 - Portable tub or basin
- 5 - Other

H-16. IS THERE A TELEVISION SET IN THIS UNIT?
- 1 - Yes
- 2 - No

H-17. WHAT KIND OF LIGHTING DOES THIS UNIT HAVE?
- 1 - Electric
- 2 - Gas
- 3 - Kerosene
- 4 - Other or none

H-18. WHAT KIND OF FUEL IS USED MOST IN THIS UNIT FOR COOKING?
- 1 - Electricity
- 2 - Gas
- 3 - Kerosene or oil
- 4 - Coal or charcoal
- 5 - Wood
- 6 - Other or none

H-19. ARE ANY ARTICLES PRODUCED IN THIS UNIT FOR SALE OR BARTER - NOT INCLUDING AGRICULTURAL PRODUCTS?
- 1 - Yes
- 2 - No

Exhibit F-5-1. OMR FORM

FORM **PH-2 (1980)**
JULY 1979

Phase 1 Questionnaire
POPSTAN 1980 CENSUSES OF POPULATION AND HOUSING

MINISTRY OF ECONOMY
NATIONAL STATISTICAL OFFICE

DISTRICT NUMBER	BLOCK NUMBER	SERIAL NUMBER	EA NUMBER

CONFIDENTIAL — This inquiry is required by law. The information is accorded confidential treatment and cannot be used for taxation, investigation, or regulation.

Inmates
① Yes
② No

Continuation questionnaire
① Yes
② No

PRINCIPAL RESPONDENT

Line number of principal respondent

Date of interview

REMARKS

ALL HOUSING UNITS

H-1. Type of living quarters
① Conventional HU
② Improvised HU (makeshift shelter of waste materials, barn, cave, warehouse, etc.)
③ Mobile HU (tent, boat, wagon, etc.)
④ Collective quarters } *Skip to H-19*
⑤ None (for homeless) }

H-2. Number of housing units in this house (building)
① 1 unit, detached
② 1 unit, attached
③ 2 to 4 units
④ 5 to 9 units
⑤ 10 units or more

H-3. WALLS - Main construction material of outer walls of this house (building)
① Stone, cement, stucco, brick
② Metal
③ Wood
④ Bamboo, leaves, reed, mud
⑤ Other material

H-4. ROOF - Main construction material of roof on this house (building)
① Stone, cement, brick, tile, slate
② Metal
③ Wood
④ Bamboo, leaves, reed, mud
⑤ Other material or none

H-5. FLOORS — Main construction material of floors in this housing unit
① Stone, cement, lime
② Tile, slate, marble
③ Wood
④ leaves, reed, mud
⑤ Other material

H-6. How many rooms are in this unit?
rooms

H-7. Is this unit owned by someone living in it or is it rented?
Occupied:
① Owned or being bought — *Skip to H-11*
② Rented for cash
③ Rent free or other arrangement — *Skip to H-11*
④ **Vacant** — *Skip to H-9*

H-8. What is the amount of rent paid monthly?
Round to the nearest goldar
Skip to H-11

VACANT UNITS

H-9. Vacancy status
① For rent or sale
② Not for rent or sale

H-10. Condition of vacant unit?
① Habitable for year-round use
② Habitable for seasonal use
③ Not habitable

OCCUPIED UNITS

H-11. Was this unit constructed in the last 12 months?
① Yes
② No

H-12. Is there piped running water for this unit?
① Piped water inside building (structure)
② Piped water outside building (structure) within 100 meters
③ Piped water outside building (structure) beyond 100 meters
④ No piped water

H-13. What is the source of the drinking water for this unit?
① Pipe or pump inside building
② Pipe or pump outside building
③ Bottled or canned water
④ Closed well or closed spring
⑤ Open well or spring
⑥ River, lake, other

H-14. What type of toilet facilities are available for this unit?
① Flush toilet for exclusive use of the occupants of this HU
② Flush toilet shared with occupants of another HU
③ Out-house or out-building
④ Covered pit
⑤ Open pit, ditch, pail
⑥ Other

H-15. What type of bathing facilities are available for this unit?
① Bathtub or shower with piped water for exclusive use of the occupants of this HU
② Bathtub or shower with piped water shared with occupants of another HU
③ Hand basin with piped water
④ Portable tub or basin
⑤ Other

H-16. Is there a television set in this unit?
① Yes
② No

H-17. What kind of lighting does this unit have?
① Electric
② Gas
③ Kerosene
④ Other or none

H-18. What kind of fuel is used most in this unit for cooking
① Electricity
② Gas
③ Kerosene or oil
④ Coal or charcoal
⑤ Wood
⑥ Other or none

H-19. Are any articles produced in this unit for sale or barter — not including agricultural products?
① Yes
② No

Exhibit F-5-1. OMR FORM--Continued

LINE NUMBER	ALL PERSONS										PERSONS 5 YEARS AND	
	What is the name of the head of this household? Names of all other persons who usually live here? Be sure to include babies and elderly persons. Members of the household away on vacation? On business? In the hospital? Any persons who have no usual residence elsewhere? List in this order Head Spouse of head Never married children of head or spouse (by age) Ever married children of head or spouse and their families (by age) Other relatives of head or spouse Nonrelatives of head or spouse	What is . . . 's relationship to the head of this household? 1—Head 2—Spouse of head 3—Son or daughter of head/spouse 4—Other relative 5—Nonrelative	Sex 1—Male 2—Female	Age as of last birthday *If under 1 year, fill 00.*	Marital status 1—Married 2—Divorced 3—Separated 4—Widowed 5—Never-married *If MALE or code 2, 3, 4, or 5, skip to P-7*	Married females *If husband lives in this HU, enter his line number* *If husband lives elsewhere, fill 00.*	Is . . . 's mother living? 1—Yes 2—No 3—Does not know *If No or DOES NOT KNOW, skip to P-9*	*If mother lives in this HU, enter her line number* *If mother lives elsewhere, fill 00*	In what country was . . . born? 1—Popstan 2—Endar 3—Victoria 4—Other	How many years has . . . lived continuously in this district? *If under 1 year, fill 00.* *If always, fill 98.*	Does . . . attend school? 1—Yes 2—No *Omit if 35 years or older*	What is the highest grade or year that . . . has completed? None (N) 00 Primary (P) 01 02 03 04 05 Middle (M) 06 07 08 High school (H) 09 10 11 12 University (U) 13 14 15 16 17+ *If 04 or more, skip to P-14*
	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
01		1 Head 2 Spouse 3 Son/Dau 4 Oth rel 5 Nonrel	1 M 2 F	00 01 02 03 04 05 06 07 08 09	1 Mar 2 Div 3 Sep 4 Wid 5 NM	00 01 02 03 04 05 06 07 08 09	1 Yes 2 No 3 Does not know	00 01 02 03 04 05 06 07 08 09	1 Pop 2 End 3 Vic 4 Other	00 01 02 03 04 05 06 07 08 09	1 Yes 2 No	00 N 01 02 03 04 05 P 06 07 08 M 09 10 11 12 H 13 14 15 16 17+ U
02		1 Head 2 Spouse 3 Son/Dau 4 Oth rel 5 Nonrel	1 M 2 F	00 01 02 03 04 05 06 07 08 09	1 Mar 2 Div 3 Sep 4 Wid 5 NM	00 01 02 03 04 05 06 07 08 09	1 Yes 2 No 3 Does not know	00 01 02 03 04 05 06 07 08 09	1 Pop 2 End 3 Vic 4 Other	00 01 02 03 04 05 06 07 08 09	1 Yes 2 No	00 N 01 02 03 04 05 P 06 07 08 M 09 10 11 12 H 13 14 15 16 17+ U
03		1 Head 2 Spouse 3 Son/Dau 4 Oth rel 5 Nonrel	1 M 2 F	00 01 02 03 04 05 06 07 08 09	1 Mar 2 Div 3 Sep 4 Wid 5 NM	00 01 02 03 04 05 06 07 08 09	1 Yes 2 No 3 Does not know	00 01 02 03 04 05 06 07 08 09	1 Pop 2 End 3 Vic 4 Other	00 01 02 03 04 05 06 07 08 09	1 Yes 2 No	00 N 01 02 03 04 05 P 06 07 09 M 09 10 11 12 H 13 14 15 16 17+ U
04		1 Head 2 Spouse 3 Son/Dau 4 Oth rel 5 Nonrel	1 M 2 F	00 01 02 03 04 05 06 07 08 09	1 Mar 2 Div 3 Sep 4 Wid 5 NM	00 01 02 03 04 05 06 07 08 09	1 Yes 2 No 3 Does not know	00 01 02 03 04 05 06 07 08 09	1 Pop 2 End 3 Vic 4 Other	00 01 02 03 04 05 06 07 08 09	1 Yes 2 No	00 N 01 02 03 04 05 P 06 07 08 M 09 10 11 12 H 13 14 15 16 17+ U

P-1a. If listing continued on page 4, fill the oval. ○

P-1b. TOTAL PERSONS — I have listed _____ persons. Have I missed anyone?
If YES and person is usual resident or has no usual residence elsewhere, add to P-1

PAGE 2

Exhibit F-5-1. OMR FORM--Continued

OLDER	FEMALES 12 YEARS AND OLDER					PERSONS 12 YEARS AND OLDER							
Can . . . read and write? 1—Yes 2—No	How many children have been born alive to . . . ? *If NONE, fill 00 and skip to P-19*	How many of these children are now — Living in this HU? *If NONE, fill 00.*	Living elsewhere? *If NONE, fill 00.*	Dead? *If NONE, fill 00.*	How many children have been born alive to . . . in the last 12 months? *If NONE, fill 00. Omit if 50 years or older*	During June what was . . . doing most of the time — 1—Working *Skip to P-23.* 2—Own housework 3—Going to school 4—Something else	During June — Did . . . do any work at all not including own housework? 1—Yes *Skip to P-23* 2—No	Was . . . looking for work? 1—Yes. *End interview for this person.* 2—No	Why was . . . not looking for work? 1—Had job *Skip to P-24* 2—Job not available 3—Did not want job 4—Unable to work *If 2, 3, or 4, end interview for this person*	How many hours per week did . . . usually work at all jobs, not including own housework?	In what business or industry was . . . 's principal job? 1—Farming, forestry, fishing, hunting 2—Mining, quarrying 3—Manufacturing 4—Electricity, gas, water 5—Construction 6—Wholesale or retail trade, restaurant, hotel 7—Transport, storage, communication 8—Financial, business services 9—Public administration 0—Social or personal services	Was . . . — 1—Employer (ER) 2—Employee for pay (EE) 3—Self-employed or own account worker 4—Unpaid family worker 5—Armed Forces (AF)	LINE NUMBER
P-13	P-14	P-15	P-16	P-17	P-18	P-19	P-20	P-21	P-22	P-23	P-24	P-25	
1 Yes 2 No	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	1 Work 2 Own hswk 3 School 4 Something else	1 Yes 2 No	1 Yes 2 No	1 Job 2 No job available 3 Not want 4 Unable	00 11 22 33 44 55 66 77 88 99	1 Farming 2 Mining 3 Manufg 4 Elec 5 Constr 6 Wholesale or retail 7 Transp 8 Finance 9 Public adm 0 Social ser	1 ER 2 EE 3 Self 4 Unpaid 5 AF	01
1 Yes 2 No	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	1 Work 2 Own hswk 3 School 4 Something else	1 Yes 2 No	1 Yes 2 No	1 Job 2 No job available 3 Not want 4 Unable	00 11 22 33 44 55 66 77 88 99	1 Farming 2 Mining 3 Manufg 4 Elec 5 Constr 6 Wholesale or retail 7 Transp 8 Finance 9 Public adm 0 Social ser	1 ER 2 EE 3 Self 4 Unpaid 5 AF	02
1 Yes 2 No	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	1 Work 2 Own hswk 3 School 4 Something else	1 Yes 2 No	1 Yes 2 No	1 Job 2 No job available 3 Not want 4 Unable	00 11 22 33 44 55 66 77 88 99	1 Farming 2 Mining 3 Manufg 4 Elec 5 Constr 6 Wholesale or retail 7 Transp 8 Finance 9 Public adm 0 Social ser	1 ER 2 EE 3 Self 4 Unpaid 5 AF	03
1 Yes 2 No	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	00 11 22 33 44 55 66 77 88 99	1 Work 2 Own hswk 3 School 4 Something else	1 Yes 2 No	1 Yes 2 No	1 Job 2 No job available 3 Not want 4 Unable	00 11 22 33 44 55 66 77 88 99	1 Farming 2 Mining 3 Manufg 4 Elec 5 Constr 6 Wholesale or retail 7 Transp 8 Finance 9 Public adm 0 Social ser	1 ER 2 EE 3 Self 4 Unpaid 5 AF	04

P-1c. If only nonresidents are living here fill in the oval. *Complete items H-1 to H-10 on front.* ○

NSO USE ONLY

P-26. Number of lines canceled.

P-27. Population count after edit

PAGE 3

Exhibit F-5-1. OMR FORM--Continued

	ALL PERSONS										PERSONS 5 YEARS AND	
LINE NUMBER	What is the name of the head of this household? Names of all other persons who usually live here? Be sure to include babies and elderly persons. Members of the household away on vacation? On business? In the hospital? Any persons who have no usual residence elsewhere? List in this order Head Spouse of head Never married children of head or spouse (by age) Ever married children of head or spouse and their families (by age) Other relatives of head or spouse Nonrelatives of head or spouse	What is . . . 's relationship to the head of this household? 1—Head 2—Spouse of head 3—Son or daughter of head/spouse 4—Other relative 5—Nonrelative	Sex 1—Male 2—Female	Age as of last birthday *If under 1 year, fill 00.*	Marital status 1—Married 2—Divorced 3—Separated 4—Widowed 5—Never-married *If MALE or code 2, 3, 4, or 5, skip to P-7*	Married females *If husband lives in this HU, enter his line number* *If husband lives elsewhere, fill 00.*	Is . . . 's mother living? 1—Yes 2—No 3—Does not know *If No or DOES NOT KNOW, skip to P-9*	*If mother lives in this HU, enter her line number.* *If mother lives elsewhere, fill 00.*	In what country was . . . born? 1—Popstan 2—Endar 3—Victoria 4—Other	How many years has . . . lived continuously in this district? *If under 1 year, fill 00.* *If always, fill 98.*	Does . . . attend school? 1—Yes 2—No *Omit if 35 years or older.*	What is the highest grade or year that . . . has completed? None (N) 00 Primary (P) 01 02 03 04 05 Middle (M) 06 07 08 High school (H) 09 10 11 12 University (U) 13 14 15 16 17+ *If 04 or more, skip to P-14*
	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
05		① Head ② Spouse ③ Son/Dau ④ Oth rel ⑤ Nonrel	① M ② F	⓪⓪⓪ ①①① ②②② ③③③ ④④④ ⑤⑤⑤ ⑥⑥⑥ ⑦⑦⑦ ⑧⑧⑧ ⑨⑨⑨	① Mar ② Div ③ Sep ④ Wid ⑤ NM	⓪⓪ ①① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Yes ② No ③ Does not know	⓪⓪ ①① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Pop ② End ③ Vic ④ Other	⓪⓪ ①① ②② ③③ ④④ ⑤⑤ ⑥⑥ ⑦⑦ ⑧⑧ ⑨⑨	① Yes ② No	00 N ○ 01 02 03 04 05 P ○○○○○ 06 07 08 M ○○○ 09 10 11 12 H ○○○○ 13 14 15 16 17+ U ○○○○○○
06		① Head ② Spouse ③ Son/Dau ④ Oth rel ⑤ Nonrel	① M ② F	⓪⓪⓪ ①①① ②②② ③③③ ④④④ ⑤⑤⑤ ⑥⑥⑥ ⑦⑦⑦ ⑧⑧⑧ ⑨⑨⑨	① Mar ② Div ③ Sep ④ Wid ⑤ NM	⓪⓪ ①① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Yes ② No ③ Does not know	⓪⓪ ①① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Pop ② End ③ Vic ④ Other	⓪⓪ ①① ②② ③③ ④④ ⑤⑤ ⑥⑥ ⑦⑦ ⑧⑧ ⑨⑨	① Yes ② No	00 N ○ 01 02 03 04 05 P ○○○○○ 06 07 08 M ○○○ 09 10 11 12 H ○○○○ 13 14 15 16 17+ U ○○○○○○
07		① Head ② Spouse ③ Son/Dau ④ Oth rel ⑤ Nonrel	① M ② F	⓪⓪⓪ ①①① ②②② ③③③ ④④④ ⑤⑤⑤ ⑥⑥⑥ ⑦⑦⑦ ⑧⑧⑧ ⑨⑨⑨	① Mar ② Div ③ Sep ④ Wid ⑤ NM	⓪⓪ ①① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Yes ② No ③ Does not know	⓪⓪ ①① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Pop ② End ③ Vic ④ Other	⓪⓪ ①① ②② ③③ ④④ ⑤⑤ ⑥⑥ ⑦⑦ ⑧⑧ ⑨⑨	① Yes ② No	00 N ○ 01 02 03 04 05 P ○○○○○ 06 07 08 M ○○○ 09 10 11 12 H ○○○○ 13 14 15 16 17+ U ○○○○○○
08		① Head ② Spouse ③ Son/Dau ④ Oth rel ⑤ Nonrel	① M ② F	⓪⓪⓪ ①①① ②②② ③③③ ④④④ ⑤⑤⑤ ⑥⑥⑥ ⑦⑦⑦ ⑧⑧⑧ ⑨⑨⑨	① Mar ② Div ③ Sep ④ Wid ⑤ NM	⓪⓪ ①① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Yes ② No ③ Does not know	⓪⓪ ①① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Pop ② End ③ Vic ④ Other	⓪⓪ ①① ②② ③③ ④④ ⑤⑤ ⑥⑥ ⑦⑦ ⑧⑧ ⑨⑨	① Yes ② No	00 N ○ 01 02 03 04 05 P ○○○○○ 06 07 08 M ○○○ 09 10 11 12 H ○○○○ 13 14 15 16 17+ U ○○○○○○

