24. *Ibid.*

25. David F. Bergwall, Philip N. Reeves, and Nina B. Woodside, *Introduction to Health Planning*, Washington, D.C.: Information Resources Press, 1974, pp. 170-171.

26. Donabedian, *op. cit.*, p. 264.

27. James R. Jeffers and Calvin D. Siebert, "Measurement of Hospital Cost Variation: Case Mix, Service Intensity, and Input Productivity Measures," *Health Services Research*, Vol. 9, Winter 1974, p. 294.

28. Jeffers and Siebert, *op. cit.*; Donabedian, *op. cit.*, p. 298; Feldstein (footnote 30), p. 39.

29. Donabedian, *op. cit.*, p. 252.

30. Martin S. Feldstein, *Economic Analysis for Health Service Efficiency*, Amsterdam: North-Holland Publishing Company, 1967, p. 129.

31. *Ibid.*, p. 134.

32. Donabedian, *op. cit.*, p. 256; Hirsch S. Ruchlin and Irving Leveson, "Measuring Hospital Productivity," *Health Services Research*, Vol. 9, Winter 1974, p. 322.

33. Donabedian, *op. cit.*, p. 551; Drosness, *op. cit.*, p. 45.

34. Drosness, *op. cit.*

35. *Ibid.*, p. 48.

36. P.L. 93-641, Section 1513 (a).

37. *Ibid.*, Section 1513 (a) (4).

38. Mara Minerva Melum, *Assessing the Need for Hospital Beds: A Review of Current Criteria*, Minneapolis: InterStudy, December, 1975, p. 32.

39. P.L. 93-641, Section 1513 (a).

40. P. Feldstein and J. German, "Predicting Hospital Utilization: An Evaluation of Three Approaches." *Inquiry*, June, 1965.

41. *Reference Manual for Project Review: Standards and Criteria*, New Orleans: Tulane University School of Public Health and Tropical Medicine, January 7, 1974, p. 91.

42. Drosness, *et al.*, *op. cit.*

43. Reference Manual for Project Review: Standards and Criteria, op. cit., p. 92.

44. F. R. Normile and H. A. Ziel, Jr., "Too Many OB Beds?" Hospitals, The Journal of the American Hospital Association, Vol. 44, July 16, 1970, p. 61.

45. Reference Manual for Project Review: Standards and Criteria, op. cit., p. 92.

Chapter 8

1. Bureau of Health Manpower, Criteria and Procedures for Identification of Critical Health Manpower Shortage Areas. Report No. 76-29, September 1975, p. 3.

2. Anthony Oreglia, et al., A Guide to the Development of Health Resource Inventories, West Lafayette, Indiana: Health Planning Data Project, Purdue University, 1976.

3. Ibid., Section III, D.

4. Federal Register, Vol. 40, No. 170, September 2, 1975, pp. 40315-415.

5. Ibid.

6. Telephone conversation with Mr. Richard C. Lee, Manpower Analysis Branch, Bureau of Health Manpower.

7. Federal Register, op. cit.

8. Bureau of Health Manpower, op. cit.

9. See footnote 6.

10. Bureau of Health Manpower, op. cit., p. 4.

11. Ibid., p. 5.

12. Federal Register, Vol. 40, No. 38, February 25, 1975, pp. 8155-8161.

13. Uwe E. Reinhardt, "Health Manpower Forecasting: Current Methodology and Its Impact on Health Manpower Policy," Manpower for Health Care. IOM Publication 74-003, National Academy of Sciences, Washington, D.C., May 1974, p. 22.

14. Roger I. Lee and Lewis W. Jones, The Fundamentals of Good Medical Care, Chicago: University of Chicago Press, 1933.

15. Marsha G. Goldfarb, A Critique of the Health Manpower Planning Literature. Working Paper Series No. 73-2, Office of Regional Activities and Continuing Education, Yale University School of Medicine, 1973.

16. W. Lee Hansen, "An Appraisal of Physician Manpower Projections," Inquiry, Vol. VII, No. 1, March 1970, pp. 102-113.

17. Reinhardt; Lee and Jones; Goldfarb; and Hansen, op. cit.; and Hyman K. Schonfeld, et al., "Numbers of Physicians Required for Primary Medical Care," New England Journal of Medicine, Vol. 286, No. 11, March 16, 1972, pp. 571-76; Henry R. Manson, "Manpower Needs by Specialty," Journal of the American Medical Association, Vol. 219, No. 12, March 20, 1972, pp. 1621-26; and "How Many People to Support a Specialist," Medical Economics, October 1967, pp. 65-80.

18. National Center for Health Statistics, "Current Estimates from the Health Interview Survey: United States - 1974," Vital and Health Statistics, Series 10, No. 100, DHEW Publication No. (HRA) 76-1527, September 1975, pp. 50-51.

19. Ibid., p. 25.

20. NCHS, "Physician Visits: Volumn and Interval Since Last Visit, United States - 1971," Vital and Health Statistics, Series 10, No. 97, DHEW Publication No. (HRA) 75-1524, March 1975, pp. 28-29.

21. Ibid., p. 17.

22. Ibid., p. 9.

23. C. N. Theodore, et al., Data Base of the AMA Center for Health Services Research and Development, Chicago: American Medical Association, October 16, 1971.

24. Center for Health Services Research and Development, Reference Data on Profile of Medical Practices: 1974 Edition, Chicago: American Medical Association, 1974, p. 183.

25. Center for Health Services Research and Development, Reference Data on the Profile of Medical Practice: 1971 Edition, Chicago: American Medican Association, 1971, p. 59.

26. Center for Health Services Research and Development, 1974, op. cit., p. 176.

27. Center for Health Services Research and Development, 1971, op. cit.

28. Richard A. Smith, et al., "Increasing Physician Productivity and the Hospitalization Characteristics of Practices Using MEDEX-A Progress Report," Northwest Medicine, Seattle, Washington, October 1971, pp. 701-706; Uwe E. Reinhardt and Donald E. Yett, Physician Production Functions Under Varying Practice Arrangements, Technical Paper Series #11, Community Profile Data Center, DHEW, January 1971, p. 40; and Uwe E. Reinhardt, op. cit., p. 28.

Chapter 9

1. National Center for Health Statistics, "Health Manpower: United States - 1965-1967," Vital and Health Statistics, Series 14, Number 1, U.S. Department of Health, Education and Welfare, Public Health Service, 1968, p. 2.

2. National Center for Health Statistics, Health Resources Statistics: Health Manpower and Health Facilities, 1974, U.S. Department of Health, Education and Welfare, Public Health Service, 1975, p. viii.

3. Ibid., p. 3, "Health Manpower: United States - 1965-1967," op. cit., p. 1.

4. Ibid.

5. Adapted from Glossary of Hospital Terms, Chicago: American Association of Medical Record Librarians, 1969, p. 15.

6. NCHS, Nursing Homes: A County and Metropolitan Area Data Book, 1971, DHEW Publication No. (HRA) 74-1224, 1974, p. 2.

7. Appendix G in Anthony Oreglia, et al., A Guide to the Development of Health Resource Inventories, Purdue University, Department of Sociology, Health Services Research and Training Program, West Lafayette, Indiana, February 1976.

8. Adapted from Oklahoma Health Manpower Needs, Occupational Training Information System, Division of Research Planning and Evaluation, State Department of Vocational Technical Education, Stillwater, Oklahoma, 1973.

9. See, for example: Thomas N. Chirikos, Allied Health Manpower in Ohio, 1972, Ohio Advisory Council on Vocational Education; Vincent Martucci, Health Manpower in New Jersey, New Jersey Department of Health. December, 1971; Oklahoma Health Manpower Needs (op. cit.); and Kirsten Bradley, et al., Health Manpower, 1971, Virginia, Coordinated Health Survey Committee.

10. cf. American Universities and Colleges (11th Edition), Washington, American Council on Education, 1973; and American Junior Colleges, Washington, American Council on Education, 1972.

11. Allied Health Manpower in Texas, 1973: A Report on Manpower Requirements, Resources and Education, Texas Hospital Association, Texas Medical Foundation, Regional Medical Program of Texas, 1973, p. 14.

12. P.L. 93-641, Section 1532 (c) (6).

This completes volume I of the Handbook. The remaining chapters of the Handbook (i.e., chapters 10 through 15) can be found in volume II.

☆U.S. GOVERNMENT PRINTING OFFICE: 1976 240-848/43 1-3

A word is in order concerning the specific audience to which this Handbook is addressed. Primarily for clarity and consistency of presentation, the health planner addressed throughout the document is assumed to be employed in a health systems agency rather than a State Agency. However, it is recognized that several of the studies described might more feasibly be conducted at the State level. Where this is the case, the studies can be adapted for State Agency implementation with a minimum of effort.

It was decided that, for purposes of illustrating the analysis and use of data, concrete examples would be more effective than blank tables. Therefore, an example health service area, called Central HSA, was created using realistic but fictitious data. The most salient demographic characteristics of Central HSA are that it contains a population of approximately two million and that it consists of thirty counties, including four SMSAs. Therefore, Central HSA is above average in size and representative of both rural and urban dwellers.

This publication presupposes the implementation by HSAs of the inventory components described in the companion publication, A Guide to the Development of Health Resource Inventories. Many of the data compilation and collection activities and the analyses described in the Handbook depend upon the completeness of inventory coverage and the accuracy of the information obtained during the process of inventory compilation.

We welcome comments or questions concerning the Handbook. Comments should be directed to: Division of Planning Methods and Technology, Bureau of Health Planning and Resources Development, Health Resources Administration, Parklawn Building, Room 12-14, 5600 Fishers Lane, Rockville, Maryland 20852.

Z. Erik Farag, Ph.D.
Director, Division of Planning
Methods and Technology
Bureau of Health Planning and
Resources Development

PREFACE

This publication is divided into four sections. Section I, the Introduction, contains two chapters. The first chapter explains the report's objectives and the second discusses general methodological issues. The remaining sections (II-IV) correspond to the categories of an information framework for health planning (see Chapter 1, Section I). Within each of these three sections, the first chapter discusses the need for the type of data being covered in that section with reference to purposes and uses for the data primarily as required by legislatively defined agency functions. Subsequent chapters detail procedures for generating a portion of the type of data designated by the section title (e.g., in Section III, Chapter 7 describes how to obtain and analyze hospital data which fulfill a portion of the need for health resources data).

All the activities based upon assemblage of extant data are contained in Sections II and III (Volume I); all the studies based upon primary data collection, with the exception of the Study of the Demand for Health Manpower in Hospitals and Nursing Homes, are contained in Section IV (Volume II). A particular temporal execution is dictated neither by the ordering of Sections II through IV with respect to each other, nor by the within-section ordering of chapters. Thus, an agency may very well conduct the study of ambulance services described in Chapter 15 of Section IV before consulting sources of extant data on hospitals (Chapter 7, Section III).

Each chapter is intended to be substantially complete within itself; therefore, a certain amount of redundancy or repetition will be found. On the other hand, cross-references to useful information contained in another chapter or in A Guide to the Development of Health Resource Inventories have been included wherever appropriate.

The within-chapter arrangement of topics differs for Sections II and III, on the one hand, and Section IV, on the other, due to the fact that the former deal with extant data and the latter with primary data collection.

In Sections II and III, the topics contained in each chapter are as follows: selection and definition of data items, description of data sources, and data analysis and use. In Section IV, the topical organization of chapters is necessarily more complex, reflecting the greater complexity inherent in a primary data collection effort. Part I describes the study methodology, covering the definition of concepts and variables, the study instrument, the data collection design, field procedures, and data processing and storage. Part II discusses data analysis and use, usually with reference to specific examples of planning, resource development or project review decisions to be made by HSAs.

Appendices are found at the end of most chapters. These contain supplementary materials such as instructions to data collection personnel, codebooks, and other relevant documentation. When content footnotes are used, they appear at the bottom of the appropriate page. Bibliographic references appear at the end of each section.