P-1a. If listing continued on page 8, fill the oval. ○

P-1b. TOTAL PERSONS — I have listed _____ persons. Have I missed anyone?
If YES and person is usual resident or has no usual residence elsewhere, add to P-1

PAGE 4

Exhibit F-5-1. OMR FORM--Continued

OLDER	FEMALES 12 YEARS AND OLDER					PERSONS 12 YEARS AND OLDER							
		How many of these children are now —					During June —						
Can . . . read and write? 1—Yes 2—No	How many children have been born alive to . . . ? *If NONE, fill 00 and skip to P-19*	Living in this HU? *If NONE, fill 00.*	Living elsewhere? *If NONE, fill 00.*	Dead? *If NONE, fill 00.*	How many children have been born alive to . . . in the last 12 months? *If NONE, fill 00. Omit if 50 years or older.*	During June what was . . . doing most of the time — 1—Working *Skip to P-23.* 2—Own housework 3—Going to school 4—Something else	Did . . . do any work at all not including own housework? 1—Yes *Skip to P-23* 2—No	Was . . . looking for work? 1—Yes, *End interview for this person.* 2—No	Why was . . . not looking for work? 1—Had job *Skip to P-24* 2—Job not available 3—Did not want job 4—Unable to work *If 2, 3, or 4, end interview for this person*	How many hours per week did . . . usually work at all jobs, not including own housework?	In what business or industry was . . . 's principal job? 1—Farming, forestry, fishing, hunting 2—Mining, quarrying 3—Manufacturing 4—Electricity, gas, water 5—Construction 6—Wholesale or retail trade, restaurant, hotel 7—Transport, storage, communication 8—Financial, business services 9—Public administration 0—Social or personal services	Was . . . — 1—Employer (ER) 2—Employee for pay (EE) 3—Self-employed or own account worker 4—Unpaid family worker 5—Armed Forces (AF)	LINE NUMBER
P-13	P-14	P-15	P-16	P-17	P-18	P-19	P-20	P-21	P-22	P-23	P-24	P-25	
① Yes ② No	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Work ② Own hswk ③ School ④ Something else	① Yes ② No	① Yes ② No	① Job ② No job available ③ Not want ④ Unable	⓪⓪ ①① ②② ③③ ④④ ⑤⑤ ⑥⑥ ⑦⑦ ⑧⑧ ⑨⑨	① Farming ② Mining ③ Manufg ④ Elec ⑤ Constr ⑥ Wholesale or retail ⑦ Transp ⑧ Finance ⑨ Public adm ⓪ Social ser	① ER ② EE ③ Self ④ Unpaid ⑤ AF	05
① Yes ② No	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Work ② Own hswk ③ School ④ Something else	① Yes ② No	① Yes ② No	① Job ② No job available ③ Not want ④ Unable	⓪⓪ ①① ②② ③③ ④④ ⑤⑤ ⑥⑥ ⑦⑦ ⑧⑧ ⑨⑨	① Farming ② Mining ③ Manufg ④ Elec ⑤ Constr ⑥ Wholesale or retail ⑦ Transp ⑧ Finance ⑨ Public adm ⓪ Social ser	① ER ② EE ③ Self ④ Unpaid ⑤ AF	06
① Yes ② No	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Work ② Own hswk ③ School ④ Something else	① Yes ② No	① Yes ② No	① Job ② No job available ③ Not want ④ Unable	⓪⓪ ①① ②② ③③ ④④ ⑤⑤ ⑥⑥ ⑦⑦ ⑧⑧ ⑨⑨	① Farming ② Mining ③ Manufg ④ Elec ⑤ Constr ⑥ Wholesale or retail ⑦ Transp ⑧ Finance ⑨ Public adm ⓪ Social ser	① ER ② EE ③ Self ④ Unpaid ⑤ AF	07
① Yes ② No	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	⓪⓪ ①① ②② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Work ② Own hswk ③ School ④ Something else	① Yes ② No	① Yes ② No	① Job ② No job available ③ Not want ④ Unable	⓪⓪ ①① ②② ③③ ④④ ⑤⑤ ⑥⑥ ⑦⑦ ⑧⑧ ⑨⑨	① Farming ② Mining ③ Manufg ④ Elec ⑤ Constr ⑥ Wholesale or retail ⑦ Transp ⑧ Finance ⑨ Public adm ⓪ Social ser	① ER ② EE ③ Self ④ Unpaid ⑤ AF	08

P-1c. If only nonresidents are living here fill in the oval. *Complete items H-1 to H-10 on front.* ○

NSO USE ONLY

P-26. Number of lines canceled

P-27. Population count after edit

PAGE 5

Exhibit F-5-1. OMR FORM--Continued

	ALL PERSONS										PERSONS 5 YEARS AND	
LINE NUMBER	What is the name of the head of this household? Names of all other persons who usually live here? Be sure to include babies and elderly persons. Members of the household away on vacation? On business? In the hospital? Any persons who have no usual residence elsewhere? List in this order Head Spouse of head Never married children of head or spouse (by age) Ever married children of head or spouse and their families (by age) Other relatives of head or spouse Nonrelatives of head or spouse	What is . . . 's relationship to the head of this household? 1—Head 2—Spouse of head 3—Son or daughter of head/spouse 4—Other relative 5—Nonrelative	Sex 1—Male 2—Female	Age as of last birthday *If under 1 year, fill 00.*	Marital status 1—Married 2—Divorced 3—Separated 4—Widowed 5—Never-married *If MALE or code 2, 3, 4, or 5, skip to P-7*	Married females *If husband lives in this HU, enter his line number* *If husband lives elsewhere, fill 00.*	Is . . . 's mother living? 1—Yes 2—No 3—Does not know *If No or DOES NOT KNOW, skip to P-9*	*If mother lives in this HU, enter her line number* *If mother lives elsewhere, fill 00*	In what country was . . . born? 1—Popstan 2—Ender 3—Victoria 4—Other	How many years has . . . lived continuously in this district? *If under 1 year, fill 00.* *If always, fill 98.*	Does . . . attend school? 1—Yes 2—No *Omit if 35 years or older*	What is the highest grade or year that . . . has completed? None (N) 00 Primary (P) 01 02 03 04 05 Middle (M) 06 07 08 High school (H) 09 10 11 12 University (U) 13 14 15 16 17+ *If 04 or more, skip to P-14*
	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
09		① Head ② Spouse ③ Son/Dau ④ Oth rel ⑤ Nonrel	① M ② F	⓪ ⓪ ⓪ ① ① ① ② ② ② ③ ③ ③ ④ ④ ④ ⑤ ⑤ ⑤ ⑥ ⑥ ⑥ ⑦ ⑦ ⑦ ⑧ ⑧ ⑧ ⑨ ⑨ ⑨	① Mar ② Div ③ Sep ④ Wid ⑤ NM	⓪ ⓪ ① ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Yes ② No ③ Does not know	⓪ ⓪ ① ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Pop ② End ③ Vic ④ Other	⓪ ⓪ ① ① ② ② ③ ③ ④ ④ ⑤ ⑤ ⑥ ⑥ ⑦ ⑦ ⑧ ⑧ ⑨ ⑨	① Yes ② No	00 N ○ 01 02 03 04 05 P ○○○○○ 06 07 08 M ○○○ 09 10 11 12 H ○○○○ 13 14 15 16 17+ U ○○○○○○
10		① Head ② Spouse ③ Son/Dau ④ Oth rel ⑤ Nonrel	① M ② F	⓪ ⓪ ⓪ ① ① ① ② ② ② ③ ③ ③ ④ ④ ④ ⑤ ⑤ ⑤ ⑥ ⑥ ⑥ ⑦ ⑦ ⑦ ⑧ ⑧ ⑧ ⑨ ⑨ ⑨	① Mar ② Div ③ Sep ④ Wid ⑤ NM	⓪ ⓪ ① ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Yes ② No ③ Does not know	⓪ ⓪ ① ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Pop ② End ③ Vic ④ Other	⓪ ⓪ ① ① ② ② ③ ③ ④ ④ ⑤ ⑤ ⑥ ⑥ ⑦ ⑦ ⑧ ⑧ ⑨ ⑨	① Yes ② No	00 N ○ 01 02 03 04 05 P ○○○○○ 06 07 08 M ○○○ 09 10 11 12 H ○○○○ 13 14 15 16 17+ U ○○○○○○
11		① Head ② Spouse ③ Son/Dau ④ Oth rel ⑤ Nonrel	① M ② F	⓪ ⓪ ⓪ ① ① ① ② ② ② ③ ③ ③ ④ ④ ④ ⑤ ⑤ ⑤ ⑥ ⑥ ⑥ ⑦ ⑦ ⑦ ⑧ ⑧ ⑧ ⑨ ⑨ ⑨	① Mar ② Div ③ Sep ④ Wid ⑤ NM	⓪ ⓪ ① ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Yes ② No ③ Does not know	⓪ ⓪ ① ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Pop ② End ③ Vic ④ Other	⓪ ⓪ ① ① ② ② ③ ③ ④ ④ ⑤ ⑤ ⑥ ⑥ ⑦ ⑦ ⑧ ⑧ ⑨ ⑨	① Yes ② No	00 N ○ 01 02 03 04 05 P ○○○○○ 06 07 08 M ○○○ 09 10 11 12 H ○○○○ 13 14 15 16 17+ U ○○○○○○
12		① Head ② Spouse ③ Son/Dau ④ Oth rel ⑤ Nonrel	① M ② F	⓪ ⓪ ⓪ ① ① ① ② ② ② ③ ③ ③ ④ ④ ④ ⑤ ⑤ ⑤ ⑥ ⑥ ⑥ ⑦ ⑦ ⑦ ⑧ ⑧ ⑧ ⑨ ⑨ ⑨	① Mar ② Div ③ Sep ④ Wid ⑤ NM	⓪ ⓪ ① ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Yes ② No ③ Does not know	⓪ ⓪ ① ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨	① Pop ② End ③ Vic ④ Other	⓪ ⓪ ① ① ② ② ③ ③ ④ ④ ⑤ ⑤ ⑥ ⑥ ⑦ ⑦ ⑧ ⑧ ⑨ ⑨	① Yes ② No	00 N ○ 01 02 03 04 05 P ○○○○○ 06 07 08 M ○○○ 09 10 11 12 H ○○○○ 13 14 15 16 17+ U ○○○○○○

P-1a. If listing continued on the next form, fill the oval. ○

P-1b. TOTAL PERSONS — I have listed _____ persons. Have I missed anyone?
If YES and person is usual resident or has no usual residence elsewhere, add to P-1

PAGE 6

Exhibit F-5-1. OMR FORM--Continued

OLDER	FEMALES 12 YEARS AND OLDER						PERSONS 12 YEARS AND OLDER							
Can . . . read and write? 1—Yes 2—No	How many children have been born alive to . . . ? *If NONE, fill 00 and skip to P-19*	How many of these children are now — Living in this HU? *If NONE, fill 00.*	Living elsewhere? *If NONE, fill 00.*	Dead? *If NONE, fill 00.*	How many children have been born alive to . . . in the last 12 months? *If NONE, fill 00. Omit if 50 years or older*	During June what was . . . doing most of the time — 1—Working *Skip to P-23* 2—Own housework 3—Going to school 4—Something else	During June — Did . . . do any work at all not including own housework? 1—Yes *Skip to P-23* 2—No	Was . . . looking for work? 1—Yes. *End interview for this person* 2—No	Why was . . . not looking for work? 1—Had job *Skip to P-24* 2—Job not available 3—Did not want job 4—Unable to work *If 2, 3, or 4, end interview for this person*	How many hours per week did . . . usually work at all jobs, not including own housework?	In what business or industry was . . . 's principal job? 1—Farming, forestry, fishing, hunting 2—Mining, quarrying 3—Manufacturing 4—Electricity, gas, water 5—Construction 6—Wholesale or retail trade, restaurant, hotel 7—Transport, storage, communication 8—Financial, business services 9—Public administration 0—Social or personal services	Was . . . — 1—Employer (ER) 2—Employee for pay (EE) 3—Self-employed or own account worker 4—Unpaid family worker 5—Armed Forces (AF)	LINE NUMBER
P-13	P-14	P-15	P-16	P-17	P-18	P-19	P-20	P-21	P-22	P-23	P-24	P-25	
	00	00	00	00	00					00	1 Farming	1 ER	
	11	11	11	11	11	1 Work			1 Job	11	2 Mining		
1 Yes	22	22	22	22	22		1 Yes	1 Yes		22	3 Manufg	2 EE	
	3	3	3	3	3	2 Own hswk			2 No job available	33	4 Elec		
	4	4	4	4	4					44	5 Constr	3 Self	
	5	5	5	5	5	3 School			3 Not want	55	6 Wholesale or retail		09
2 No	6	6	6	6	6		2 No	2 No		66	7 Transp	4 Unpaid	
	7	7	7	7	7	4 Something else			4 Unable	77	8 Finance		
	8	8	8	8	8					88	9 Public adm	5 AF	
	9	9	9	9	9					99	0 Social ser		
	00	00	00	00	00					00	1 Farming	1 ER	
	11	11	11	11	11	1 Work			1 Job	11	2 Mining		
1 Yes	22	22	22	22	22		1 Yes	1 Yes		22	3 Manufg	2 EE	
	3	3	3	3	3	2 Own hswk			2 No job available	33	4 Elec		
	4	4	4	4	4					44	5 Constr	3 Self	
	5	5	5	5	5	3 School			3 Not want	55	6 Wholesale or retail		10
2 No	6	6	6	6	6		2 No	2 No		66	7 Transp	4 Unpaid	
	7	7	7	7	7	4 Something else			4 Unable	77	8 Finance		
	8	8	8	8	8					88	9 Public adm	5 AF	
	9	9	9	9	9					99	0 Social ser		
	00	00	00	00	00					00	1 Farming	1 ER	
	11	11	11	11	11	1 Work			1 Job	11	2 Mining		
1 Yes	22	22	22	22	22		1 Yes	1 Yes		22	3 Manufg	2 EE	
	3	3	3	3	3	2 Own hswk			2 No job available	33	4 Elec		
	4	4	4	4	4					44	5 Constr	3 Self	
	5	5	5	5	5	3 School			3 Not want	55	6 Wholesale or retail		11
2 No	6	6	6	6	6		2 No	2 No		66	7 Transp	4 Unpaid	
	7	7	7	7	7	4 Something else			4 Unable	77	8 Finance		
	8	8	8	8	8					88	9 Public adm	5 AF	
	9	9	9	9	9					99	0 Social ser		
	00	00	00	00	00					00	1 Farming	1 ER	
	11	11	11	11	11	1 Work			1 Job	11	2 Mining		
1 Yes	22	22	22	22	22		1 Yes	1 Yes		22	3 Manufg	2 EE	
	3	3	3	3	3	2 Own hswk			2 No job available	33	4 Elec		
	4	4	4	4	4					44	5 Constr	3 Self	
	5	5	5	5	5	3 School			3 Not want	55	6 Wholesale or retail		12
2 No	6	6	6	6	6		2 No	2 No		66	7 Transp	4 Unpaid	
	7	7	7	7	7	4 Something else			4 Unable	77	8 Finance		
	8	8	8	8	8					88	9 Public adm	5 AF	
	9	9	9	9	9					99	0 Social ser		

P-1c. If only nonresidents are living here fill in the oval. *Complete items H-1 to H-10 on front* ○

NSO USE ONLY

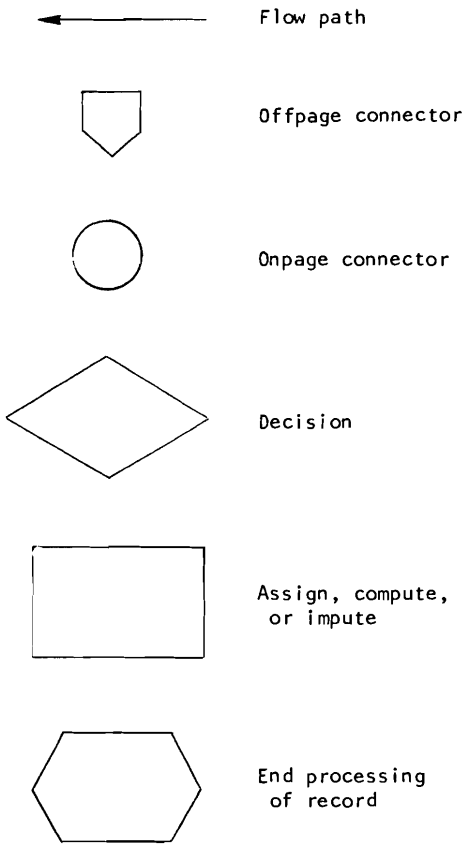
P-26. Number of lines canceled

P-27. Population count after edit

PAGE 7

Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS

LEGEND



GUIDE TO TERMS

Assign: Insert into the record being edited an imputed or computed value

First person in housing unit: First population record for a serial number

HG: Higest grade completed

Line: Line number of the record being edited

Line of head: Line number of head of household

Line of husband: Line number of husband as reported in item P-6.

Line of mother: Line number of mother as reported in item P-8

Line of spouse: Line number of any previous person in the housing unit whose relationship is reported as spouse

Married female: Any previous female in the housing unit who reported married for marital status

Married male: Any previous male in the housing unit who reported married for marital status

Married male's line: Line number of any previous male in the housing unit who reported married for marital status

MS: Marital status

Previous mother: Any previous female in the housing unit who reported children ever born

Previous mother's line: Line number of any previous mother

Previous person: Any previous person in the housing unit (with same serial number)--not simply the preceding person

Rel: Relationship

Valid entry: A value within the range of codes for an item as defined on the questionnaire. For example, the valid codes for marital status are 1 to 5:

1 = Married
2 = Divorced
3 = Separated
4 = Widowed
5 = Never married

Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

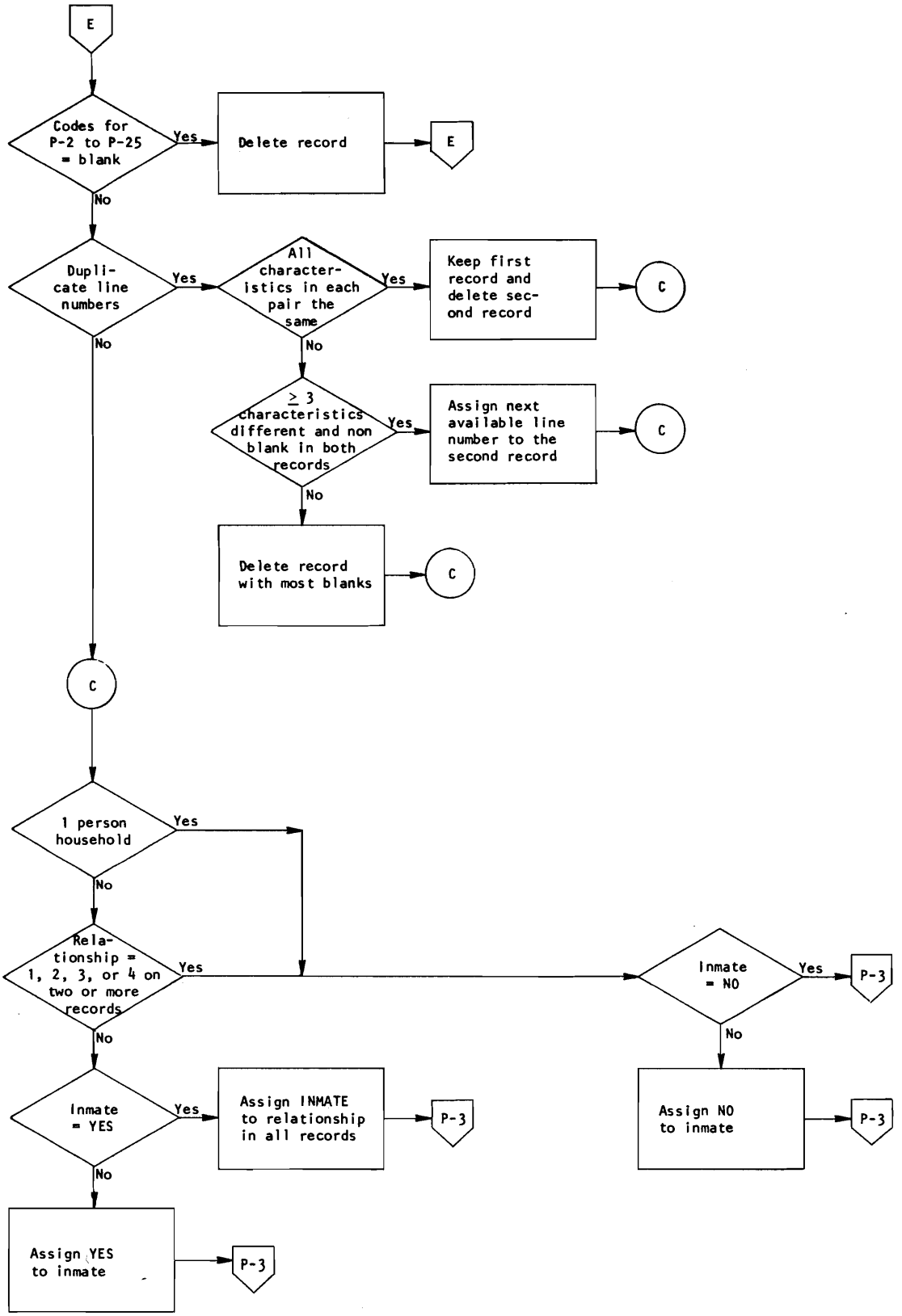


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

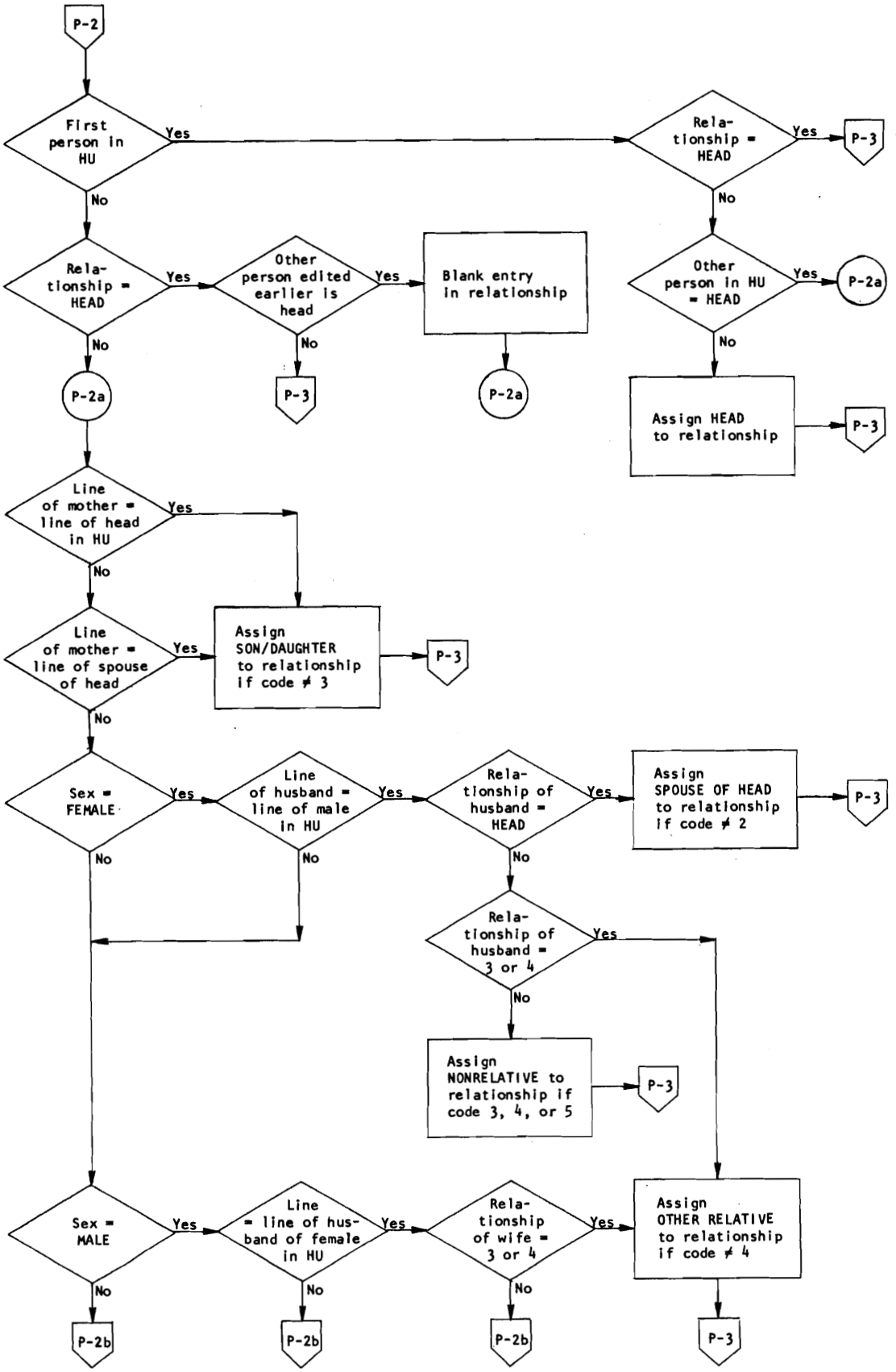


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

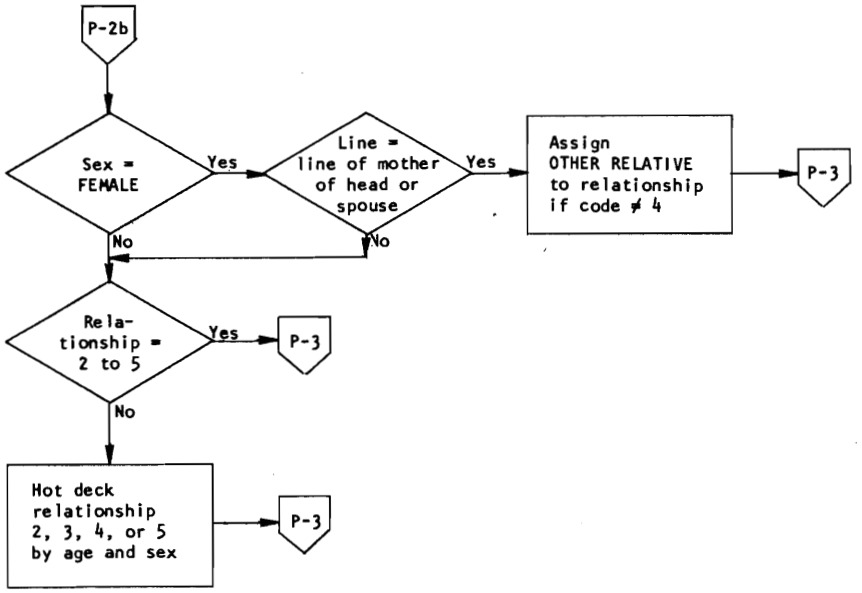


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