ACKNOWLEDGMENTS

The staff of Purdue's Health Planning Data Project, Health Services Research and Training Program, would like to thank specific individuals who contributed to our preparation of this Handbook.

In particular we wish to acknowledge the active support and continuing advice of our Project Officer, Alan R. Boissy of the Bureau of Health Planning and Resources Development, Health Resources Administration. His concern with detail and accuracy has helped to keep high quality continuously before us as a goal to be achieved. Co-Project Officer Robert H. Mugge of the National Center for Health Statistics also has given us valuable assistance.

We also wish to thank Julia A. James, Hyman Luden, Peter D. Mott, Philip N. Reeves, Caesar A. Ricci II, and George E. Schwarz. As consultants to the Health Planning Data Project, these persons have generously given their advice and recommendations concerning health planning data issues

Our staff acknowledges the help of the above-named individuals while retaining full responsibility for the accuracy of the information contained in this Handbook.

CONTENTS

Page

FOREWORD. ii
PREFACE v
ACKNOWLEDGMENTS vii

SECTION I. INTRODUCTION. 1

CHAPTER 1. THE NEED FOR DATA IN HEALTH PLANNING. . . 1

THE PROBLEM. 3
LEGISLATIVE REQUIREMENTS OF P.L. 93-641. 4
HEALTH PLANNING AGENCY FUNCTIONS 5
INFORMATION FRAMEWORK. 5
STUDY SELECTION 6

CHAPTER 2. METHODOLOGICAL ISSUES RELATED TO DATA COLLECTION AND ANALYSIS 9

INTRODUCTION 11

THE RESEARCH PROCESS I: PRELIMINARY PHASE . . . 13

The Statement of Research Objectives. . . . 13
Feasibility Assessment. 14

THE RESEARCH PROCESS II: STUDY IMPLEMENTATION . 18

The Study Plan. 18
Definitions 19
The Consistency Issue 20
Defining the Universe and Selecting Units of Analysis 22
Methods of Enhancing Cooperation. 26
Instrument Design 28
Data Collection 30
Data Processing and Storage 31

THE RESEARCH PROCESS III: DATA ANALYSIS AND USE 33

Data Display 33
Data Summarization. 35
The Confidence Interval 43
Testing Differences 46

Page

FOOTNOTES. 53

SECTION II. POPULATION, SOCIOECONOMIC AND HEALTH STATUS DATA. 55

CHAPTER 3. THE NEED FOR POPULATION, SOCIOECONOMIC AND HEALTH STATUS DATA 55

POPULATION AND SOCIOECONOMIC DATA 57

HEALTH STATUS DATA. 59

Health Indicators. 59
Health Indicators and Health Planning. . . . 61

CHAPTER 4. THE SOURCES AND USES OF POPULATION AND SOCIOECONOMIC DATA 65

SELECTION AND DEFINITION OF DATA ITEMS. 67

DESCRIPTION OF DATA SOURCE. 69

Subject Content. 69
Geographic Coverage. 70
Data Products. 76
The Availability of Specific Data Items. . . 81

Total Population. 82
Age and Sex 83
Race. 87
Income. 90
Education 93

DATA ANALYSIS AND USE 97

Denominator Data 97
Sampling Applications. 101

Representativeness of Survey Samples. . 102
Obtaining a Stratified Survey Sample. . 102

Page

The ADMATCH Program and the Uses of Geocoded Data. 106

The Identification of Populations-at-Risk and the Location of Health Facilities 108
The Development of Neighborhood Health Profiles 109
Geocoded Data and Computer Graphics. . . 110
The Linkage of Geocoded Health Data with Census Data 113
A Case Study of the ADMATCH Program. . . 113
Inferences Drawn from Aggregated Data and the Ecological Fallacy 119

APPENDIX A A Description and Assessment of an Actual Application of the ADMATCH Program. 123

APPENDIX B Birth, Fetal, and Death Certificates. . . . 127

APPENDIX C The Ecological Fallacy. 132

CHAPTER 5. THE SOURCES AND USES OF HEALTH STATUS DATA. 135

MORTALITY DATA

Selection and Definition of Data Items. . . . 137
Description of Data Sources 139
Data Analysis and Use 140

MORBIDITY DATA 144

Selection and Definition of Data Items. . . . 144
Description of Data Sources 146
Data Analysis and Use 152

DISABILITY DATA. 155

Selection and Definition of Data Items. . . . 155
Description of Data Sources 157
Data Analysis and Use 160

COMPOSITE INDICATORS OF HEALTH STATUS. 165

FOOTNOTES 171

Page

SECTION III. HEALTH RESOURCES DATA. 175

CHAPTER 6. THE NEED FOR HEALTH RESOURCES DATA . . . 175

HOSPITAL DATA 177

PHYSICIAN DATA. 178

INSTITUTIONAL HEALTH MANPOWER DATA. 181

CHAPTER 7. THE SOURCES AND USES OF HOSPITAL DATA. . 183

SELECTION AND DEFINITION OF DATA ITEMS. 185

DESCRIPTION OF DATA SOURCES 190

DATA ANALYSIS AND USE 196

Descriptive Uses of the Data 197

Availability of Hospital Beds 198
Availability of Hospital Facilities, Services and Special Beds 208
Distribution of Hospital Beds, Services and Facilities 218

Measures of Hospital Utilization 227

The Admission Rate. 228
The Case Flow 230
The Occupancy Ratio 232
The Length of Stay. 234
The Utilization Rate. 235

The Capacity of the Hospital to Produce Service. 237

Concepts Related to Hospital Service Capacity. 238
A Model for the Assessment of a Hospital's Service Load Relative to its Capacity to Produce Service. 239
Factors Affecting the Hospital's Maintenance of Standby Reserve. . . 241

Page

Measures of Service Load and Service Capacity. 246
Analysis and Use of Service Load and Service Capacity Data in Project Review and Resource Development Decisions 254

Bed Need Projections 268

The Hill-Burton Formula 268
Utilization-Adjusted Hill-Burton Formula 270
Utilization/Occupancy-Adjusted Hill-Burton Formula. 274
The Normile Model 277
The Commission on Hospital Care Method. 283
Summary of Bed Need Projections . . . 283

CHAPTER 8. THE SOURCES AND USES OF PHYSICIAN DATA . 287

SELECTION AND DEFINITION OF DATA ITEMS. 289

DESCRIPTION OF DATA SOURCES 292

DATA ANALYSIS AND USE 294

Assessing the Availability of Physicians and Identifying Shortage Areas 295

Identification of Medically Underserved Areas (MUAs) 295
Identification of Critical Medical Shortage Areas (CMSAs). 301

Estimating Future Requirements for Primary Care Physicians. 307

APPENDIX A Specialty Categories of Physicians 323

CHAPTER 9. A STUDY OF THE DEMAND FOR HEALTH MANPOWER IN HOSPITALS AND NURSING HOMES. 325

STUDY METHODOLOGY 327

Page

Definition of Concepts. 327
Definition of Variables 329
Study Instruments 337
Data Collection Design. 347
Field Procedures. 350
Data Processing and Storage 352

DATA ANALYSIS AND USE. 355

Identifying Manpower Shortages. 355
Assessing Hospital and Nursing Home Staffing Patterns 361

FOOTNOTES 367

SECTION I

INTRODUCTION

Chapter 1

The Need for Data in Health Planning

At this point in the development of health planning there is no need to justify the use of data in the planning process. It has become an implicit assumption of planners and policymakers that better, if not more, data will permit more rational and more acceptable planning and decision-making about health care delivery. A second implicit assumption is that more rational decision-making regarding the organization of the health care system and the delivery of health services will ultimately lead to an improvement of the system—improvement defined variously as healthier people, increased access to the health care system by medically underserved populations, deceleration in the rate of increase of health care delivery costs, increased satisfaction of the population, or all four.

The purpose, then, of this introduction is not so much to justify the need for, or use of, data in health planning as it is to explain or justify the need for a data collection and analysis handbook for health planners, and for this one in particular. It will, therefore, discuss the question of why this Handbook was written, specifically covering: 1) the problem as presently understood; 2) the general legislative requirements of P.L. 93-641, which form the present, pragmatic decision setting within which the problem is to be addressed; 3) health systems agency (HSA) and State Agency functions as specified in P.L. 93-641, which provide a further context for addressing the problem; 4) the framework selected for classifying the types of information needed by planners in view of required agency functions and decisions, and 5) the use of this information framework to suggest particular studies for inclusion in the Handbook.

The Problem

One important problem facing health planners in carrying out their data-related activities can be formulated rather precisely: data related to health care delivery exist or are obtainable but there is a fundamental lack of understanding about how to integrate the intervening steps between the selection of a particular study and the decisions that are derived once the analyzed data from such a study are presented, for example, to the board of a health systems agency. The most important of these intervening steps, (i.e., study design, data collection, data analysis, interpretation and decision-making) are discussed in detail in this Handbook.

It is recognized and acknowledged, however, that decisions are not merely the consequence of interpreting analyzed data. Actions inconsistent with the interpretation of study data are often taken. It is the political process which is the chief contender with "objective" data for influence on decision-making. It was, for the most part, beyond the scope of this Handbook to deal with the implications of the social and political context of planning and decision-making. The objective of the Handbook, then, is solely to improve the execution of the steps from study selection to data use, and to explore the implications of standard types of decisions, as called for by law and experience, upon these steps.

Legislative Requirements of P.L. 93-641

On January 4, 1975, the President signed into law the National Health Planning and Resources Development Act (P.L. 93-641) which established health systems agencies and State Agencies. Its requirements were chosen as the basis of efforts to link data and decisions by means of the information framework described below. P.L. 93-641 contains, among other things, directives regarding: 1) the general functions of HSAs and State Agencies; 2) the policy purposes of carrying out these functions; 3) the specific areas, related to population needs and system resources, about which HSAs will assemble and analyze data for planning and resource development, and 4) the criteria to be used by agencies in conducting project reviews. Therefore, while planning agency decisions are not specifically stated in the law, they can be inferred from the above, as well as from historical precedents established by planners due to planning and review experience under Section 1122 of the Social Security Amendments of 1972 (P.L. 92-603) and the Comprehensive Health Planning and Public Health Service Amendments of 1966 (P.L. 89-749). In cases where these were relevant, the decisions implied or required by other pieces of legislation were also considered.

Health Planning Agency Functions

P.L. 93-641 specified several functions for HSAs. Although these functions overlap and are interrelated, they fall into one of three broad categories: planning, resource development and project review. Some specific examples of actions based upon these functions are the development of a health systems plan (HSP); the development of an annual implementation plan (AIP) which describes objectives to achieve the goals of the HSP; the review and approval of proposed uses in the health service area served by the HSA of certain Federal funds; the making of recommendations to the State Agency regarding the need for proposed new institutional health services, and the periodic review of existing institutional health services. State Agency functions are similarly set forth in P.L. 93-641. Examples of particular decisions of the sort which make up these types of actions are discussed in the chapters which follow.

Information Framework

Planning, resource development and project review decisions should form the basis for data collection, if the data are ultimately to be relevant to action. Therefore, our approach within each section which follows has been to first specify the general purposes or uses of data in the context of such decisions. The data themselves are classified with reference to types of health information.

The framework selected for the organization of the content of this Handbook is a simple one consisting of three major parts: Population, Socioeconomic and Health Status Data; Health Resources Data; and Health Services Utilization Data. The reasons for the selection of this framework were its comparability with the majority of frameworks for classifying health data supported by expert opinion (national committees, federal agencies, and academics); its consistency with the provisions of P.L. 93-641 and preceding health planning legislation, and the fact that it contained categories by which health data are normally collected and disseminated by federal agencies and other national organizations. It should be noted that the three information categories are not discrete in actual use. For example, population data are used as denominators in calculating rates of utilization for types of health resources and services.

Study Selection

The particular chapters included in the Handbook represent an operational definition of the information types specified above. Population, socioeconomic, and health status data are discussed primarily in terms of procedures for obtaining and analyzing extant data available from the U.S. Bureau of the Census, the National Center for Health Statistics, and state boards of health.

Health resources are discussed in terms of selected types of health facilities and health manpower. The type of facility selected for emphasis in the Handbook is the general hospital, since decisions related to the allocation of these resources are acknowledged by planners to be of the greatest significance. Hospitals are discussed in terms of the analysis of extant data obtained from one or more of several sources. The compilation and analysis of resource data on specialty hospitals, nursing homes and other health facilities are discussed in the companion publication to this Handbook, A Guide to the Development of Health Resource Inventories. The health manpower types selected are physicians in a chapter which discusses the analysis of extant data, and those categories of manpower employed in health care facilities in a study based upon primary data collection.

In considering the use of health services, these services are defined as 1) those rendered in ambulatory care settings (hospital outpatient departments and emergency rooms and physicians' offices), 2) those provided in general hospitals, and 3) those provided in nursing homes. A fourth study uses ambulance services as an example of a type of special service in which considerable interest on the part of health planners exists. A final study of the use of health services, "A Family Health Survey," approaches utilization from the standpoint of the population rather than the provider. The unit of analysis in all these studies is the individual case, discharge, etc., rather than the health resource perspective of the facility. All are based upon primary data collection.

In recognition of the fact that HSAs may wish to obtain or collect data by means other than those specifically discussed in this Handbook, Chapter 2 of this section, "Methodological Issues Related to Data Collection and Analysis," presents criteria to be applied when deciding whether or not to undertake a particular data compilation or collection activity.

Chapter 2

Methodological Issues Related to Data Collection and Analysis

The availability and use of data in the health planning process will not necessarily enhance the appropriateness of decisions and actions. The value of data in such an objective is a function not merely of their presence, but of the extent to which they are collected (or assembled), analyzed and interpreted in a manner that is as consistent as possible with current standards of adequate research methodology.

Certain features of the context in which health planning takes place operate to complicate this seemingly obvious dictum. In the first place there is no single discipline of health services research and hence no structured body of theory and methodology that is de facto appropriate for use by the health planner.[1] Rather, the theoretical and methodological perspectives of a variety of academic disciplines are "borrowed" for such use. For the most part these borrowed theoretical/methodological perspectives derive from the social sciences and are therefore themselves the subject of continued research and development—they are not absolute even in the context of the "purest" of scholarly research and become even less so when transferred to a policy setting. This is partly because of the decision constraints that operate in a context such as the HSA. For example, the planner will typically not have available the level of expertise, facilities and resources necessary to perform the kind of detailed analysis of very narrow issues that are favored by academic researchers; he will not in all likelihood have the time to do so; and in most instances will have little use for the limited conclusions that can be derived from such research. In addition, the HSA will not have the luxury of stating the favorite caveat of the researcher ("more research is needed before definitive answers will be possible"). On the contrary, the planner will be forced to make decisions on the basis of such information as he has at any given time. This problem is further complicated by the legislative requirement that HSA's use extant data to the maximum extent possible.[2] As a result, many of the questions regarding techniques of data collection will not be subject to control by the HSA.

Given these issues no definitive set of methodological "rules" governing data analysis in the HSA is possible. The present chapter is therefore designed not as an introduction to social sciences research methodology but as a discussion of certain basic issues which can be expected to arise in the use of data for health planning. Persons who have experience or familiarity with social science research methodology may find these materials unduly basic. Those unfamiliar with such procedures will wish to read this chapter with care and refer to the sources noted for additional information on topics which remain unclear. Virtually all of the data collection and use described throughout this Handbook will be comprehensible only if the reader has a complete grasp of the principles presented here.

It should be noted that the single most important data resource of an HSA may well have little to do with methodological issues of collection and analysis. This is the detailed, first hand knowledge of local conditions and circumstances that each member of an HSA staff must develop over the course of the agency's activities. Although part of this information may be derived from data analysis (especially the completion of resource inventories and the examination of extant data) much of it will be gained in an experiential manner, through conversations with local providers and consumers, as well as early planning and management efforts. Since the structure and nature of the 200 plus HSAs will be diverse, this base of knowledge about idiosyncratic local conditions will be valuable in its own right, and will in many cases dictate the utility of certain methods of data collection discussed throughout this Handbook.

A. THE RESEARCH PROCESS I: PRELIMINARY PHASE

Before any particular research project or study can be undertaken certain preliminary issues must be dealt with. Generally these are 1) the statement of research objectives and 2) the assessment of study feasibility.

1. The Statement of Research Objectives

In the planning and management setting of an HSA, data are collected and analyzed for use as a resource in some present or anticipated decision-making process. As a result, the general research problem is easily identified. However, there needs to be a clear statement of the specific questions to be answered by the research. Whether the research problem is as narrow as determining the physician to population ratio in the HSA or as broad as describing the health services utilization patterns of residents in the area, specific research objectives must be clearly defined and stated before the research process itself begins. Vague suggestions (e.g., "Let's do an EMS study") will undoubtedly return to haunt an HSA which proceeds too quickly in the implementation of such a study. At best, that type of procedure will result in some inefficiency—the investment of resources in a study which is not relevant to the most pressing problems in the HSA, or the unnecessary collection of information which is superfluous to the most important dimensions of a problem. At worst, the result might be a failure to gather the type of information of critical relevance to the health problem under investigation.

The seemingly straightforward example of the physician to population ratio serves as a case in point. It requires, for example: 1) the specification of areal units—is the HSA interested in a single ratio for the entire area, or a series of ratios describing counties or other specified sub-regions which might be compared? 2) the specification of temporal units—is the HSA interested simply in current physician to population ratios, or in projecting population change in order to estimate the number of physicians which will be required to maintain or improve such ratios in the future? 3) the definition of the term "physician"—for example, are osteopaths, interns or physicians who provide no patient care (e.g., those involved in research or who are retired) to be included? It is only after questions such as these are answered that specific research objectives can be stated.

Two mechanisms are available to facilitate this process of objective specification. The first of these is an examination of related information. Although somewhat analogous to the academic researcher's "literature review" this procedure is more extensive, including reviews of extant data sources, related state and federal regulations, government reports, documents and publications as well as the relevant research literature. At the same time, staff should precisely outline at least the essential uses of the study data, as these are presently perceived by the staff. By virtue of these two procedures the research objectives can be more easily specified and they can be phrased in a manner consistent with related research; thus facilitating comparisons therewith. In addition, previous methods of data collection and analysis can be examined, useful techniques and procedures can be borrowed, and errors avoided.

2. Feasibility Assessment

Having determined and set forth study objectives, the agency will, of necessity, assess the feasibility of completing the research. In this context, feasibility must be viewed as a relative term, incorporating more than the narrow issue of whether or not a study to provide answers for some specific set of questions is possible. The relative difficulty of obtaining the necessary information, available resources, the precision with which the data will address the research questions, and related issues regarding the HSA's information needs are also relevant.

One portion of the feasibility question is the issue of the data source. At the most general level, this is the question of whether study objectives allow the use of existing data or require primary data collection. P.L. 93-641 clearly specifies that "the agency shall, to the maximum extent practicable use existing data...,"[3] and it is therefore assumed that primary data collection will not typically be undertaken. This dictum implies, however, that there is some absolute and clear division between extant data and those which must be primarily collected. In fact this is not the case. It is more accurate to view the distinction as a continuum ranging from data which exist in some immediately useful form to data which must be gathered entirely "from scratch". Between these extremes are data which exist but will be useful to the HSA only after some degree of manipulation. Operationally, extant data can be considered those which are available at reasonable cost, either in published form or in a manner that allows abstraction by a single HSA staff person working at one location. In spite of this definition, it is acknowledged that some issues as to the relative availability will remain.

For example, such extant data as are available were undoubtedly collected for purposes not precisely equivalent to those specified by the HSA. They will, therefore, be more or less suitable in meeting the latter's objectives, and the degree of suitability will have to be evaluated. For instance, there may be more than one potential data source, and the advantages and disadvantages of each will have to be weighed. Perhaps the study will require the simultaneous use of data from several sources, in which case the consistency of definitions, coding schemes, areal and temporal coverage, etc., will have to be carefully reviewed. The available data may be in a form that will require the allocation of considerable resources to necessary pre-analysis manipulation. In view of these issues, the existing vs. primary data question will typically require an assessment rather than any clear cut decision.

This assessment will in part evolve directly from the statement of objectives—for some research questions, the precise data requirements might well be met by the direct use of existing data, while for some purposes there will be no previously collected information of even peripheral relevance. For the most part, a situation between these extremes will obtain. Although it is impossible to define an absolute set of criteria for use in assessing the relative value of an existing data source the following issues are clearly relevant:

- Data availability: The fact that data "exist" does not make them ready for immediate utilization. The amount of manipulation required may be as little as collapsing categories or as much as the complete abstraction of a sub-set of data cases. In the latter type of situation the value of the data and the importance of the research questions must be carefully assessed.

- Data completeness: The data may or may not cover the precise population and areal boundaries of interest to the HSA; data cases may be more or less complete; data items of interest to the HSA may not be included. The planner should carefully examine whether or not gaps in the data will be significantly detrimental to their utility.

- Timeliness: Existing data are often made widely available at a point in time when their immediate relevance is questionable. When all other things are equal, more recent data sets are preferable. In some cases (e.g., Census data) even the most recent information may be significantly out of date.

- Comparability: Extant data have been gathered by diverse organizations for a variety of reasons. As a result, the comparability of various data sets ranges from substantial to nil. To the extent that projected data uses include comparisons, or the synthesis of several data sets, the comparability issue is critical.

In cases where an examination of such issues leads the agency to the conclusion that no extant data will serve the stated purposes of the study, three options are available: 1) the study can be abandoned; 2) the objectives can be restated in such a way that extant data are useful; or 3) the agency can proceed to collect the required data. When one of the latter two options is selected, certain other dimensions of the feasibility issue remain. These refer primarily to the resources of the agency in question and include:

- Staff resources: Whether or not the agency staff includes persons with the training and expertise to conduct the study of interest is of primary consideration. Without such personnel, the agency will be required to enlist the services of outside consultants or arrange to contract the entire study to a research organization.

- Facilities: Depending upon the nature of the study, various specialized facilities may be required. The most obvious example is a study which uses large quantities of numerical data and hence requires extensive use of computers.

- Cost: The HSA will also have to estimate the cost of study execution in order to determine whether or not its results merit the required allocation of resources. Obviously, the cost will in part be determined by the feasibility questions previously discussed—if in-house staff have the required expertise the cost of sub-contracted consultants will be avoided; the availability of required facilities will prevent the HSA from having to contract for such things as keypunching service. However, even studies that can be completed by in-house staff will often require substantial man hour investment. For example, a study of Emergency Medical Services such as is described in Chapter 5 of Section IV may well require 6-8 months and the full-time commitment of two or more professional staff members as well as substantial clerical support. Very straightforward calculations of relevant areal figures from easily obtainable extant data (such as the physician to population ratios mentioned earlier) will also have certain costs—obtaining the required data may necessitate a library search, the calculations themselves will consume time, etc.

At the conclusion of the preliminary phase of the research process, then, the HSA will have set forth study objectives, selected a data source or sources, and completed a detailed projection of data uses. In addition, the cost and relative feasibility of the study will have been estimated and its benefits weighed. A decision as to whether or not study implementation is warranted will be made. It should be acknowledged that for studies which involve fairly clear use of extant data, this entire preliminary phase may take but a few hours; the costs involved will probably be minimal and the study will be completed as a matter of course. In more complex studies a more careful step by step movement through this procedure will be required. The following discussion refers to the latter situation—an extensive study whose results have been deemed to have a value which outweighs the costs of obtaining them.