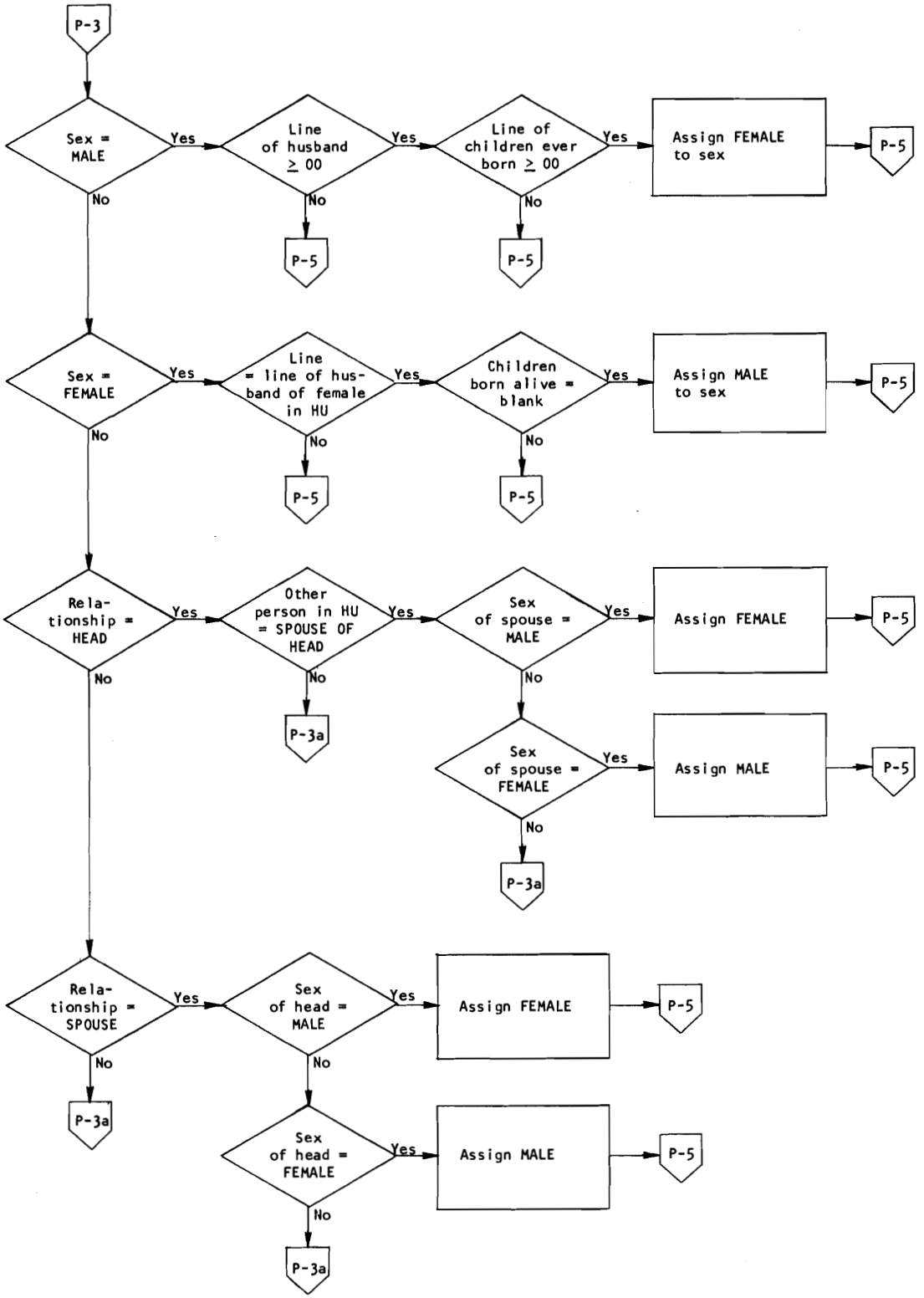


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

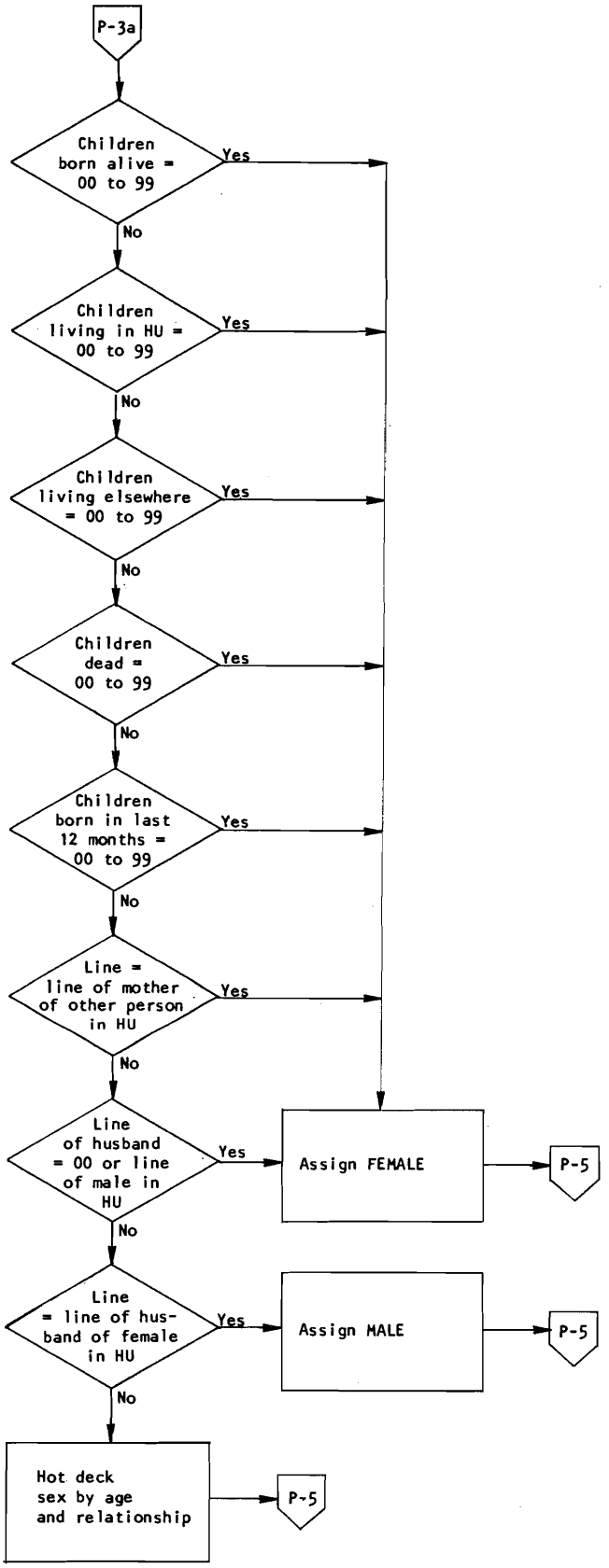


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

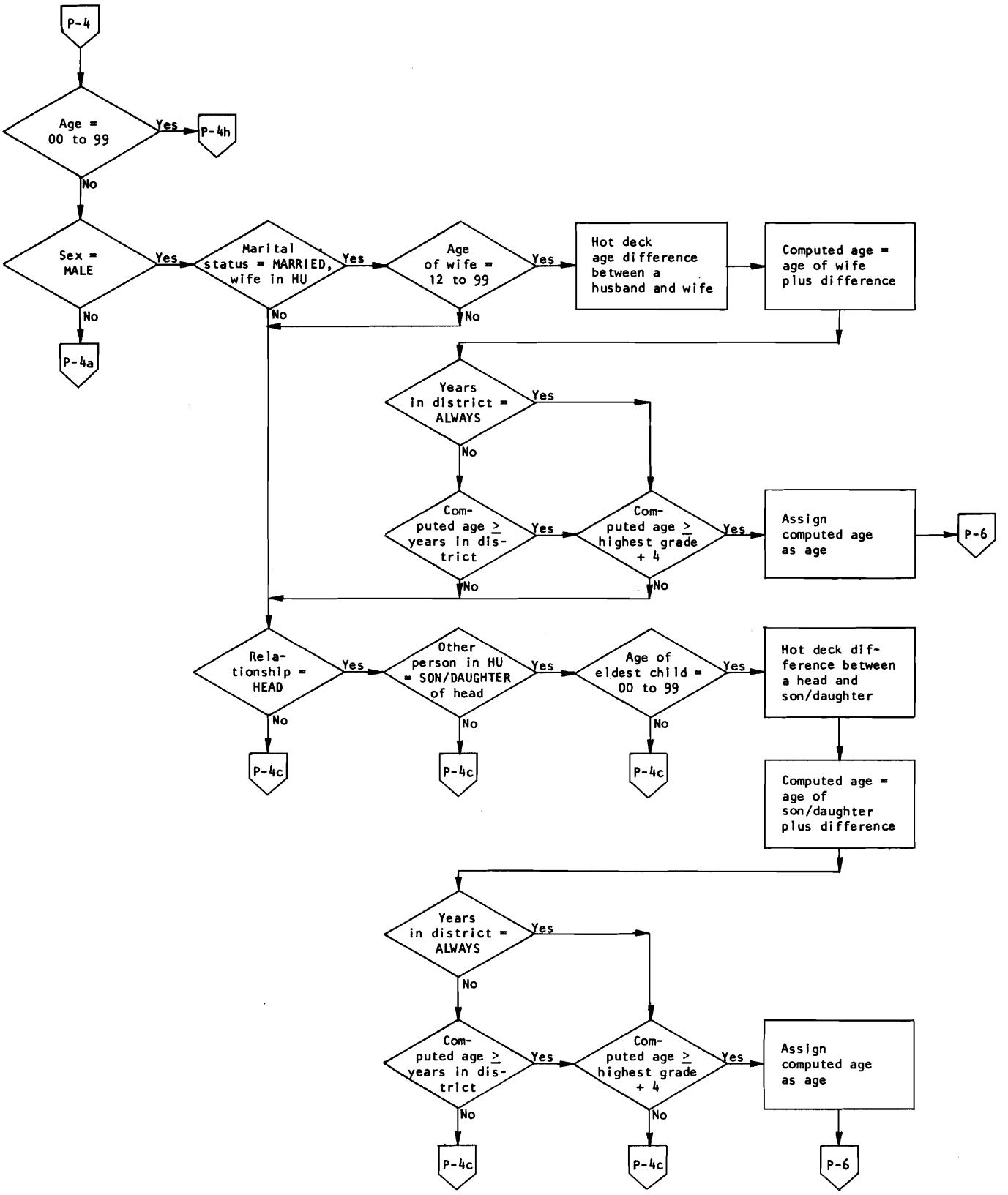


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

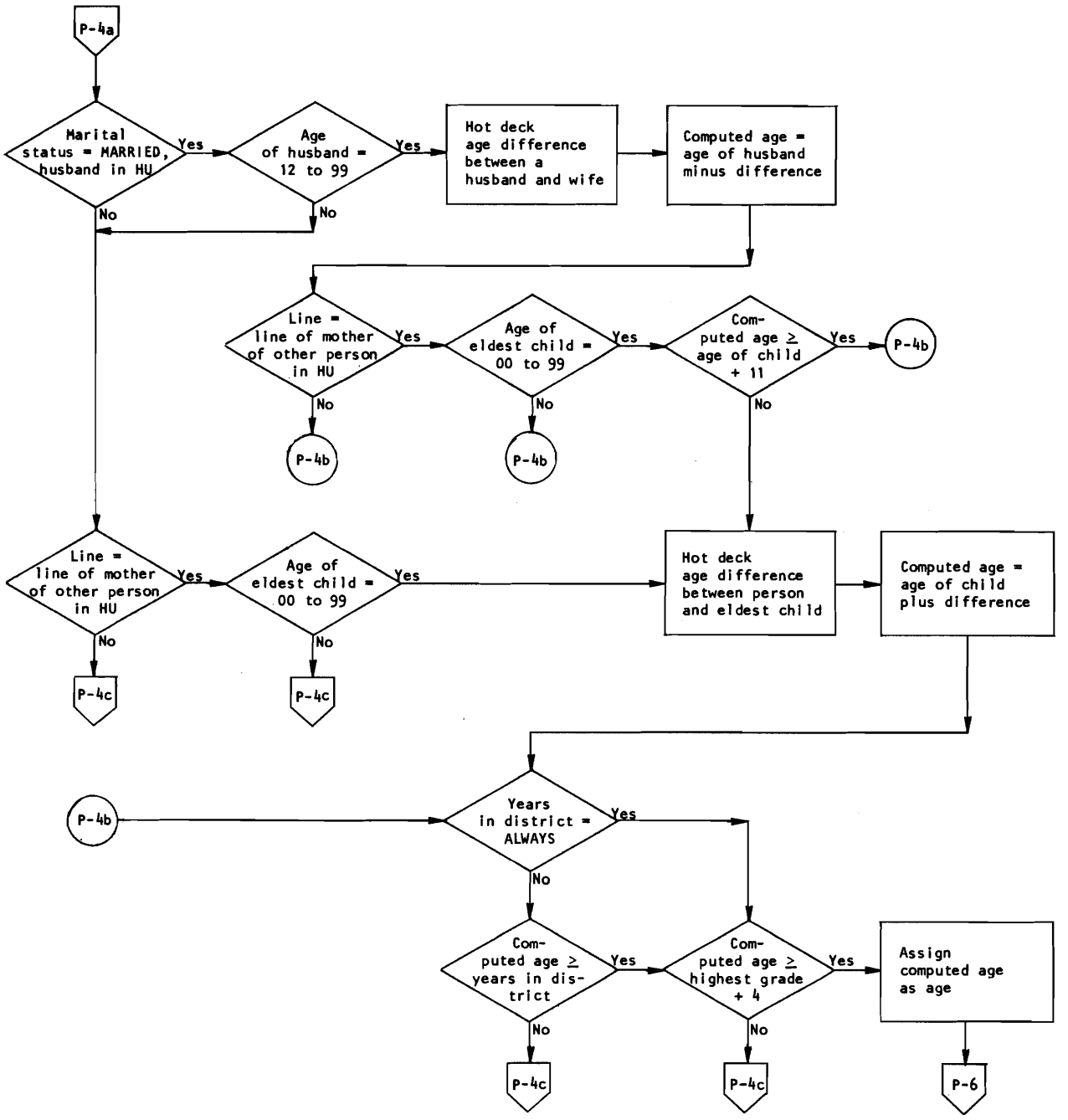


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

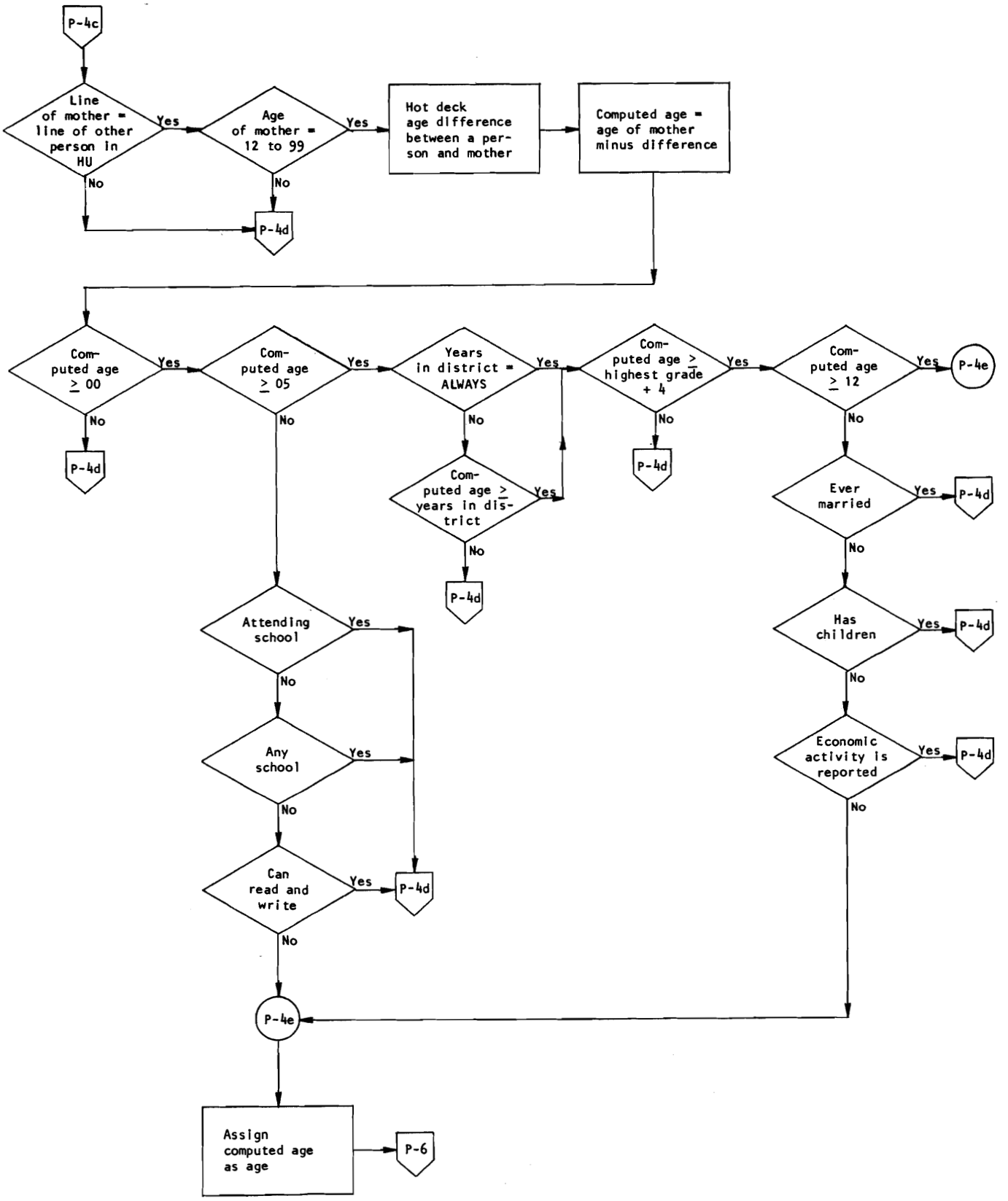


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

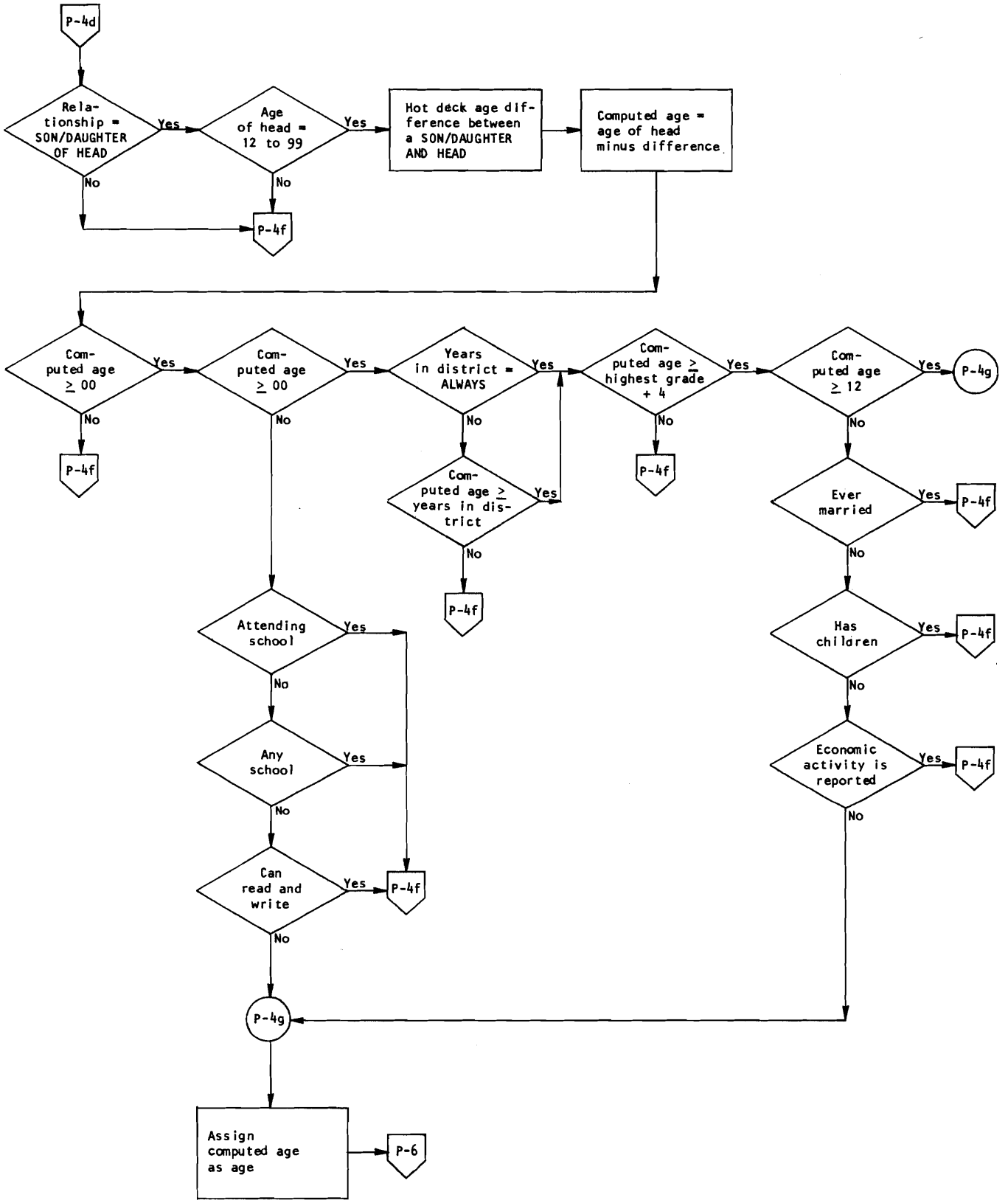


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

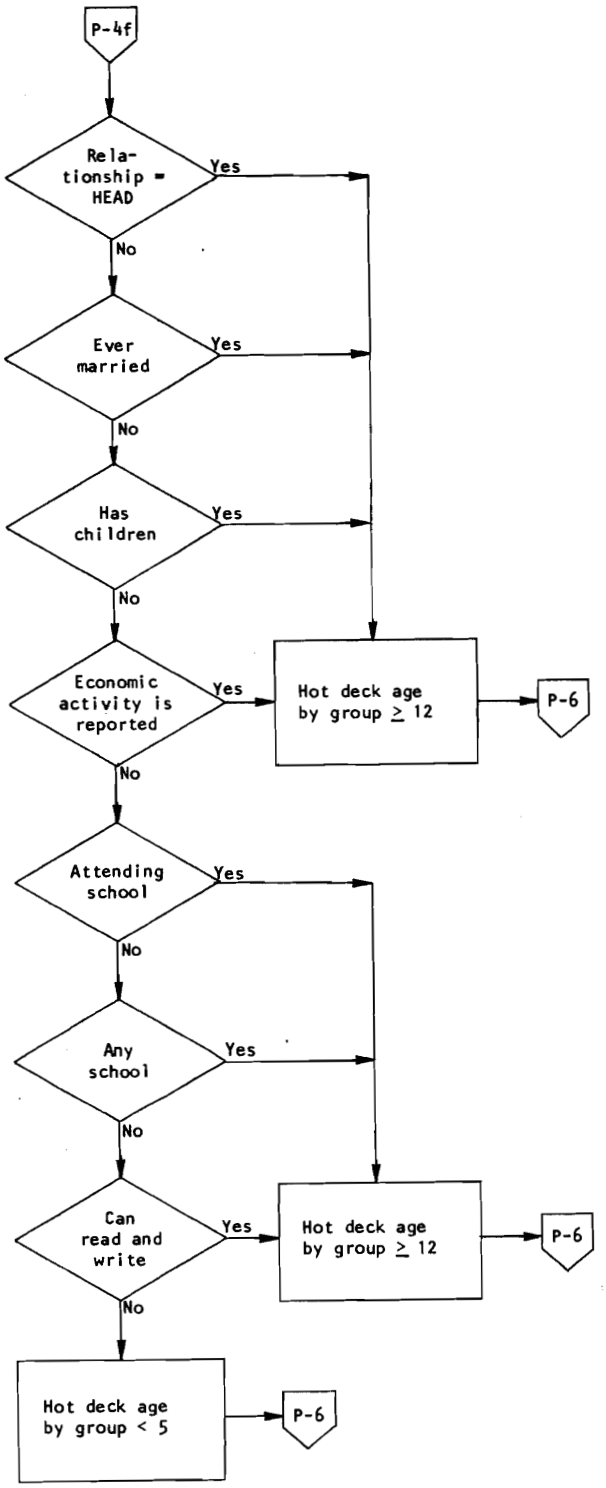


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

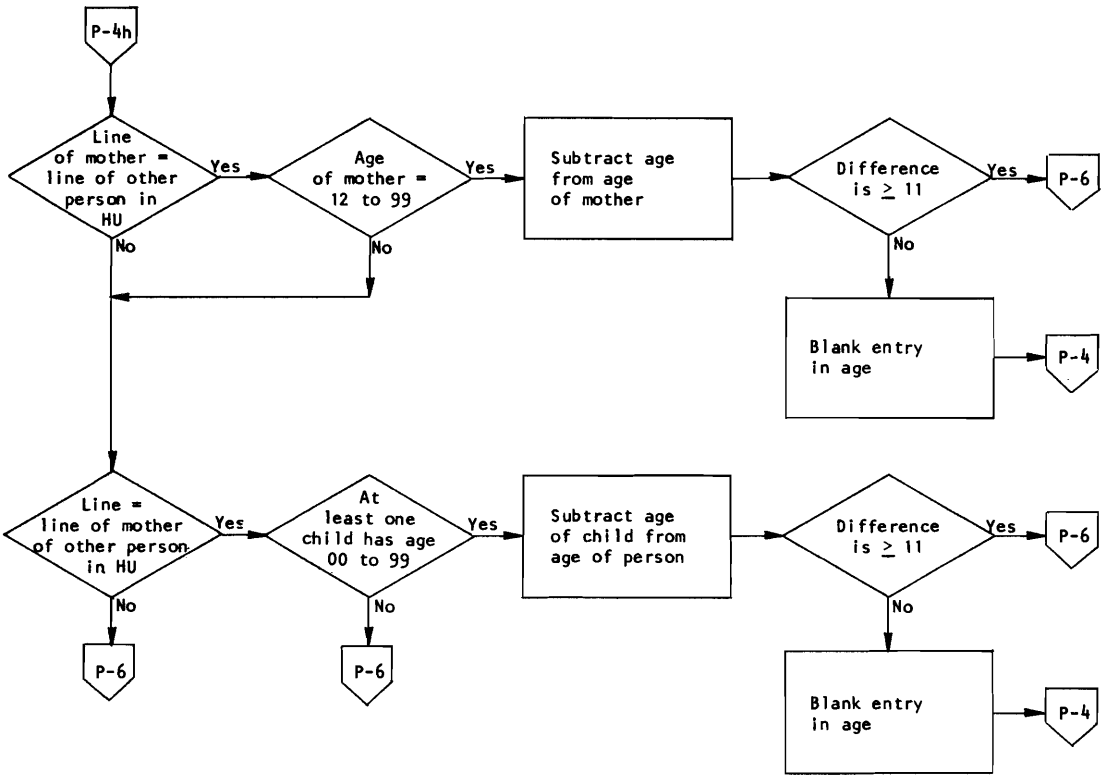


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

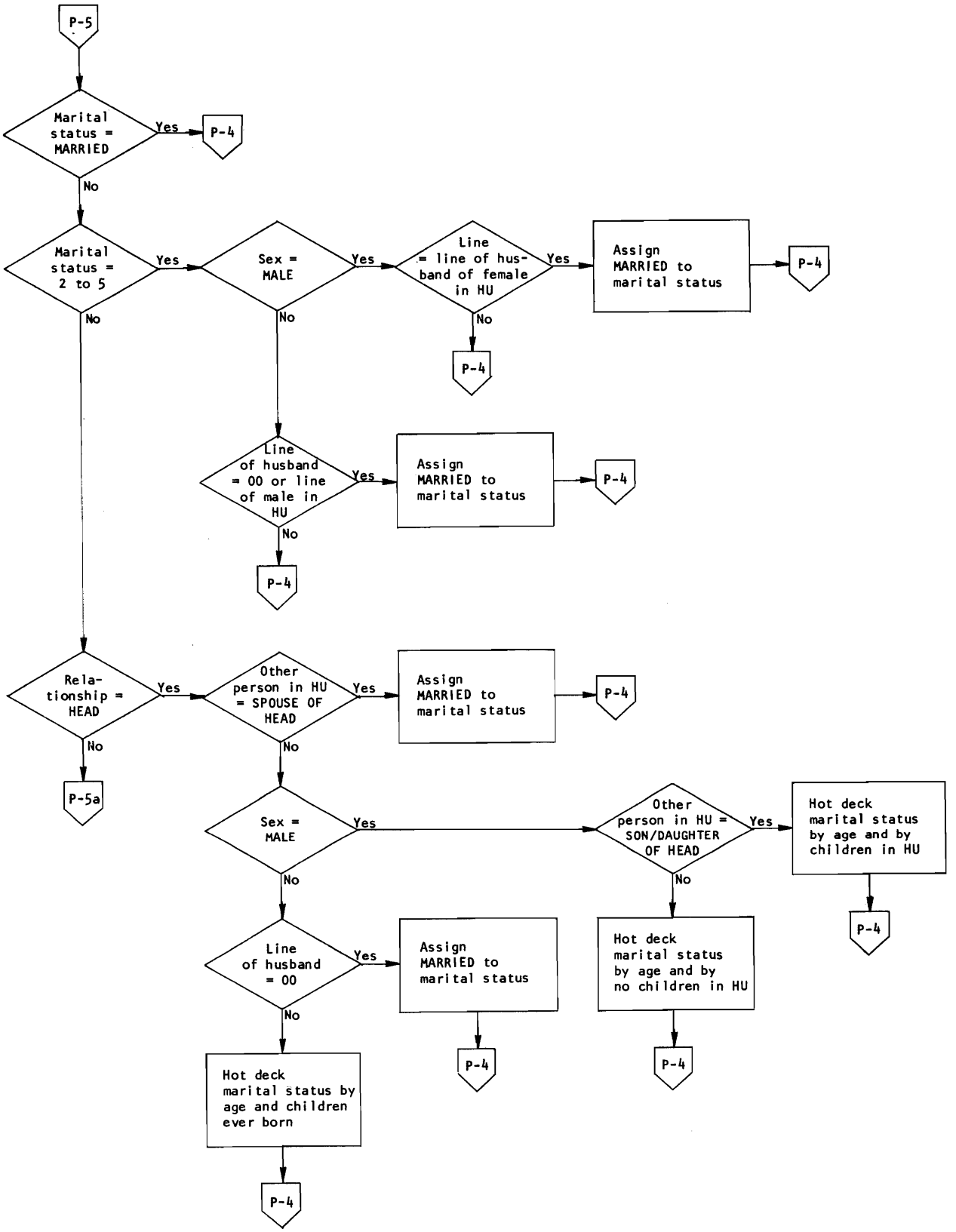


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

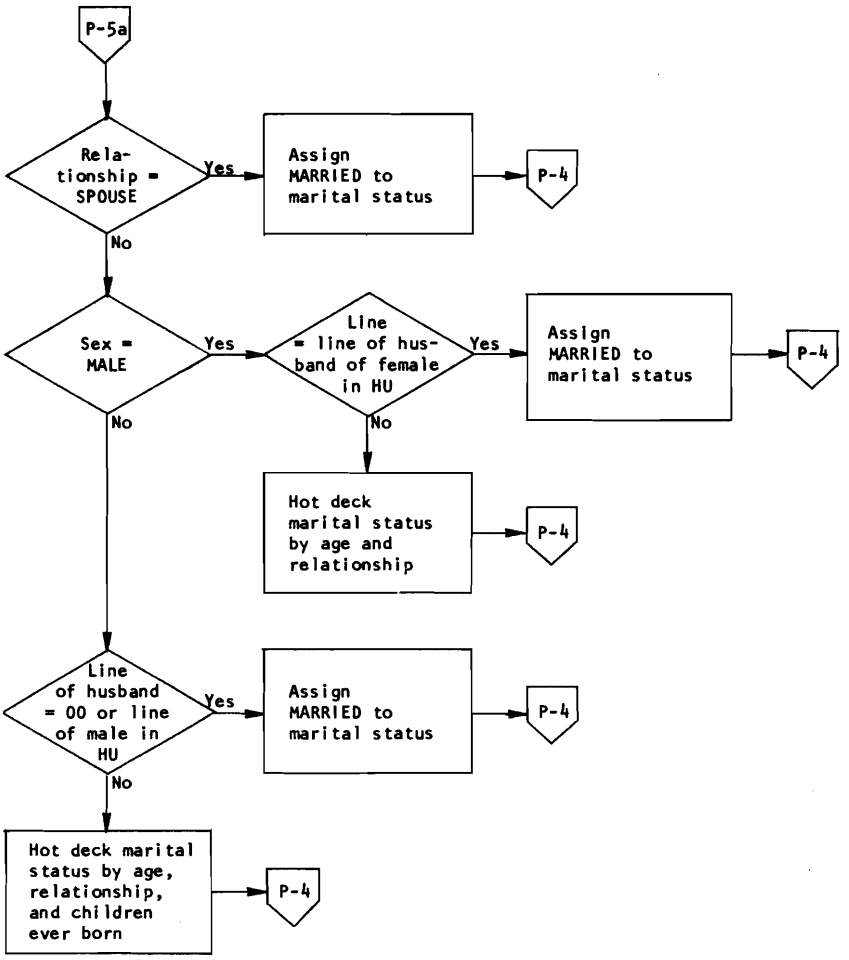


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

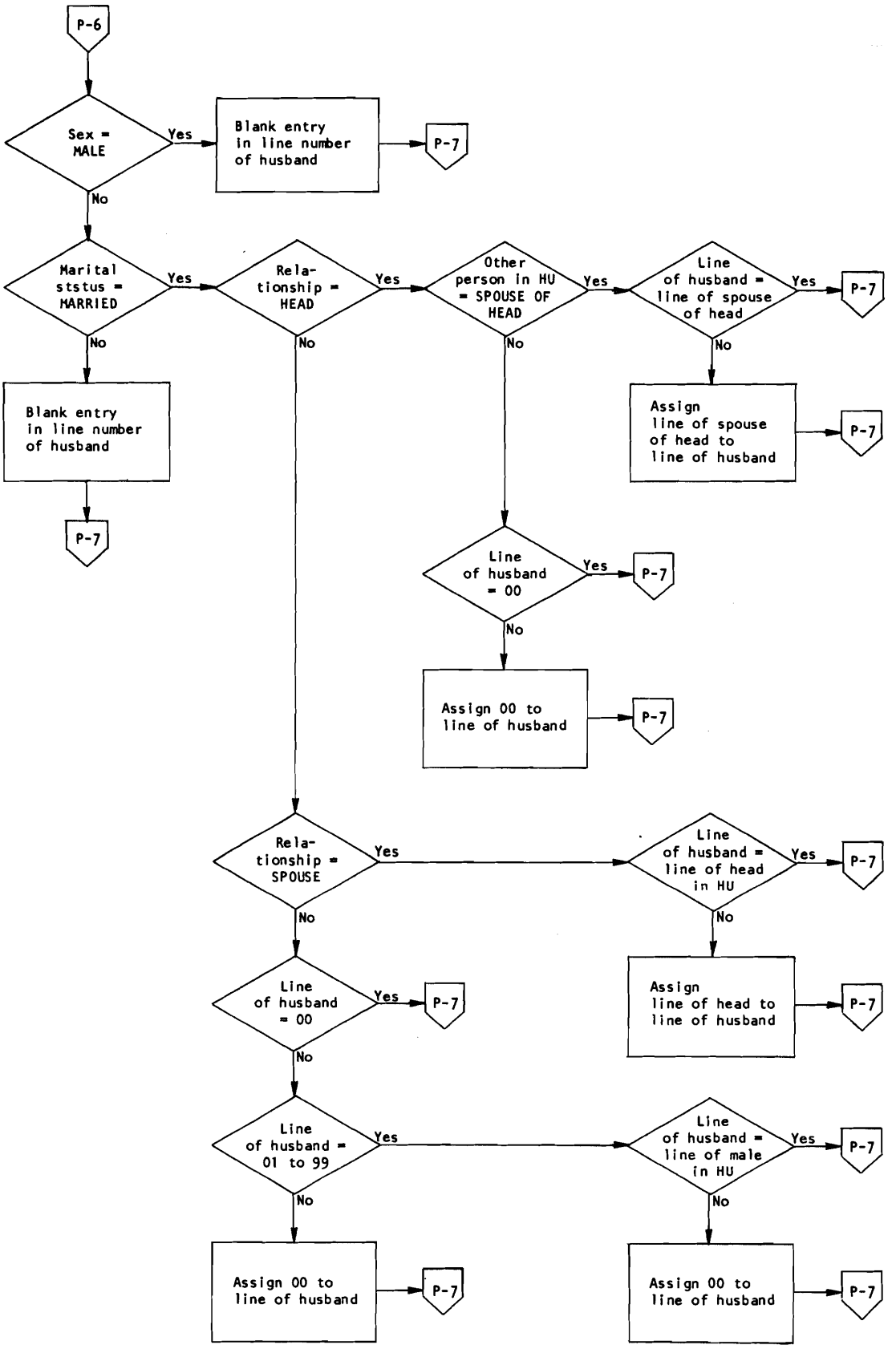


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

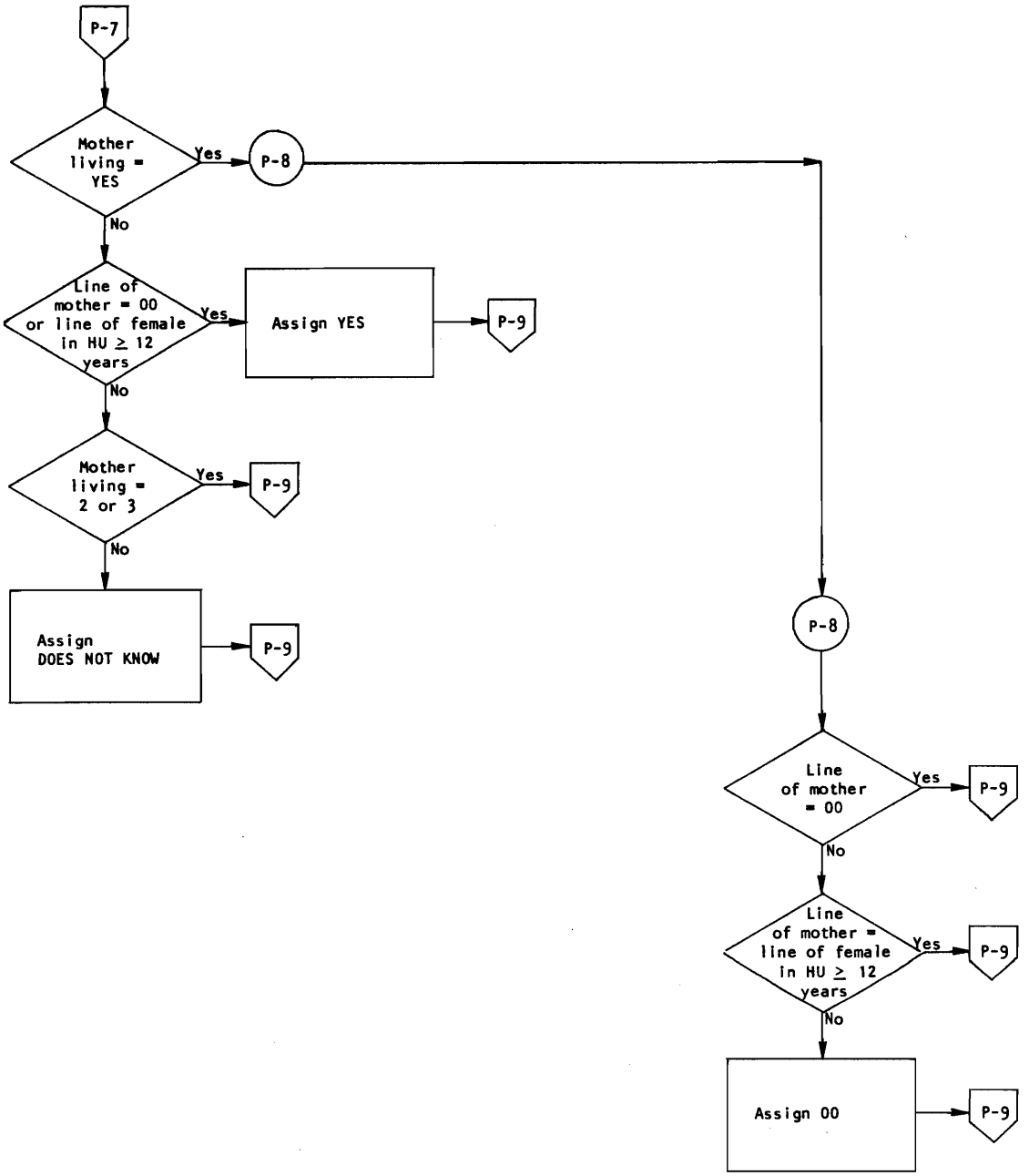


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

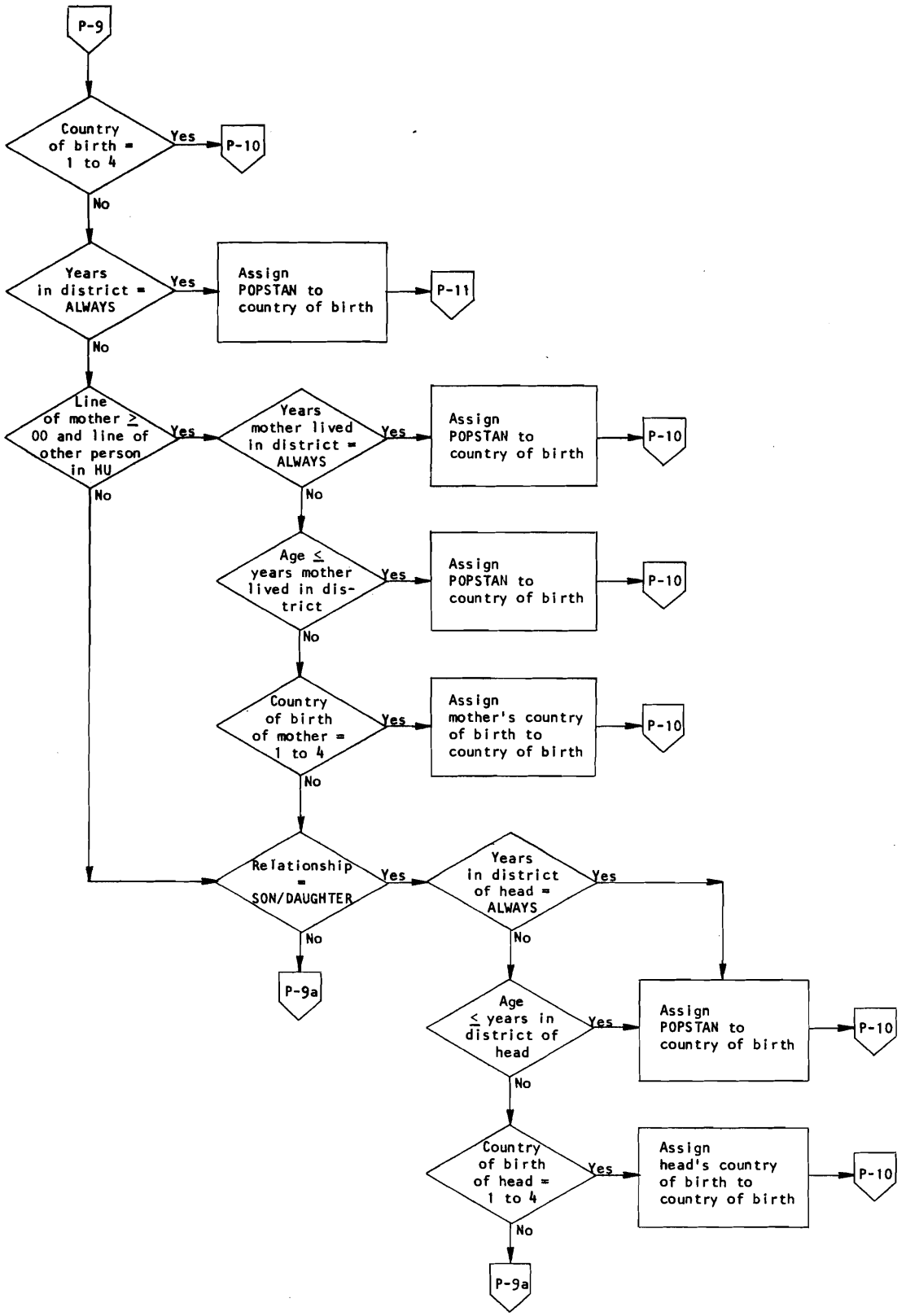


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

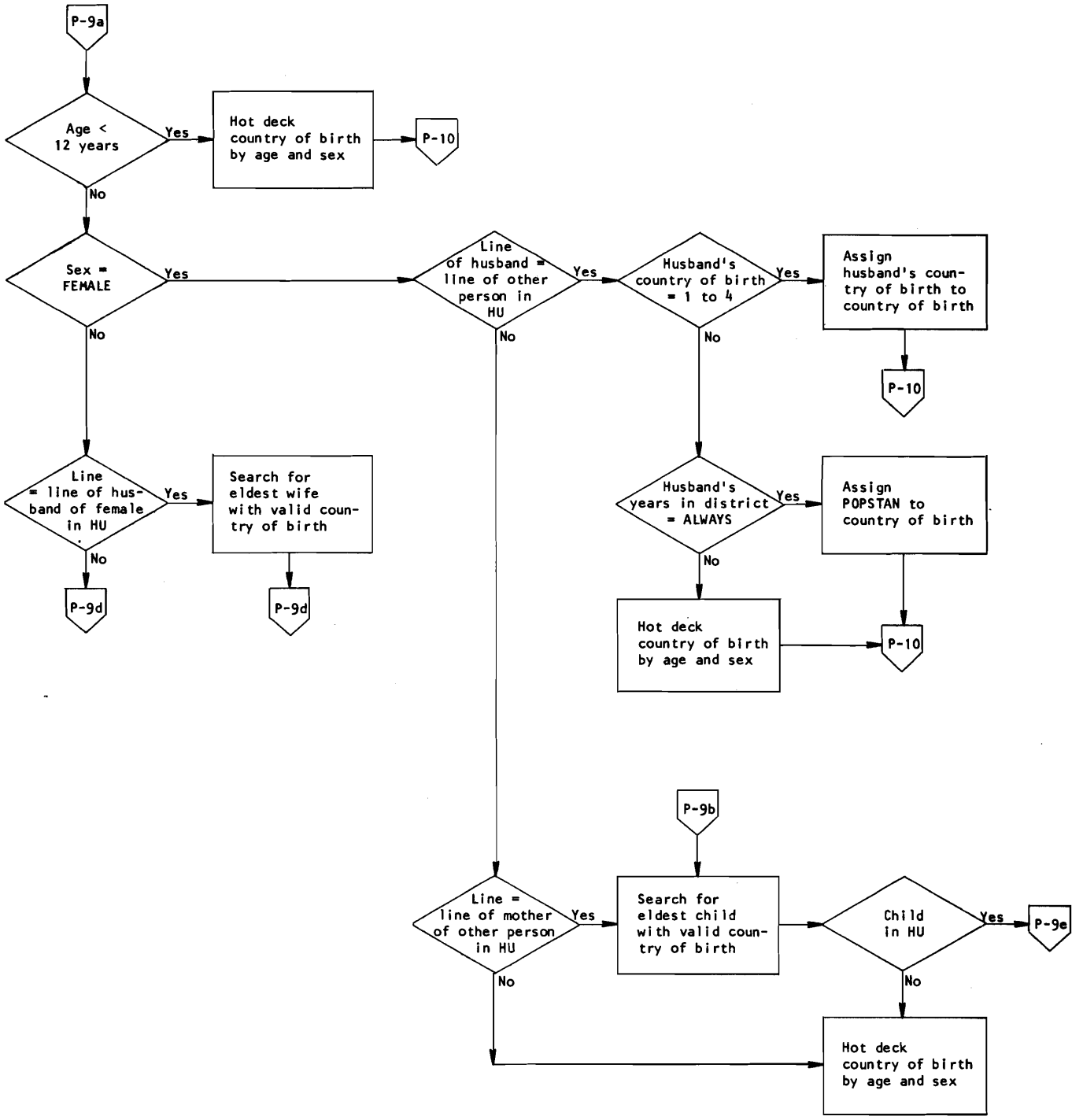
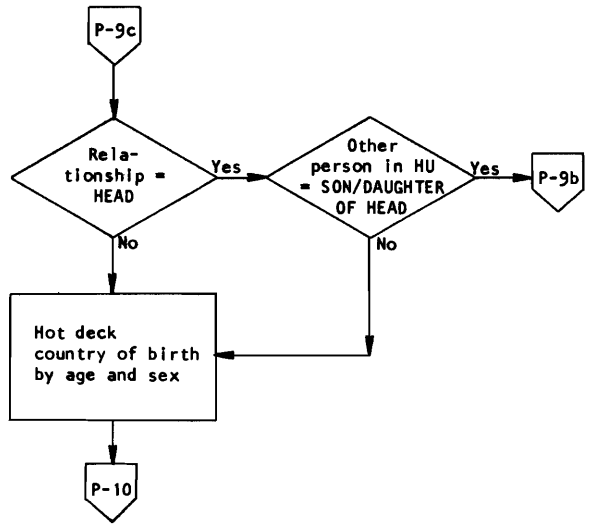
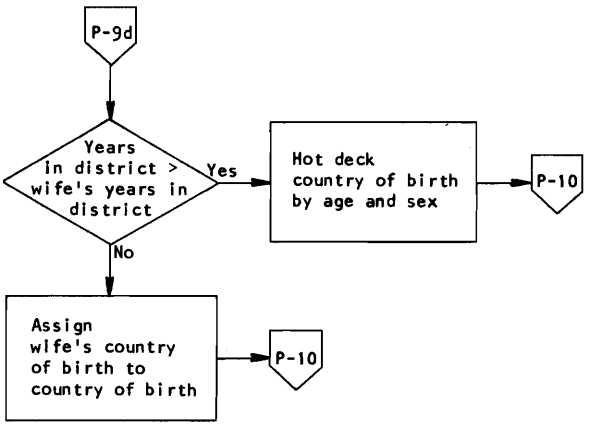