B. THE RESEARCH PROCESS II: STUDY IMPLEMENTATION

1. The Study Plan

The first step in actual implementation is the development of a study plan. This should outline the expected research procedure, beginning with the statement of objectives and concluding with intended data analysis and use. It will also include those of the following sections which are appropriate:

- definitions of key concepts, terms and variables
- the consistency issue
- the definition of the universe and selection of analytic units
- methods for enhancing cooperation
- instrument design
- data collection
- data processing and storage

Each of these topics will be discussed in the following pages in somewhat more detail than they would appear in a study plan. However, it is emphasized that each issue should be reviewed in advance to insure the continuity of a study. Few experiences are more frustrating than the discovery, late in the research process, of a flaw which severely limits the utility of a study and cannot, at such time, be corrected. While a study plan cannot be expected to prevent all problems it can help avoid clear difficulties such as the failure to adequately define key variables or coding data in a way which precludes the completion of required analyses.

Part of the study plan should include attempts to enlist the services of persons or organizations outside of the HSA who might be able to make valuable contributions to the study. These might range from enlisting the aid of local service clubs to assist in the leg work of some studies, to obtaining complimentary data processing from local firms, hospitals or universities. Obviously, the potential value of such assistance is immense, but its availability will depend upon the nature of the study. Local universities are a particularly valuable resource of this type; faculty and advanced graduate students may serve as consultants, often simply in exchange for an opportunity to utilize data in research endeavors. Undergraduate students are often available for such tasks as interviewing, coding, library research, etc. Such persons are often relatively sophisticated in that type of work yet can be hired without the expenditure of large amounts of funds.

2. Definitions

The explicit definition of key concepts, terms and variables is, of course, a standard requirement of the research enterprise.[4] It is accentuated in the HSA context by the fact that much of the expected data analysis will be comparative, involving an assessment of various situations in the HSA as they compare to similar phenomena in some referent population. As a result, concepts, terms and variables must be defined in a manner that is not only useful within the context of the particular study, but also consistent with analogous measures for referent groups. For example, one measure of health services utilization that has been frequently used is the number of physician visits per unit population per unit time. Whether physician visits include telephone consultations, mass service provision (e.g., school vaccinations or TB skin tests), inpatient contacts, etc., and the meaning of the term "physician", are all questions of definition and must be specified. Without such specification a comparison of physician utilization in an HSA with that in the nation may have little meaning.

Since much of the data to be used will be derived from extant sources, the definition issue will in these instances be less subject to HSA control—the definitions employed in the original data collection process will have to be used except in cases where data are regrouped (e.g., into different age categories). Whenever more than one data source is used, definitions must be consistent or acknowledged to be otherwise.* In any event, all fundamental terms must be provided with explicit operational definitions.

3. The Consistency Issue

As indicated above, we assume a predominant role of comparative analysis. This demands that the various data used in any analysis be consistent. It should be noted that the term consistency here refers to issues beyond terminology and includes the following:

- temporal consistency
- analytic unit consistency
- measurement consistency
- data linkage

Temporal consistency is of clear significance in the analysis of data referring to an entity which is as dynamic as the health services delivery system. For example, the physician to population ratio was noted earlier as a clear and useful measure of service availability. It requires that the numerator (number of physicians) and denominator (population) refer to a similar point in time, especially when either number is known to be undergoing change. In addition, since the ratio will be most useful when combined with a referent ratio, both ratios must refer to a comparable time period.

*This problem will occur with some frequency in the HSA context where multiple sources of extant data will be utilized. In some cases it simply cannot be helped and comparisons will either be impossible or so heavily qualified that they have little meaning. However, definitions may have a degree of similarity that allows reasonable levels of comparison, for example,in the definitions of "short stay general hospital" adopted by NCHS and the AHA (see discussion of this problem in Section IV, Chapter 12, describing "A Study of the Utilization of Hospital Inpatient Services".

Analytic units must also be consistent across the data sources of interest. If the unit of analysis in one data source is the hospital patient, comparison with studies of hospital discharges will be of questionable accuracy. The same issue applies to studies which use different levels of aggregation (see Section II, Chapter 4, of this Handbook which discusses the ecological fallacy).

Finally, attempts to integrate several data sets (for example, where studies are to be repeated in order to obtain time series data) are dependent upon the presence of consistency in units of analysis.

Measurement techniques have long been acknowledged as a significant factor in data comparability. The wording of questions, the response categories provided, whether the data are collected by mailed questionnaires or face to face interviews, etc., are all measurement issues which will have an impact upon ensuing analysis. For all extant data such problems will not be subject to control by the HSA. As a result, where inconsistencies are apparent, appropriate caution in the interpretation and use of the data must be exercised.

The question of data linkage will be of particular importance in HSAs where, over a period of time, a number of special studies requiring the collection of primary data are implemented. In such instances, the potential utility of any given data set is dramatically increased if it is also amenable to linkage with other data sets. The complete linkage of relevant data files represents the elusive "Health Services Data System" (HSDS).

The "elusive" nature of such a system is in part a reflection of the lack of any integrated body of theory in the area of health services.[5] In that sense, HSDS will continue to elude planners as well as researchers for some time to come. On the other hand, certain mechanisms are available to enhance data linkage from a very practical rather than theoretical perspective. Many of these have little or no real cost associated with them and should therefore be utilized. Primarily, the mechanisms include the use of coding systems which insure that location variables (county, address, zip code, etc.) and identification codes (e.g., for hospitals, physicians or other units) are kept constant in form and repeated in all applicable data files. Such procedures also enhance the potential for utilizing data from a single source in more than one study. For example, the two studies discussed in Chapters 11 and 14 of Section IV require data which would be collected in hospital emergency rooms. Careful planning may well allow a single collection of data to be used for both purposes. Similarly, it may be desirable to merge two data files and thereby expand a data base. This will be possible only if appropriate attention has been paid to coding procedures, etc. Throughout the various studies described in this Handbook, some attention has been paid to this issue. Nevertheless, careful planning in the specific contexts of various HSAs will enhance these procedures.

4. Defining the Universe and Selecting Units of Analysis

The term "universe" is used to denote the total of all units to which a study refers, e.g., all short-stay hospitals in Central HSA, all physicians in Central HSA, etc. When the analytic units are persons, the term "population" is often substituted to describe the universe. In all social research situations a clear definition of the universe is required. This insures that there is no misunderstanding regarding the generalization of findings and aids in any sampling decisions.

In the area of health services research, the definition of the universe is particularly important. This is because practical exigencies often make it necessary to examine a universe which is not equivalent to that which might be intuitively assumed from a single term. For example, HSAs will undoubtedly gather and analyze data referring to the hospitals in the area. However, the actual universe under consideration may be:

- all hospitals located in the HSA;

- all hospitals serving residents of the HSA;
- all general hospitals located in or serving residents of the HSA,

or any one of a variety of specific hospital groups. Similarly, an EMS study may refer to providers of emergency medical care whose center of operations is located in the HSA or to those (regardless of location) who provide service in the HSA. The selection of the universe of interest is a theoretical/substantive question which is affected by the study objectives, and the explicit definition thereof is required regardless of the reasons for its selection.

Having defined the universe of interest, the selection of analytic units (cases) follows logically. The first issue is whether all members of the universe or some sample thereof should be studied. In some situations the universe cannot be examined in its entirety because of accessibility, limited time, or prohibitive cost. Alternatively, the universe may contain a number of cases too small to permit any sampling—some HSAs may contain very few nursing homes, for example, and a sample of them would not be useful. This decision will often not be unequivocal—clear advantages may accrue from either alternative and each must be weighed in cost/benefit terms.

As a general rule, sampling is a method of enhancing efficiency by examining some portion of the universe in order to draw conclusions about its entirety. The degree to which the researcher has confidence that the sample data are in fact indicative of phenomena in the universe is a function of the representativeness of the sample. Since it is often not possible to empirically check representativeness, the researcher must generally rely upon confidence in the nature of the sampling procedure.[6]

The degree of confidence in a sample can b measured only with probability samples, i.e., those selected in a manner that gives each element in the universe a known, non-zero probability of becoming an element in the sample. Such non-probability samples as fortuitous samples (e.g., man-on-the street interviews), expert choice samples (e.g., where an element is selected as "typical" based on expert knowledge), etc., <u>may</u> be representative but it is impossible to measure the degree to which the researcher is confident of that representativeness. In some instances this uncertainty must simply be acknowledged and "lived with."

All probability samples are, in one way or another, based on the concept of simple random sampling (SRS).[7] However, relatively few practical applications of SRS are likely to occur. More frequently, such techniques as cluster sampling, stratified sampling, multistage sampling or some combination of these and/or other techniques will in fact be utilized.

- <u>Cluster sampling</u> simply means that a cluster or group of the elements to be sampled is selected first. For example, if one wishes to obtain a sample of the persons in a city it will probably be impossible to list all of the elements (people) and use a simple random selection procedure. In such cases a sample of clusters, for example, city blocks or census tracts, might be selected and the sampling units derived therefrom.

- <u>Stratified sampling</u> is similar to cluster sampling except that the clusters are selected according to some purposive procedure. For example, suppose an HSA is interested in selecting a sample of discharge records from hospitals in its area. Time and cost factors prohibit the selection of a sample of records from every hospital so a random sample of hospitals is to be drawn. However, the HSA has theoretical or substantive reasons to believe that the information of interest on the discharge records may vary depending upon the size of the hospital. In such a case, the sampling procedure might first involve the construction of three lists of hospitals (small, medium and large) and the random selection of specific hospitals from each of these strata. This would insure the inclusion of discharge records from each type of hospital.

- <u>Multistage sampling</u> simply describes a process whereby elements are selected from sampling units (city blocks, or hospitals etc.) in two or more stages.

The basic decision of whether to sample or examine the entire universe will be made on the basis of very practical (typically time and money) issues. The more specific decision about the sampling technique will be influenced by these as well as other factors. According to Kish, a good sample design "requires the judicious balancing of four broad criteria:"[8]

- Goal orientation: "The entire design...should be oriented to the research objectives, tailored to the survey design and fitted to the survey conditions....Although these admonitions are vague, they are not trivial. For example, too often a neat sample of a single unit is used, such as a town, or a tract, or a class of students, when the research objectives would be better met with a less neat sample of a larger population."[9]

- Measurability: The design should "allow the computation, from the sample itself, of valid estimates or approximations of its sampling variability."[10]

- Practicality: The design must be such that it can be accomplished as designed. It must be susceptible to description in "simple, clear, practical and complete" terms to the field personnel who are actually implementing the design.

- Economy: The design should reflect the minimum cost required to satisfy the objectives. A study directed at an issue where substantial non-sampling problems exist (e.g., measurement problems) is only marginally enhanced by an exceptionally good sample. An excellent example of this last criteria is the issue of sample size.

  > A sample is too small if its results are not precise enough to make appreciable contributions to decisions. For example, it would be generally useless to forecast that one of two parties in an election will win from 40 to 60 percent of the votes. On the other hand, a sample is too large if its results are more precise than is warranted by their likely uses, or if the non-sampling errors overwhelm the sampling precision.[11]

This brief discussion makes it clear that the myriad of decisions to be made in the process of sampling are, while governed by general rules, affected by the specific, idiosyncratic situation of the study. All of the possible issues cannot, therefore, be discussed here, but study staff should be aware that sampling decisions will fundamentally affect the utility of any given data file and should be made with appropriate caution. One excellent source of information regarding general issues in sampling is:

Leslie Kish, Survey Sampling, New York: Wiley, 1965

A valuable source of information more specifically related to some of the problems likely to be faced by HSAs is:

Irene Hess, Donald C. Riedel and Thomas B. Fitzpatrick, Probability Sampling of Hospitals and Patients, Ann Arbor: University of Michigan, 1961.