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued





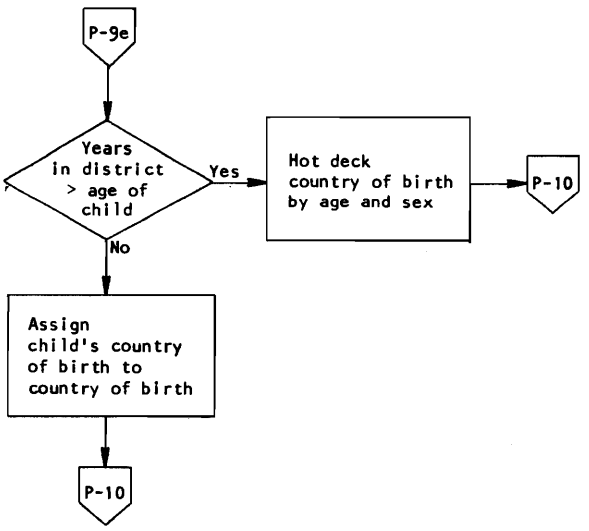


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

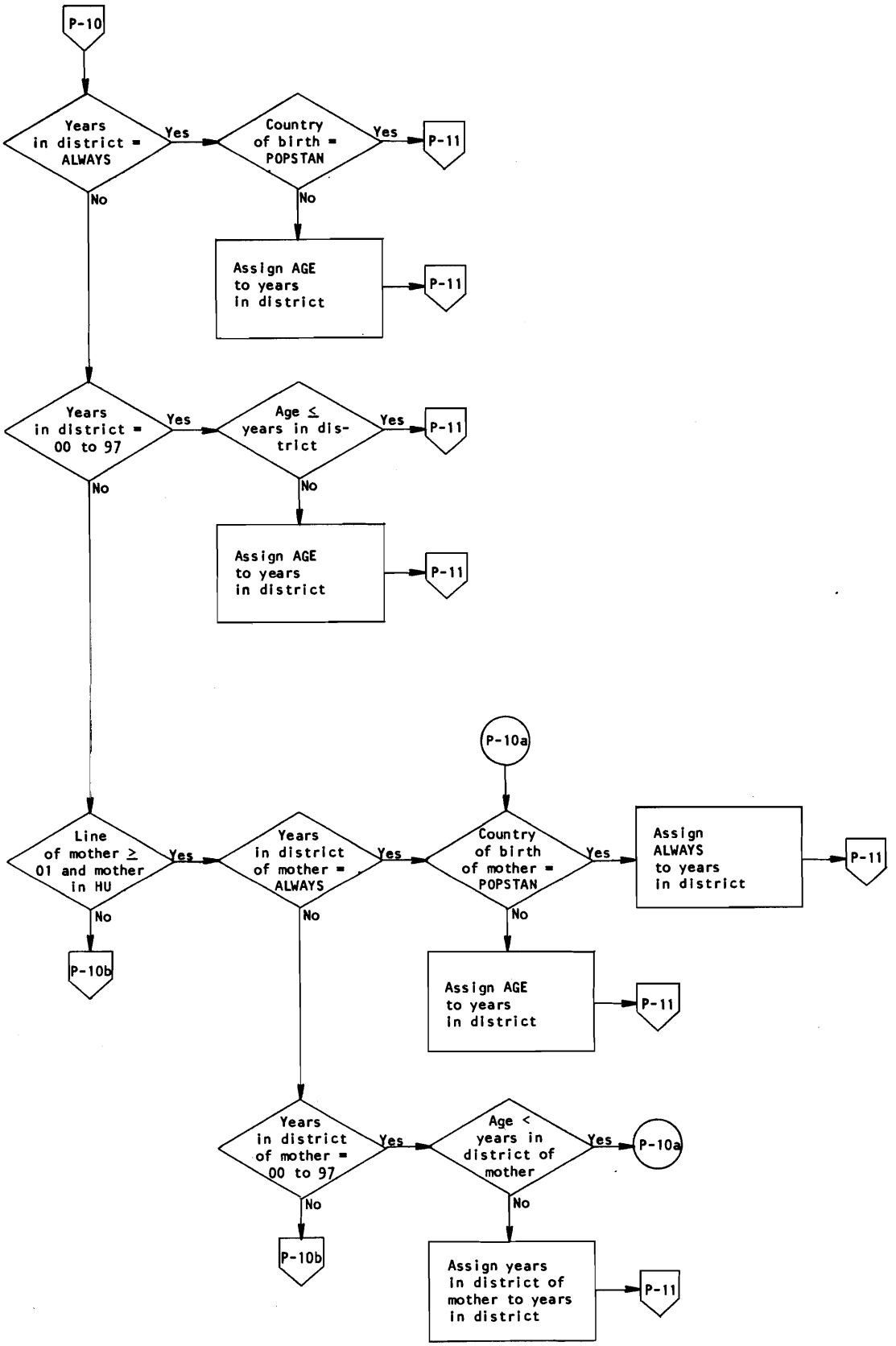
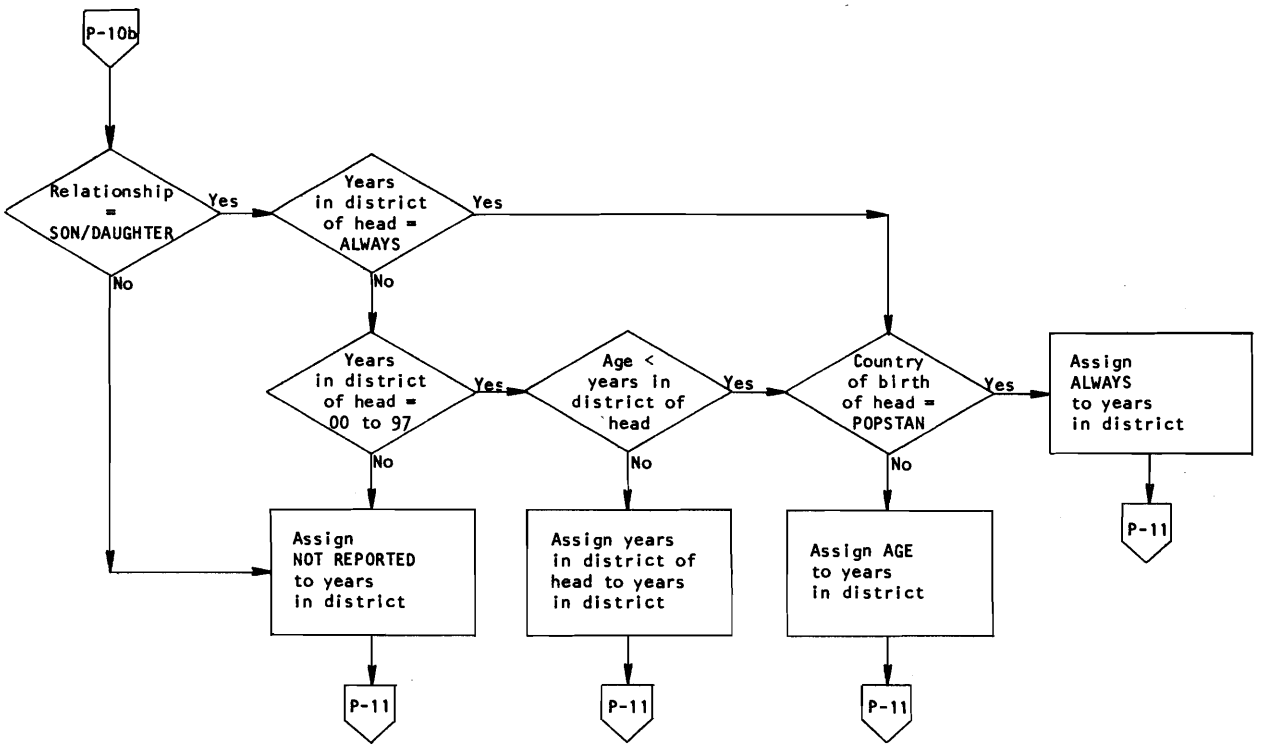


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued



OTHER EXHIBITS FOR REFERENCE

Exhibit C-2-1. INDIVIDUAL CENSUS REPORT

FORM PH-3
(JANUARY 1980)

MINISTRY OF ECONOMY
NATIONAL STATISTICAL OFFICE

INDIVIDUAL CENSUS REPORT
POPSTAN 1980 CENSUSES OF POPULATION AND HOUSING

CONFIDENTIAL – This inquiry is required by law. The information is accorded confidential treatment and cannot be used for taxation, investigation, or regulation.

ENUMERATOR INSTRUCTIONS: Complete items A through O.

A. Enumerator's name		**B.** Date	
C. Province	Code	**D.** District	Code
E. Barrio/village	Code	**F.** Place	Code
G. Ward/tract code	**H.** Block number	**I.** CL area number	
J. Type-of-place code	**K.** Metro area	Code	**L.** EA number
M. Number or name of street or road		**N.** Serial number	
O. House number, apartment number, location or description			

The National Statistical Office is making every effort to take a complete and accurate 1980 Census. The Census Taker was not able to obtain from you personally the information required for the Census. Please complete this form (items 1 through 15) as soon as you can. PLEASE PRINT. If this form is filled by someone else, the term "your" applies to the person for whom this form is intended. Seal the completed form in the envelope provided. Do not mail the form; the Census Taker will return in the next few days to pick it up. Thank you for your cooperation.

PLEASE PRINT

1. What is your full name? *Enter family name first.*	**4. District of your usual residence**
2. Where do you usually live; that is, where do you live most of the time? 1 – Usually live here } *Go to 11* 2 – Have no usual place of residence } 3 – Usually live elsewhere – *Continue with 3.*	**5. Barrio/village of your usual residence**
3. Province of your usual residence	**6. Urban place/hamlet of your usual residence**

Continue on back

PREVIOUS PAGE BLANK

Exhibit C-2-1. INDIVIDUAL CENSUS REPORT--Continued

7. Address of your usual residence *Give house number, apartment number, street or road name, or other description.*

8. At your usual residence, what are the names of nearby intersecting streets *(roads)?*

9. At your usual residence, what are the names of nearby landmarks *(rivers, churches, schools, etc.)?*

10. At your usual residence, what are the names of two of the nearest neighbors?

11. Who is head of the household at your usual place of residence? *Enter full name.*

12. What is your relationship to the head of household where you usually live? *Circle one.*

1 – Head

2 – Spouse of head

3 – Son or daughter of head/spouse

4 – Other relative

5 – Nonrelative

13. Sex *Circle one.*

1 – Male

2 – Female

14. What is your age as of last birthday? *If under 1 year, enter 00.*

_____ Years

15. What is your marital status? *Circle one.*

1 – Married

2 – Divorced

3 – Separated

4 – Widowed

5 – Never married

OFFICE USE ONLY	*Transcribed by* ☐ Enumerator ☐ DCO ☐ NSO	Name	Date
	INSTRUCTIONS: If code 1 or 2 in question 2, it is assumed that the information for the person is transcribed to the EA and Serial Number indicated in items C to N. If code 3 in question 2, enter in items C to N, in RED, the identification of the EA where the information was transcribed.		

Figure D-7e. MATCHING FORM

FORM PH-681 (NOVEMBER 1980)

MATCHING FORM

Name:

Date:

a. STRAT	PROV	DIST-C	DIST-RC	BARR/V	PLACE	WARD/TR	BLK	TOPL	MA	EA NO.	SAMPL SEC	SER NO.
391	42	05	053	006	079	127	16	2	9	143	10.1	2

b. PHASE II ADDRESS: 240 LOMA ROAD

c. PHASE I ADDRESS: 240 LOMA STREET

d. MATCH CODE FOR ADDRESS: 1

Line (1)	X in col. A (2)	Phase I location (3)	Initial match status (4)	Final match status (5)	Phase I Rel. (6)	Sex (7)	Age (8)	Mar. (9)	Comments (10)
01	X	EA 143 Serial 25 Line 01	☒ Match o Possible o Non-match	(1) - Match 2 - Non-match	H	M	43	M	
02	X	EA 143 Serial 25 Line 02	☒ Match o Possible o Non-match	(1) - Match 2 - Non-match	Sp	F	39	M	
03	X	EA 143 Serial 25 Line ~~04~~ 03	o Match ☒ Possible o Non-match	1 - Match (2) - Non-match	S/O	M	20	NM	FIRST SEEMED LIKE POSSIBLE MATCH TO LINE 04. CHANGED TO LINE 03 WHEN 04 WAS MATCHED TO OUT-MOVER.
04	X	EA 143 Serial 25 Line 05	☒ Match o Possible o Non-match	(1) - Match 2 - Non-match	S/D	F	13	NM	
05	X	EA 143 Serial 25 Line 06	o Match ☒ Possible o Non-match	(1) - Match 2 - Non-match	S/D	F	9	NM	
06	X	EA 143 Serial 25 Line 07	o Match ☒ Possible o Non-match	(1) - Match 2 - Non-match	S/D	M	7	NM	
07	X	EA 143 Serial 25 Line 08	o Match ☒ Possible o Non-match	(1) - Match 2 - Non-match	S/D	F	5	NM	
08	X	EA Serial Line	o Match o Possible ☒ Non-match	1 - Match (2) - Non-match					ISAM REPORTEDLY LIVED IN HU FOR MORE THAN A YEAR.
09	—	EA Serial Line	o Match o Possible o Non-match	1 - Match 2 - Non-match					
10		EA Serial Line	o Match o Possible o Non-match	1 - Match 2 - Non-match					
5-01 ~~11~~	OUT-MOVER	EA 143 Serial 25 Line 04	☒ Match o Possible o Non-match	(1) - Match 2 - Non-match	S/D	M	18	NM	
12		EA Serial Line	o Match o Possible o Non-match	1 - Match 2 - Non-match					

STATISTICAL TRAINING DOCUMENT
ISP-TR-4F

Popstan

A CASE STUDY FOR THE
1980 CENSUSES OF
POPULATION AND HOUSING

PART F: Variations for Local Conditions

(No.1) November 1983

U.S. Department of Commerce — BUREAU OF THE CENSUS

Errata

Page 32, Figure F-3c, col. 5, line 10

Reads: 170 to 130

Should read: 170 to 230

Page 56, col. 2, (4.19)

Reads: $s^2(x'') = \sum_{h-1}^{34} s^2\left(x''_h\right).$

Should read: $s^2(x'') = \sum_{h}^{34} s^2\left(x''_h\right).$

Page 67, col. 2, para. 2, line 3

Reads: Chapter A-9, Data Processing Considerations.

Should read: Chapter A-9, Data Processing.

Page 72, col. 2, para. 2, line 1

Reads: The first three types of reports can be

Should read: The first two types of reports can be

Page 81, col. 2, para. 2, line 19

Reads: the houseing unit. To handle this situation,

Should read: the housing unit. To handle this situation,

Page 83, col. 1, para. 2, line 16

Reads: and highest grade completed; and if less than

Should read: and highest grade completed; and if more than

Page 111, entire page

Replacement page 111 attached

Page 119, entire page

Replacement page 119 attached

Page 120, entire page

Replacement page 120 attached

Page 124, entire page

Replacement page 124 attached

Page 125, entire page

Replacement page 125 attached

Page 127, entire page

Replacement page 127 attached

Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

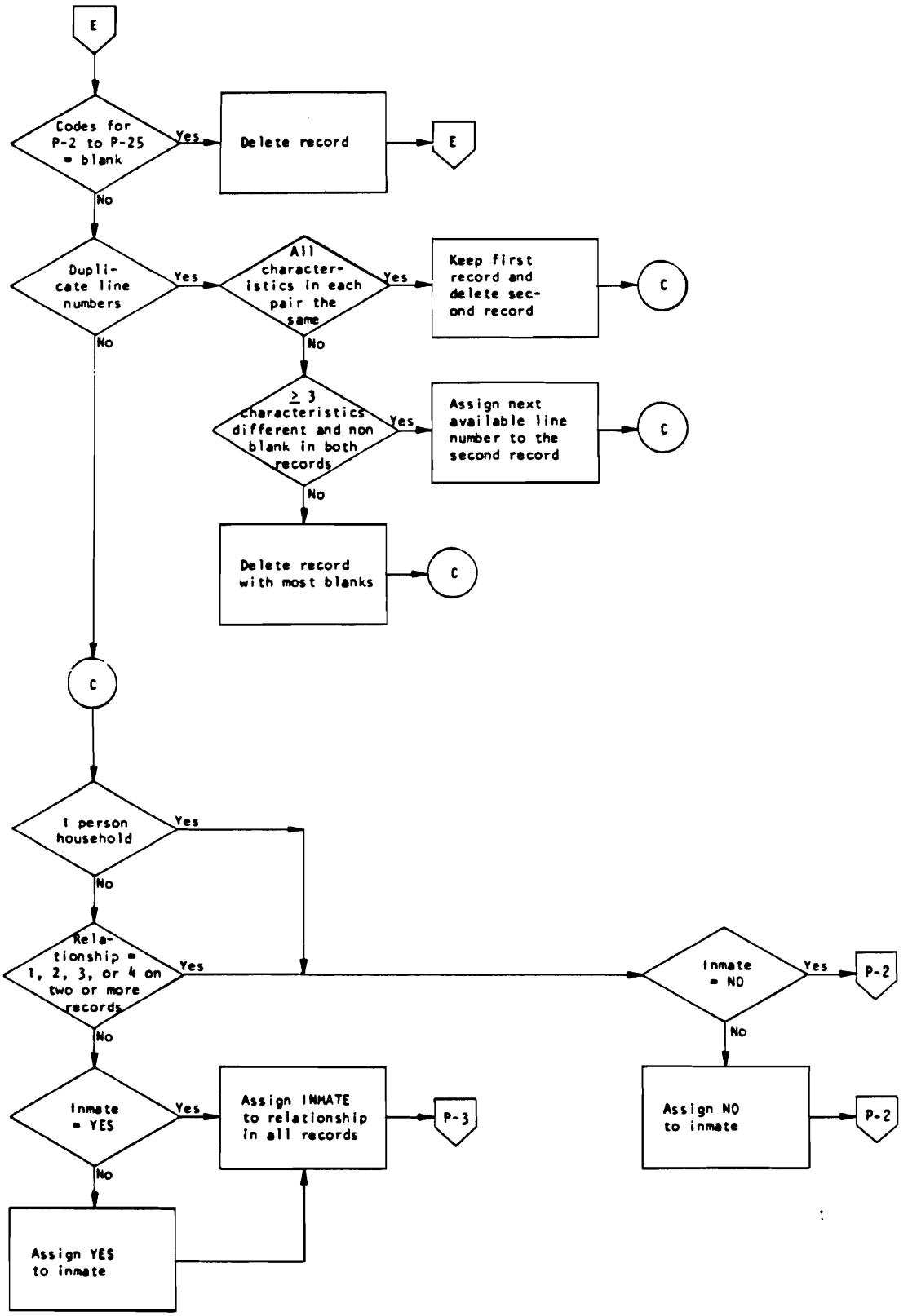


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

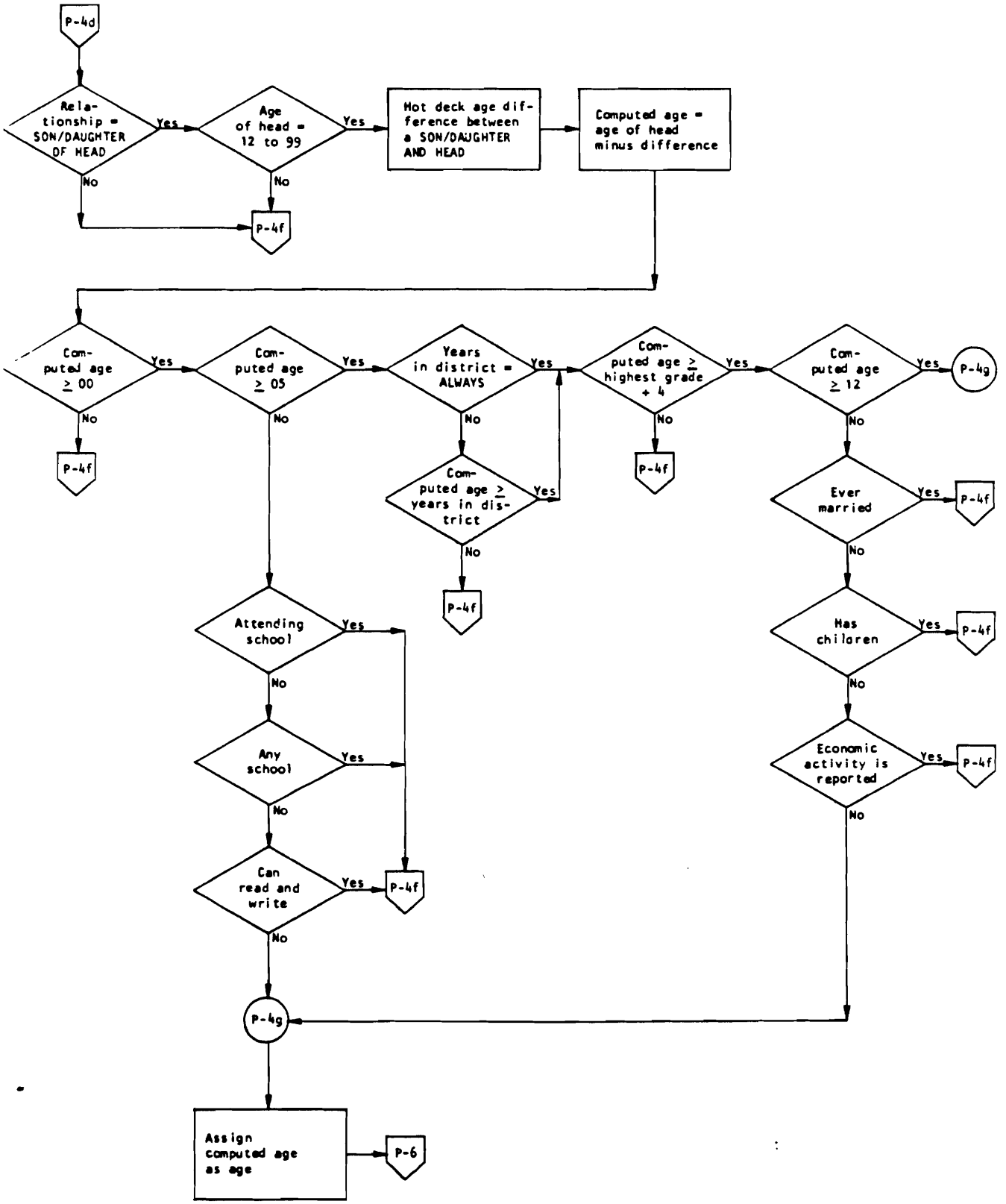


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

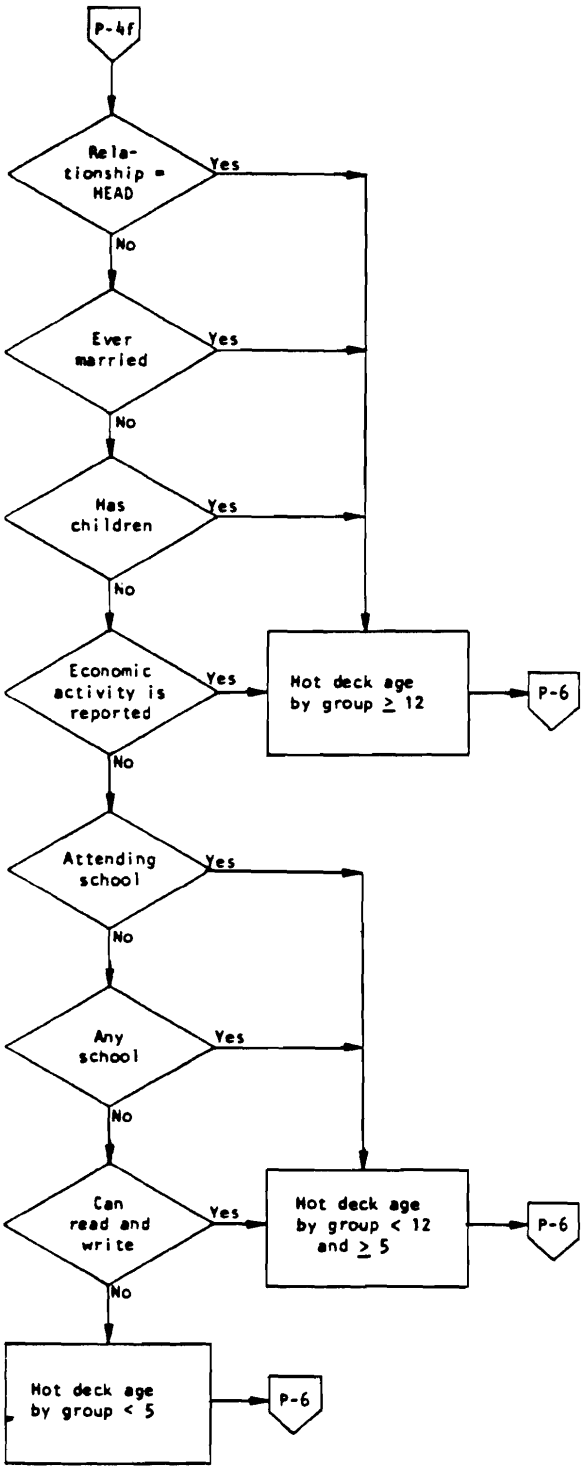


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

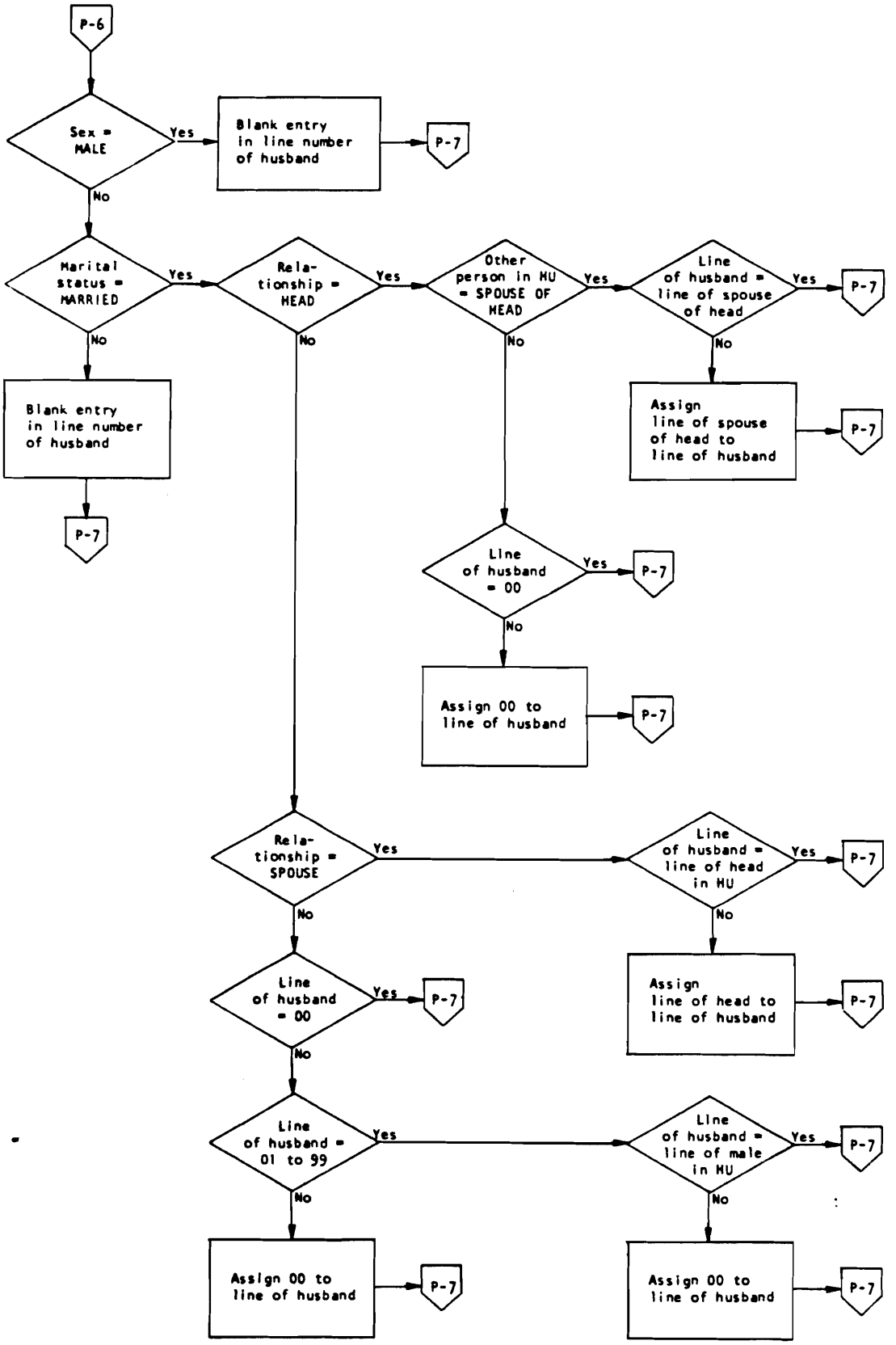


Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued

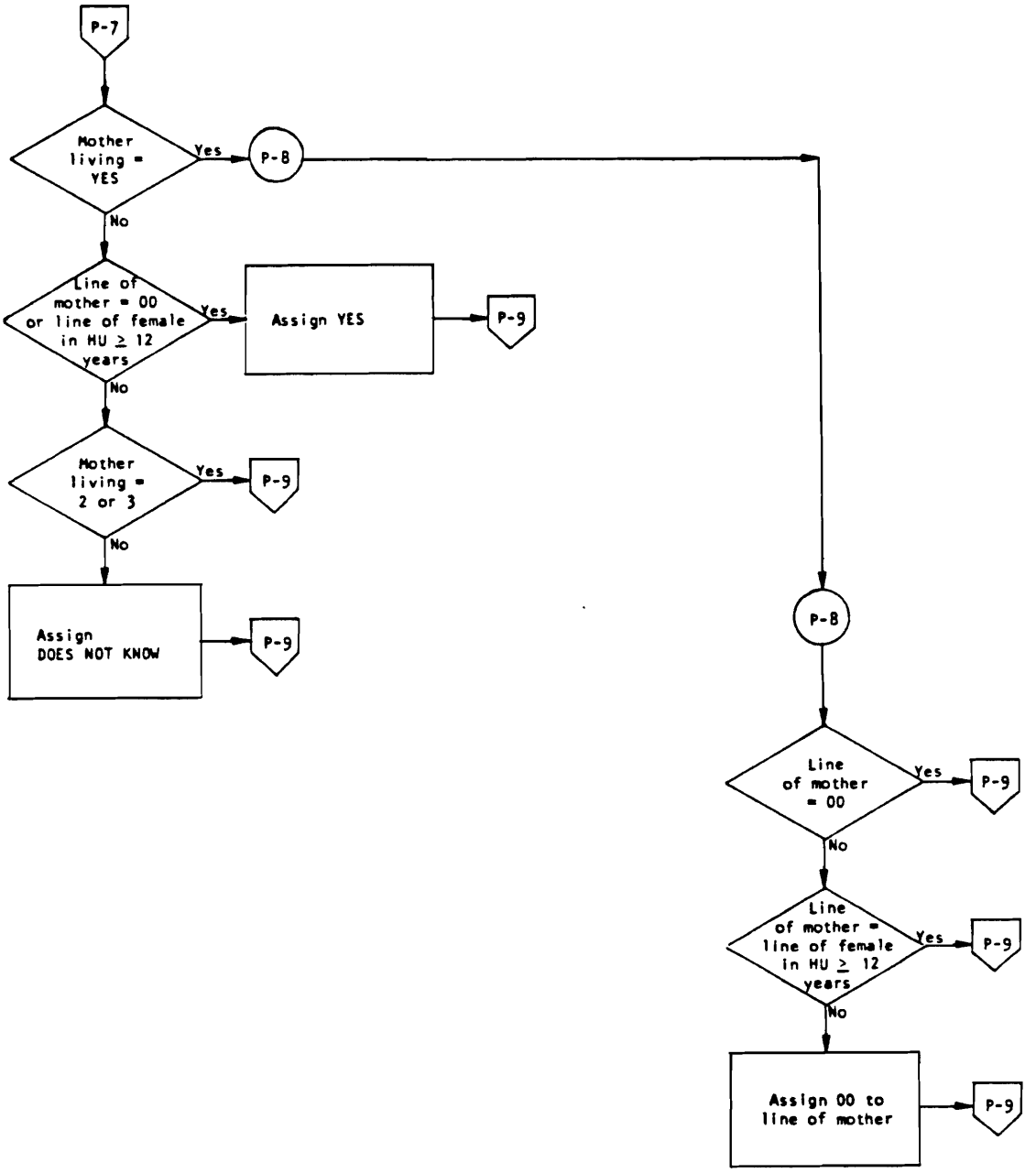
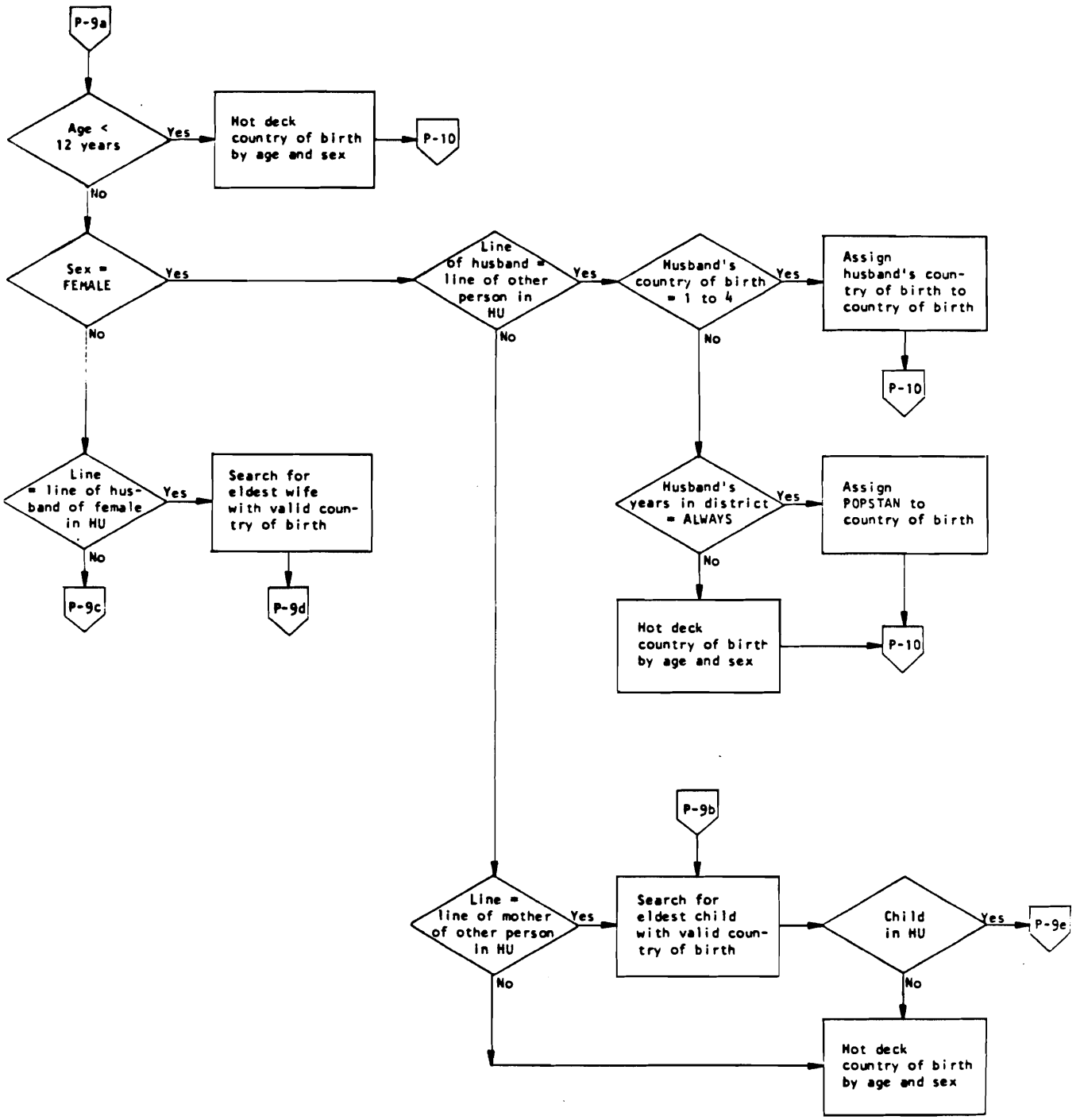



Exhibit F-6-1. FLOW CHARTS FOR POPULATION EDITS--Continued