Each of these works contains a substantial list of references which may be useful to HSAs depending upon the specific sampling issues of concern. Each also contains a clear discussion of the actual mechanics of element selection.

5. Methods of Enhancing Cooperation

One issue that is frequently given too little attention in the design of a study is the question of gaining cooperation from those persons who control access to the data to be used in projected analyses. Too often, an excellent sampling design is rendered useless when data are obtained from only a portion of the sample elements. While it is usually impossible to force cooperation from all elements (sometimes it is impossible to even find some prospective respondents), a high level of cooperation, and hence a high response rate, is desirable. For an HSA whose study objects (especially the various providers of health care) already perceive themselves to be overstudied and unnecessarily burdened with forms, questionnaires and reporting requirements, this problem is particularly acute.

The HSA, of course, has a legislative mandate to collect and analyze data, and in the case of such things as project review decisions, the provision of necessary data will be required. However, it will presumably be in the best interests of both the HSA and the various actors in the health services delivery system to exchange information in a manner characterized by cooperation rather than legislative demands. Since it is the HSA which will in most cases initiate studies, the primary responsibility for assuring such an atmosphere of cooperation must rest with them.

In each of the special studies described in the Handbook, considerable attention is paid to potential mechanisms of enhancing cooperation. These should be taken as examples and utilized in various appropriate forms depending upon the specific study to be undertaken. They are characterized by three fundamental principles. First, cooperation should not be assumed—it should be sought. Whether by providing news releases and "dear neighbor" letters to prospective respondents in a household survey, by personal or telephone contacts with hospital administrators, the solicitation of endorsement from professional organizations or whatever, the study staff must pursue the objective of gaining the necessary cooperation.

Even when an atmosphere of cooperation is established, the HSA should continuously direct efforts at enhancing it. This will include the second and third general principles: confidentiality and data sharing. All research subjects have privacy rights which must be maintained and which are of particular relevance in the area of health-related information. While to be useful, data must often contain detailed address and identification information these must be guarded and individual anonymity protected. Assurances must be provided and respected. The HSA which can point to a record of judicious use of data has a powerful argument to allay the (often justifiable) fears of potential respondents. Conversely, an agency which fails to treat information in an appropriate way will soon find that individuals and organizations are not likely to be cooperative.

The value of data sharing will depend upon the individual study. In some cases respondents will have no interest in study findings. However, institutions and organizations will often have more than a passing interest in results, and to the extent that confidentiality is not compromised, should be provided with copies of reports, etc. In many cases this is the only reward that the HSA can offer in return for cooperation. Promises of such sharing are often used to gain cooperation and should be fulfilled. Again, a record of data sharing will be valuable and a record of broken promises to share results will have a dramatic negative effect on future cooperation.

6. Instrument Design

The design of an instrument to be used in data collection will be an ongoing process that does not start at any specified time. For example, it is often advantageous to have an instrument at least tentatively laid out before beginning the necessary contacts in the process of gaining cooperation. On the other hand, during that process prospective respondents may react to any instrument and the HSA will gain information about its length, the difficulty of obtaining requested data, etc. In this regard even early and informal contacts represent a quasi-pretest of a tentative instrument, and there should not, at that point, be any profound commitment to its specific form or content.

Instrument design will be substantially facilitated by a knowledge of previous related work. It is well known that the wording and placement of questions, the response categories provided, etc., will have an impact on the data collected. If previously tested questions, categories, etc., can be used, instrument development will be easier and the product of that process more sound. While the difficulties in instrumentation will vary depending upon the particular study, certain fundamental issues bear mention.

- *Length*: The instrument should be kept as short as is consistent with the study objectives. Regardless of whether it is an interview schedule, an abstracting form, or a questionnaire, its length will influence the cost of completion, and, in the case of questionnaires especially, an inordinately long instrument will result in higher frequencies of refusals and incomplete forms. This should not, of course, be carried to the extreme of excluding questions or items of value—the completion of an abstracting form which requires five minutes may be less expensive in the long run than using a shorter instrument only to find that another data collection process is required at some later time.

- *Efficiency*: Regardless of its length, the instrument should be framed in a way that enhances the efficiency of its completion. This includes the blocking and logical ordering of related items, the provision of adequate response space, etc.

- *Complexity*: Related to efficiency and length is the question of instrument complexity. However carefully one trains the data collection personnel, interviewers, abstractors, etc., there will likely be some (hopefully minimal) level of error in the data collection process. This is especially the case when the instrument is characterized by complex skip patterns (i.e., systems of determining which question should be asked next depending upon a previous answer: "if yes, go to question 12") different patterns of sub-questions, etc. The instrument should be as simple and clear as is consistent with the study objectives.

- *Instructions*: Even the simplest, post-card length instrument will contain terms and questions that are open to various interpretations. In situations where respondents provide the data, careful instructions, including definitions and examples should be provided. Where interviewers or abstractors are employed, they must be trained in the use of the instrument. In either event, study staff will find it necessary to develop a detailed set of explanatory information regarding the instrument. This need will most efficiently be met if such materials are developed concurrently with the instrument itself.

Each of these issues in the instrumentation process will be best handled by way of pretesting. The pretest is a try-out of the instrument to see how it works under conditions similar to those which will obtain in the actual study. Even when an instrument closely resembles one which has been previously used with some success a pretest will provide useful information as to its applicability in the particular context of its intended use. When the instrument has been developed specifically for a study, the pretest will indicate which questions are difficult, unclear or resisted; whether response categories are adequate; how long it takes to complete the instrument; etc. If substantial changes are necessitated by the results of the pretest, the entire process should be repeated. A brief but excellent discussion of questionnaire development is contained in:

> Claire Selltiz, _et al._, _Research Methods in Social Relations_, New York: Holt, Rinehart and Winston, 1967 (Appendix C)

Although it deals with the questionnaire as a specific type of instrument the discussion is relevant to other forms as well.

Following the pretest and the resulting derivation of a "final" version, the instrument may be pre-coded if the resulting data are to be stored and analyzed with the aid of computers. Substantial time and money savings can be achieved if the instrument can be designed in a manner that includes the coding categories and allows direct keypunching. For relatively complex instruments, the coding scheme should also be pretested, perhaps by having a small number of instruments completed with fictional data and keypunched.

7. Data Collection

The mechanics of the actual data collection process are also subject to valuable preliminary planning. Again, while there will be immense variation from one type of study to the next, a set of generic issues should be considered. These are most critical when data collection involves a staff of interviewers or abstractors, but are also relevant in the case of mailed questionnaires or the transfer of extant data from its original form to that needed for any particular study.

The persons who will actually perform the data collection must be carefully trained for their task. This may vary from intense training sessions covering several days for prospective household survey interviewers, to the simple provision of clear instructions to a research assistant who is charged with extracting some figures from a census report and constructing a table. Also included here are instructions for the completion of questionnaires or forms where the respondents themselves are the data collection personnel. In all of these cases the same principle holds: the persons who gather the data must have a clear and complete understanding of what they are doing and why. This will obviously be enhanced if, for example, experienced interviewers can be hired, or medical records technicians are used to abstract data from hospital records.

In addition to advance training, study staff should institute specific quality control techniques which will be imposed during the process of data collection. In the case of interviews these will include spot telephone checks to insure that interviews took place, editing of completed instruments to check for accuracy and consistency and the inclusion of an interviewer identification code on each instrument. In the case of abstracting techniques, the data collection personnel should be closely supervised and spot checked. Particular attention should be paid to the provision of clear decision rules for the coding of non-categorical data, and post facto editing for logical consistency is valuable. An excellent discussion of these procedures as they apply to interview surveys is available in:

> Eva Weinberg, Community Surveys with Local Talent, Chicago: University of Chicago, 1971

8. Data Processing and Storage

Following completion of the data collection, there may be a necessity for further processing prior to analysis and use. Where data are of an inventory nature (see for example, A Guide to the Development of

Health Resource Inventories)[12] they may be collected, stored and analyzed using a single file of index cards. For studies which require tabulations of large quantities of data they will undoubtedly be keypunched and stored on some machine-readable medium. In either situation, careful editing of each data case to insure its internal consistency and maximum possible completeness is required. There are a variety of computer programs available to assist in this task for large data files but in the case of inventory data it will simply require a careful check by the relevant staff member.

If the data are to be analyzed with computers it is at this stage that a codebook must be completed. Where the instrument was precoded, much of the structure of the codebook will be determined by that fact, although additional information may be included. Codebooks have been included for data files that would emerge from the special studies described in this Handbook. These provide examples of conventions in coding that will enhance data analysis. However, as was implied in our discussion of data linkage, the specific features of the research plans and interests of the HSA must govern coding decisions.

C. THE RESEARCH PROCESS III: DATA ANALYSIS AND USE

As has been implied or noted at several points throughout this Handbook, data analysis and use are fundamentally related. Those portions of existing or primarily collected data which will be analyzed are almost by definition those for which the agency has some clear use. Data files are not likely to be the focus of "data dredging," hypothesis testing, or narrow analytic studies of specific relationships. As a result, many of the sophisticated and (in some cases) very powerful data analysis techniques now available to social scientists are simply not relevant to the HSA context. Fundamentally, the health planner will be concerned with: 1) descriptive statistics (as opposed to the various modes of relational and/or explanatory analysis) and, given our emphasis on a comparative perspective; 2) the necessary techniques for examining differences (among or between population subgroups, geographic areas, time periods or whatever).

1. Data Display

An efficient, parsimonious yet meaningful display of data is the first step in their analysis. For many of the types of information which will be used in the HSA, the basic mode of data display is the statistical table. For studies based on extant data, the information will often be derived from published tables, and regardless of their source, data will frequently be presented in tabular form. This implies that relevant HSA staff will have mastered the techniques of reading and interpreting tables as well as those of table construction. Mueller, et al., succinctly express the issues in table construction, and the same principles should be borne in mind for other modes of data display (graphs, charts, maps, etc.):

> It is clear that there is more than mere "shape" to a table, and that its construction requires more than routine skill in layout and spatial arrangement. A table is much more than a mere listing; it is an organization of the data. A well built table constitutes the pith and meaning of a study in its definitive form. Consequently, serious consideration of its form and context should not be postponed until the final stages of the study....To assure advance consideration of tabular requirements, the preparation of dummy tables is recommended.[13]

Most basic social science research methodology text books include discussions of data tables. In addition, a particularly clear statement is available in:

Hans Zeisel, Say It with Figures, 5th (revised) Edition, New York: Harper and Row, 1968

It is expected that in addition to tables, the following modes of data display will be useful in a context such as an HSA.

- Histogram: A set of contiguous columns whose heights are proportional to class frequencies and whose widths are proportional to the class intervals of a quantitative variable of interest.

- Frequency Polygon: Whereas the histogram contains a relatively small number of columns referring to broad intervals, the frequency polygon is a line joining the mid points of a series of very narrow intervals. If those intervals are narrow enough the frequency polygon will be a smooth curve reflecting a distribution much more accurately than the histogram.

- Bar Chart: This is similar in construction to the histogram except that the bars refer to classes which are nominal or categorical (e.g., sex, race, etc.). As a result, the bars are not contiguous but are usually separated by dead space. Since the variable is nominal, the width of the bars is arbitrary but their length remains proportional to the frequency of the attribute.

- Pie Chart: This is a familiar tool, often used to describe the relative allocation of funds. It consists of a circular figure divided into wedges whose size is proportional to the attribute frequency.

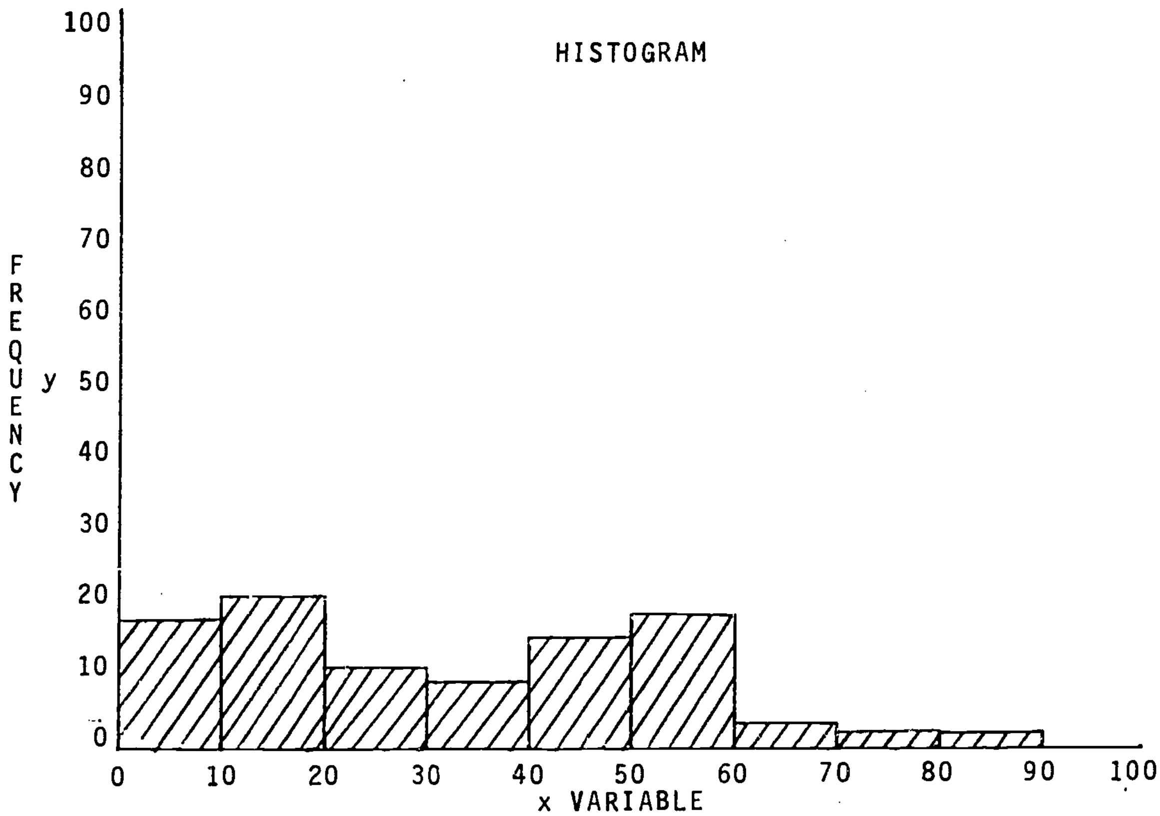
- <u>Statistical Maps</u>: These symbolize varying attribute frequencies in specified geographic areas. Typically, they utilize shades to convey various frequencies. They are particularly useful in demonstrating geographic variation in the occurrence rates of various phenomena.

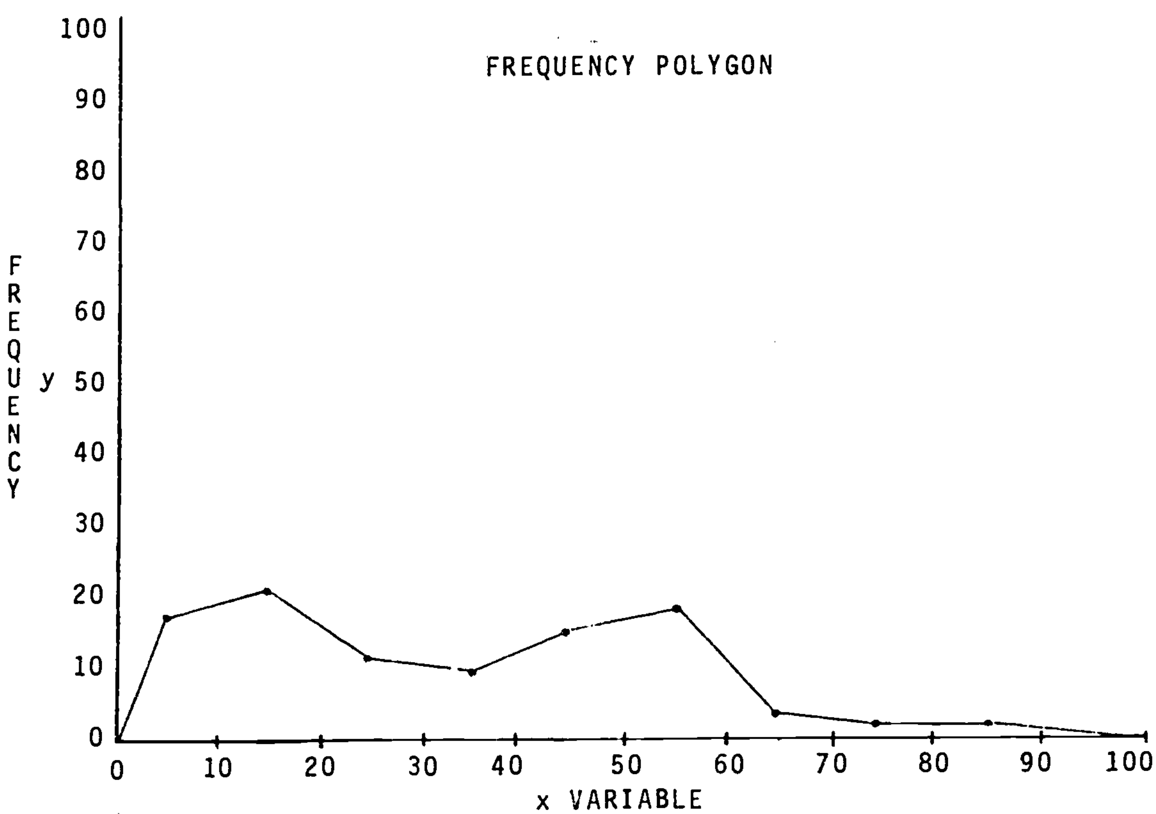
- <u>Time Graphs</u>: These may use either bars or a single line, with the latter providing more detailed information. The frequency of the event is represented on a vertical axis and the time periods of concern are displayed along the horizontal. The device has notable value in mapping trends, or changes over time—frequent objectives of the HSA.

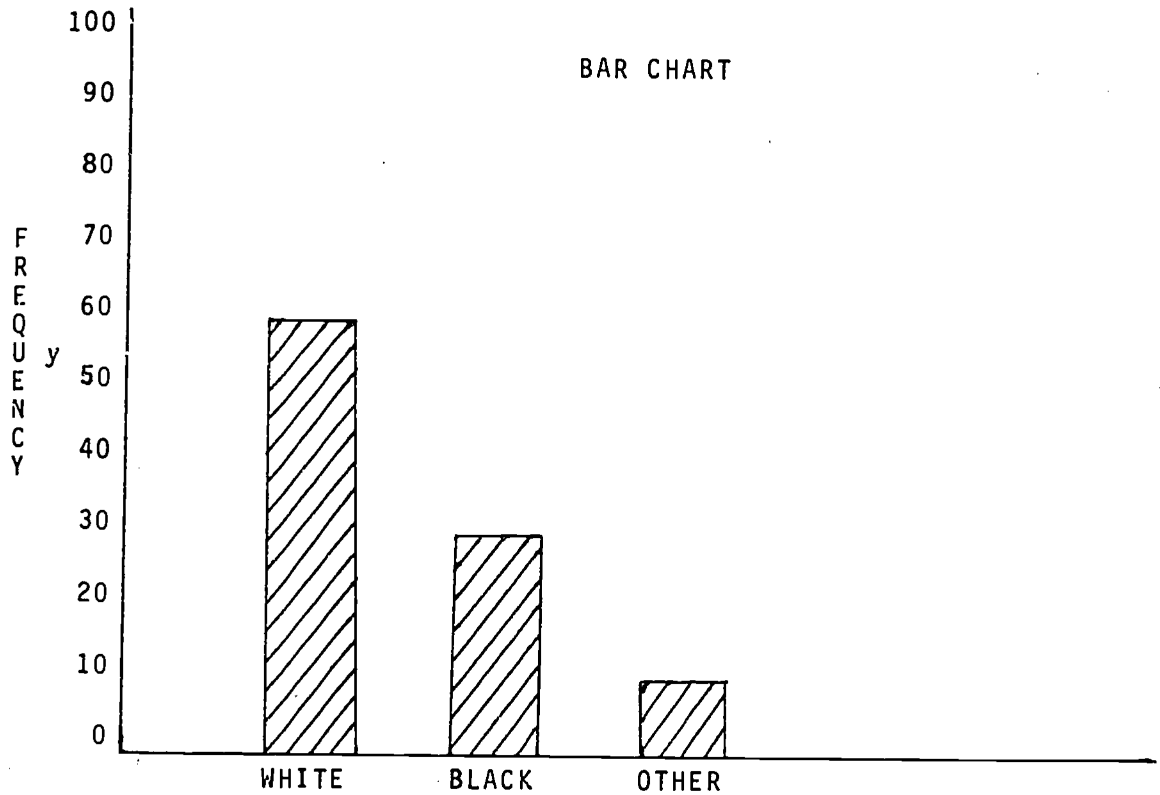
Examples of these modes of data display follow. It should be noted that two of these devices (bar charts and time graphs) are amenable to the simultaneous plotting of two or more frequency distributions. In the former case, this is achieved by pairs or groups of bars shaded to identify the groups of interest. For time graphs, multiple lines are employed. In the HSA, simultaneous plotting of frequency distributions in the HSA and some referent population will often be employed as a means of assessing where the HSA stands regarding various health related phenomena.

2. Data Summarization

In addition to the display of information, data can be <u>described</u> in various summary statements which are frequently more practical than a depiction of the complete distribution. There are a number of such summary descriptions and the appropriateness of each is determined by the level of measurement utilized. Measurement implies the concept of scales, which can be nominal, ordinal, interval or ratio in nature. Nominal scales represent simple classifications (e.g., male/female; protestant, catholic, other). There are no assumptions about relationships between such categories—none is "higher" or "better." If the categories are exhaustive and mutually exclusive, the minimum conditions necessary for the application of certain statistical procedures have been met. Ordinal scales are those which allow an ordering of categories (e.g., high, medium, low) but no inferences about the magnitude of differences between elements. While data of an


HISTOGRAM
FREQUENCY
y
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
x VARIABLE


FREQUENCY POLYGON
FREQUENCY
y
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
x VARIABLE


BAR CHART
100
90
80
70
60
50
40
30
20
10
0
FREQUENCY
y
WHITE
BLACK
OTHER

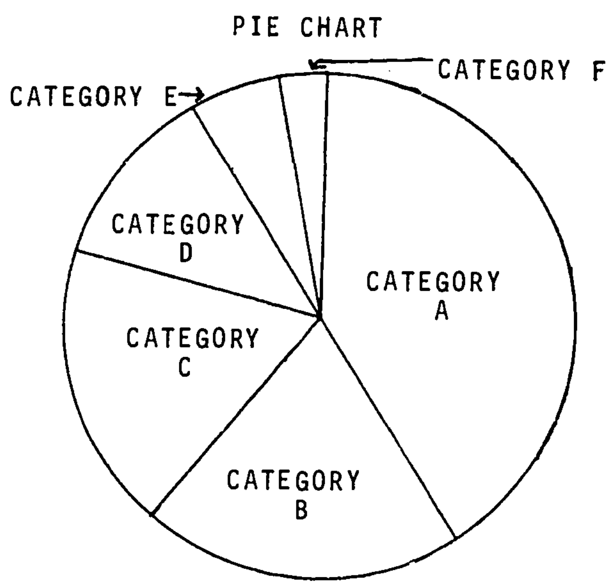
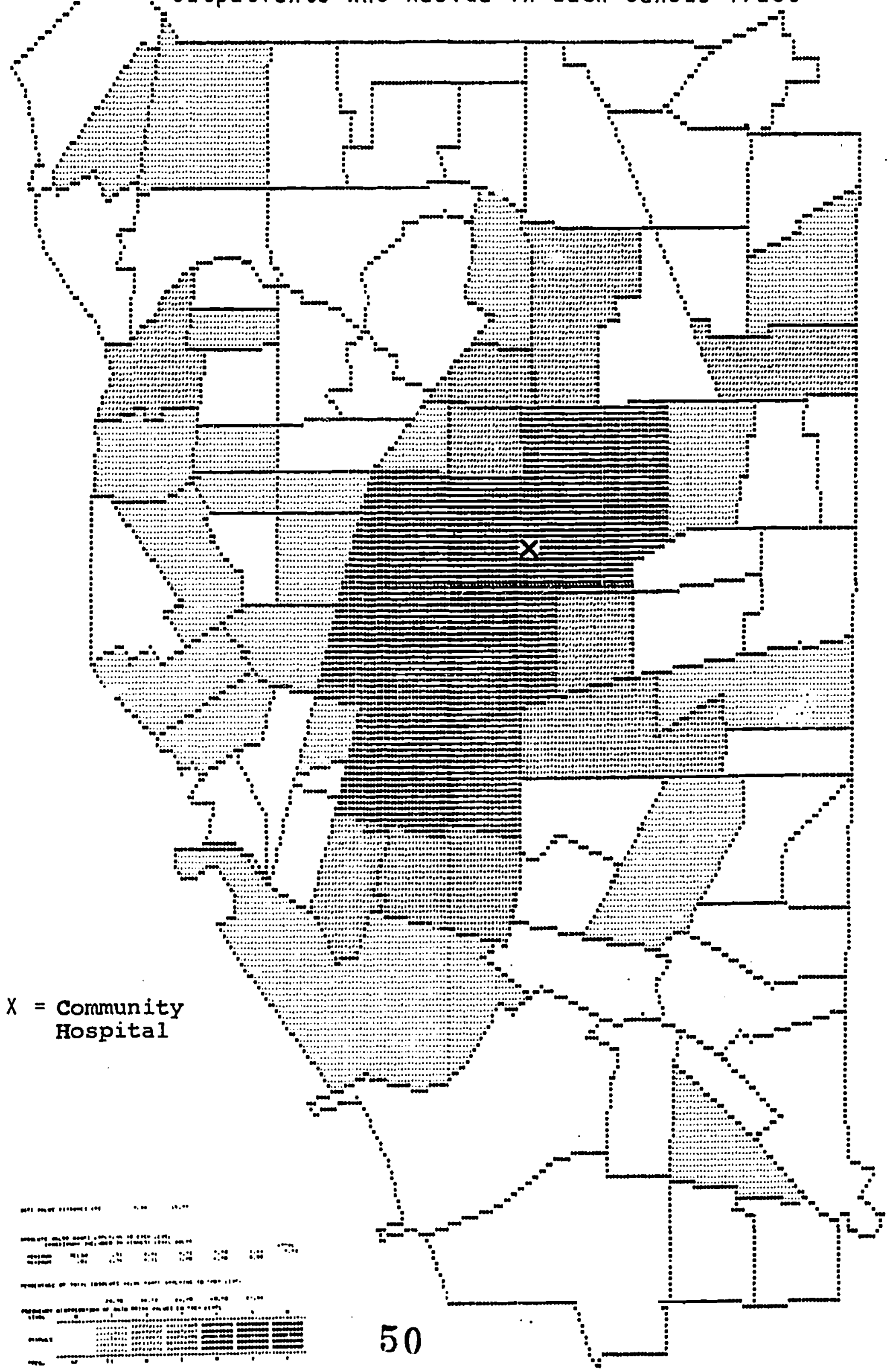
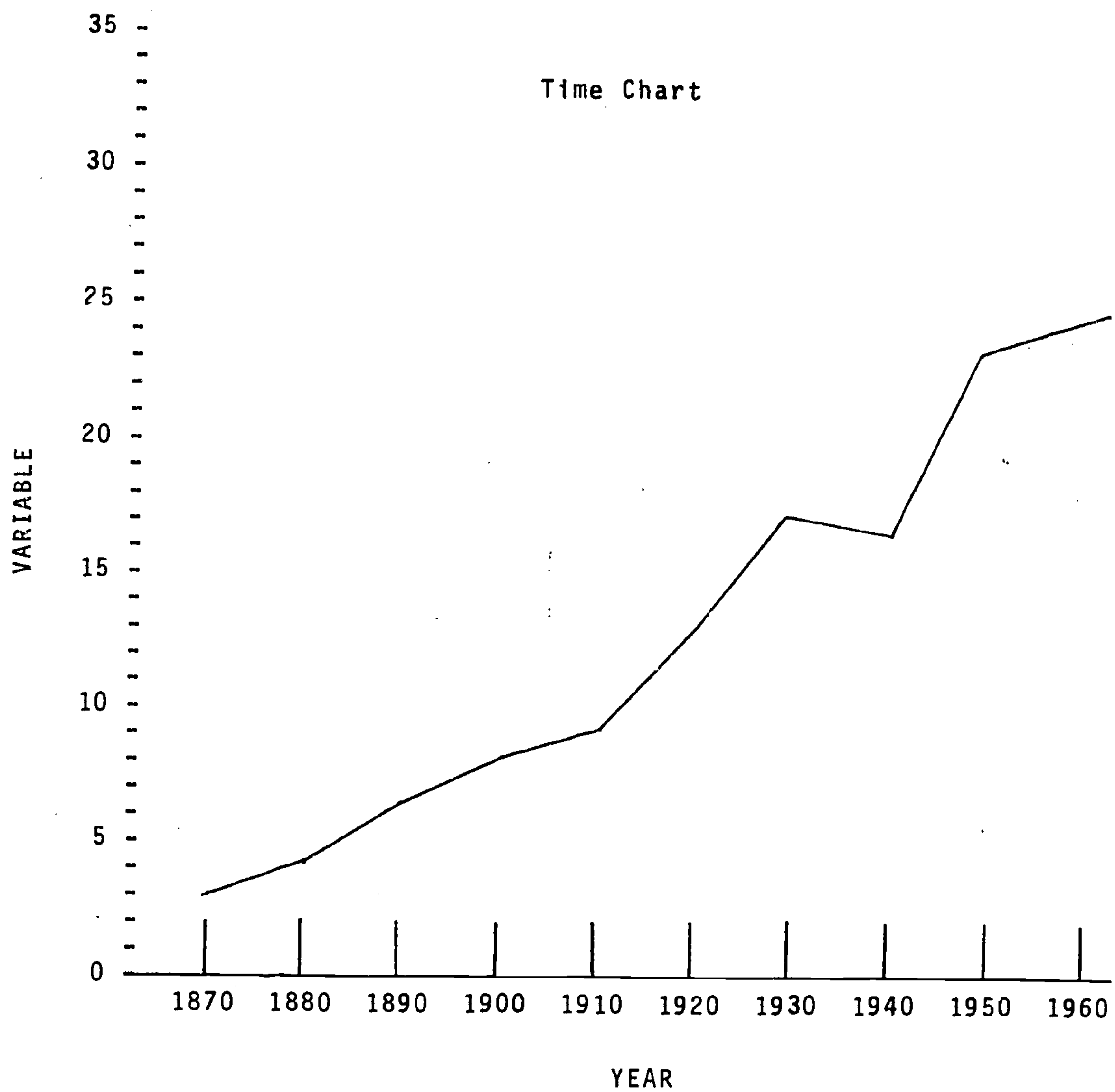

PIE CHART
CATEGORY F
CATEGORY E
CATEGORY D
CATEGORY A
CATEGORY C
CATEGORY B

Figure 3. Proportion of Community Hospital's Outpatients Who Reside in Each Census Tract




35
30
25
20
15
10
5
0
Time Chart
VARIABLE
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
YEAR

ordinal nature are amenable to "greater than" or "less than" operations, they are not subject to the usual operations of addition, subtraction, etc. Hence, for practical purposes, data of this type are often examined in a manner similar to nominal data. Interval scales are those which allow the ranking of objects in terms of the degree to which they possess certain characteristics as well as the specification of the exact distances between them. This implies a unit of measurement (e.g., inches, years, degrees) and allows the use of diverse mathematical and statistical techniques.

For data of a nominal nature, such as would be displayed in bar charts, pie charts, etc., the descriptive statistics of interest are proportions, percentages, ratios and rates. Proportions describe the relative size of any given category by expressing that frequency as the proportion of the total number of cases, hence standardizing for size. For example, suppose it is known that in two counties 5,000 and 15,000 persons, respectively, saw a physician during some specified time period. These numbers would provide no useful information on the relative number of each county's residents who had physician contacts. By expressing the figures as a proportion of each county's total population such a comparison could be achieved. If the counties had populations of 10,000 and 60,000, respectively, it would be readily apparent that the proportions of each county's residents who saw a physician are .50 and .25. Percentages are simply proportions multiplied by a constant (100) in order to further enhance their clarity. Ratios are an economical means of expressing relative frequencies if a very small number of categories are of interest—typically only one or two. In these cases, the ratio is simply number A divided by number B. These are often converted to some convenient base such as unity, 100, 100,000, to form rates. This means of describing a frequency will often be utilized by an HSA—for example in the case of death rates.

When data are at the interval level of measurement a very large number of possible values might exist. For example, in a survey data file where age is represented in years, it would not be efficient to calculate the percent aged one year, the percent aged two years, etc. The data would typically be grouped in a manageable number of intervals (e.g., 0-10 years, 11-20 years, etc.) before percentages would be useful. (Obviously, such grouping would also have been completed before this type of data could be displayed in tables, charts, or whatever.) However, data of this type can be described very efficiently by the use of certain other measures which are not applicable to nominal frequency distributions.

In the example of the age variable, it becomes possible to calculate an <u>average</u> age and describe a population in terms of that measure.

Three types of averages are frequently used for such purposes:

- <u>mode</u>: the most frequently occurring score
- <u>median</u>: the value which divides the cases into two groups of equal size, with all scores in the higher group being higher than all scores in the lower group
- <u>mean</u>: the arithmetic average score

Each of these is a measure of central tendency or typicality in a distribution, yet each has distinct properties and will be more or less useful depending upon the data. The mode will be of value when a large number of cases are clustered around the modal value (i.e., the distribution is highly peaked) and the remaining cases are asymetrically distributed. The median is frequently used to describe distributions where there are considerably more extreme cases in one direction than the other (i.e., skewed distributions). In virtually all other cases, the mean is the most valuable measure of central tendency.

No measure of central tendency will in itself serve to adequately describe a distribution. This becomes intuitively obvious when one considers that a distribution of three cases with scores 0, 50, and 100 has no mode and both the mean and median equal 50, while a distribution of three cases with scores 50, 50, and 50 has no median, but the mean and mode both equal 50. The result would be two valid measures of central tendency for each of two very different distributions, and all four measures would be equivalent. It is obvious that the difference between the two distributions lies in the manner in which the scores are dispersed, and as a result, an adequate description of either distribution requires a measure of central tendency and a measure of dispersion. The simplest and most intuitively obvious measure of dispersion is the range—the difference between the highest and lowest scores. The most useful measure of dispersion, especially for distributions that approximate the normal curve,* is the standard deviation. This is defined as the square root of the arithmetic mean of the squared deviations from the mean and is expressed as:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$$

where,

s = standard deviation

x_i = the individual score

$\bar{x}$ = the mean

n = the number of cases

---

*The normal curve is a special type of frequency distribution which has a number of remarkable mathematical properties. In form, the normal curve is a unimodal, bell shaped symetrical curve. This form is, however, theoretical—actual data will only approximate the normal curve because it is based on an assumption of an infinite number of cases. Given a reasonable large number of cases the mathematical properties of the normal curve can be utilized in spite of variation from true normality. The most important feature of the normal distribution is that its actual form can be constructed from only two pieces of information—the mean and standard deviation. As a result, a fixed proportion of the cases under a normal curve will fall between the mean and an ordinate which is a given number of standard deviation units from that mean. The percent of all cases which fall in such areas need not be calculated each time. Tables indicating such areas have been constructed and are included in appendices in virtually all statistics text books.

The calculation follows these steps: 1) take the deviation of each score from the mean, 2) square this difference, 3) sum the squared differences, 4) divide by the number of cases, and 5) take the square root. Further details describing these calculations for both grouped and ungrouped data are available in all introductory statistics texts. Of course, the measure will most frequently be applied to large data files and will be calculated by computer.

The mean and standard deviation are fundamentally useful in simply describing a distribution. In addition, however, they provide the basis whereby one can set a confidence interval about an estimate and test the statistical significance of an observed difference between two means. Since both of these objectives are to be met in much of the data analysis described throughout this Handbook, a clear understanding of the two measures is essential.

3. The Confidence Interval

It is obvious that any estimate of a population mean which is derived from a sample is not likely to be precisely equivalent to the population mean. If another sample were drawn, it is likely that a different mean would be calculated, and while both sample means would be valid estimates they would probably vary from each other and from the true (i.e., population) mean by some unknown degree. Since we undoubtedly wish to make inferences about the population on the basis of sample data, it is important that we be able to describe the extent to which we are confident that the sample mean approximates the true mean. At the most obvious level, this is achieved by setting an <u>interval</u> around our estimate. If we were forced to provide an estimate of the average age in a population we would be more confident with an interval ("the average age is between 25 and 35 years") than a single point ("the average age is 30 years"). In the same way, when a sample mean is used as the estimate we would be more confident if an interval around that mean could be described. One of the valuable properties of probability samples is that they make it possible to set such an interval around an estimate <u>and</u> to associate a specific probability with that interval. In the example of estimating average age this would allow a statement to the effect of "we are 95 percent certain that the true mean (average age) is between 25 and 35 years."

The construction of a confidence interval, i.e., a range of values carrying a specified probability of containing the estimate to be obtained is based on an ability to estimate the average deviation by which all sample estimates would vary from the true mean. This average is called the standard error, and is valuable because the distribution of sampling errors (if multiple samples are drawn) is a normal distribution about the true mean. The standard error is the standard deviation of such a sampling distribution. Of course, the researcher does not have multiple samples available—he must estimate standard error from the single sample at hand.

When the standard error is conceptualized as the average error in multiple samples it becomes clear that two features of a single sample will be related to the size of the standard error. The first feature of interest is the sample size—clearly, larger samples will yield smaller standard errors. This principle is stated clearly by Mueller, *et al.*, — "the standard error of the mean (1) will always be smaller than the standard deviation of the universe since *n* (sample size) will always be larger than 1; and (2) will decrease as sample size increases, reaching zero when n = 100% of the universe."[14] The second feature of the sample which will be related to the standard error is its heterogeneity (which is presumably a reflection of the heterogeneity of the universe). Obviously, if all elements in a sample were alike, all sample means would be alike, and there would be no standard error. With samples where there is a great deal of scatter, sampling errors, and hence the standard error will be larger.

Knowledge of these two features provides the basis for calculating the standard error of the mean:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

where,

$\sigma_{\overline{x}}$ = standard error of the mean $\overline{x}$

σ = standard deviation of the universe

n = sample size

Since the researcher will not know the standard deviation of the universe we substitute the sample standard deviation (see previous formula). However, we noted that the standard deviation of the sample will typically be smaller than that of the universe so a correction in the use of the sample standard deviation is introduced by reducing n by 1. Hence the working formula for sample standard deviation is

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n - 1}}$$

and the formula for standard error becomes:

$$s_{\overline{x}} = \frac{s}{\sqrt{n}}$$

For calculation in a single step, this is equivalent to

$$s_{\overline{x}} = \frac{\sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n - 1}}}{\sqrt{n}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n(n - 1)}}$$

At this point the confidence interval is easily constructed. It will be recalled that the distribution of sample means will be normal about the true mean. It follows, therefore, that any given mean will fall within the interval of one standard error on either side of the true mean 68.27 percent of the time, within 1.96 standard errors 95 percent of the time and within three standard errors 99.74 percent of the time. Comparable figures for any specified percent can be found in a table of the areas under the normal curve. Areas under the normal curve are presented in brief tabular form in virtually all statistics texts. The researcher is thus required to select the confidence level which he considers appropriate (95 percent is commonly used) and calculate the boundaries of the relevant interval:

$$\text{interval} = \bar{x} \pm 1.96\ (s_{\bar{x}})$$

Since much of the data to be used in HSAs (extant as well as primary data) will be based on samples, this procedure will be useful. Obviously, being confident at the .95 level that the true mean is within some specified range of the sample mean is simply another way of saying that there is a .05 probability that the true mean is outside of that range. This knowledge must be part of decision-making processes.

4. Testing Differences

Whether it be for purposes of assessing relative health status, examining health services utilization, evaluating average lengths of hospital stay or a myriad of other questions, an HSA will frequently find it useful to examine <u>differences</u> between measures of health-related phenomena in the HSA and some referent population. In some cases this will serve purely descriptive purposes—for example, when an HSA wishes to compare the physician to population ratio in its constituent counties with some standard. In other cases, such comparisons will form the basis for drawing conclusions about health services in the area and lead to specific actions consistent with those conclusions.

Suppose, for purposes of explanation, that an HSA wishes to examine the average length of stay in some hospital or group of hospitals and compare that average with a regional norm. Clearly, in practice such a comparison would involve a whole series of averages for different patient groups. However, let us concentrate on a single figure for purposes of clarity. If the HSA collected information on every patient in every hospital of concern for a year, and the regional norm was similarly computed from data referring to every patient in every hospital, the comparison would be simple. Average length of stay in the hospitals of interest would be computed, comparable figures would be derived from the regional source of data and the difference noted.

However, this situation is not likely to obtain. Regional or national norm data are likely to be derived from PAS or NCHS data. The former refer to member hospitals*(a non-random sample of all hospitals) and the latter are derived from a national probability sample of hospitals and a sample of discharges from each selected hospital.[15] Presumably, most HSAs which complete a hospital discharge survey will also be constrained to use sampling procedures as the only efficient method of gathering the data (see Chapter 12 in Section IV). As a result, the average length of stay figures which are to be compared will typically be estimates based on samples and hence subject to the ubiquitous sampling error noted previously. Assume that the sample estimate for the region is an ALS = 4.0 days and the estimate for the hospital(s) of interest to the HSA is 5.0 days. It is still possible that the two true means are not different—perhaps they are both 4.5 days. The HSA will presumably wish to avoid erroneously concluding that there is a difference in the average length of stay.

This is achieved by calculating a measure of the probability that an observed difference between two means occurred simply by chance (due

---

*PAS (Professional Activity Study) hospitals are those participating in the ongoing study by the Commission on Professional and Hospital Activities. The data are published in a series of companion volumes, displaying information for the United States, four regional subdivisions and Canada.

to sampling error). Obviously, for very crucial decisions, the HSA will wish to set a very low probability (or very high confidence level) and be willing to draw conclusions only when observed differences are very unlikely to be a result of chance factors. The probability that an observed difference occurred by chance is obtained by testing the level of significance of the difference. An observed difference between two means which could happen by chance only 1 time in 100 is said to be significant at the level of .01. The level of significance is not a measure of the magnitude of observed differences; it is simply an estimate of the probability that the difference, whatever its size, occurred by chance.

A straightforward test of the significance of an observed difference between two means can be achieved by the calculation of a statistic called "t."* The calculation requires three pieces of information from each of the two samples:

- the two means
- the variance around each mean (variance = the square of the standard deviation)
- the number of cases upon which each mean is based

The formula for t takes the form

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where,

$\bar{x}_1$ = the mean in sample 1

$\bar{x}_2$ = the mean in sample 2

*This statistic was originally developed by a statistician named W. S. Gossett, who first published his results under the pseudonym "student." Hence the statistic, and its related distribution are sometimes called "student's t."

s_1^2 = the variance in sample 1

s_2^2 = the variance in sample 2

n_1 = the number of cases in sample 1

n_2 = the number of cases in sample 2

The computed value for t can then be located in a table of areas under the normal curve at whatever level of significance had been chosen in advance, and the hypothesis that $\overline{x}_1 = \overline{x}_2$ can be rejected or not as indicated by the value t and the chosen level of significance. Perhaps an extended numerical example will serve to clarify this procedure.

Assume that an HSA has gathered appropriate data and derived the following estimates regarding ALS for a specified subset of patients in a group of hospitals in the HSA and analogous estimates for the nation.

	HSA	NATION
ALS	5.9	4.3
Variance	15	13
n	1,000	10,000

The HSA staff is concerned that the local ALS appears higher than that of the nation, but before final derivation of such a conclusion wishes to test the hypotheses that the means are not different. The HSA prefers to be somewhat cautious and has determined that it will conclude that there is in fact a difference in means only if the obtained difference is such that it could happen by chance no more than 1 time in 100. The t statistic is calculated as follows

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{5.9 - 4.3}{\sqrt{\frac{15}{1000} + \frac{13}{10000}}}$$

$$= \frac{1.6}{\sqrt{.015 + .0013}}$$

$$= \frac{1.6}{.1277}$$

$$= 12.53$$

Recall that a significance level of .01 was selected. Reference to a table of the distribution of t indicates that the value 12.53 lies well beyond the critical value of 99 percent of the curve, indicating that the observed difference would occur by chance much less often than 1 time in 100. The HSA is therefore confident that the measured ALS in the hospitals of interest is not equal to that in the nation. It bears reemphasis that this does not imply anything about the magnitude or substantive importance of the difference; statistical significance has no direct relationship to such questions. This test simply indicates that the observed difference is extremely unlikely to have occurred by chance.

It is important to note the effect of sample size on such procedures. If two samples of size 50 and 75 respectively were the basis for deriving the same estimates of average length of stay and variance, the resulting t statistic would be 2.33 which is not significant at the .01 level.

It is obvious that the scope of this chapter has been limited to a somewhat cursory discussion of a wide range of methodological and statistical issues. The statistics and statistical tests described are only examples; there are many other alternatives. Similarly, the issues which were described regarding data collection must be seen as limited to very basic considerations. In this regard the foregoing must be seen not as a "final word" on methodology, but as an introduction. We have, however, attempted to choose examples which are salient, emphasize issues that are likely to obtain in the HSA setting, and present the complete spectrum of relevant issues. It is hoped that one principle has been made clear: the research process is complicated and is characterized by an almost continuous series of decisions, each of which limits the number of options at subsequent stages. Thus, the importance of developing an overall plan for any specific study cannot be over-emphasized. Since only the most basic components of such an overview have been presented here, the reader should be sure that the principal issues are clearly understood, and refer to relevant literature for such clarification as is appropriate. The various studies described in the remainder of this Handbook assume a basic understanding of the foregoing material.

FOOTNOTES - SECTION I

Chapter 2

1. R. L. Eichhorn and T. W. Bice, "Academic Disciplines and Health Services Research" in E. Evelyn Flook, *et al.* (eds.) *Health Services Research and R & D in Perspective*, Ann Arbor: Michigan, 1973.

2. P.L. 93-641, Section 1513 (b).

3. *Ibid*.

4. cf. J. H. Mueller, K. F. Schuessler and H. R. Costner, *Statistical Reasoning in Sociology*, New York: Houghton Mifflin, 1970; C. Selltiz, M. Jahoda, M. Deutsch and S. W. Cook, *Research Methods in Social Relations*, New York: Holt, Rinehart and Winston, 1967; E. R. Babbie, *Survey Research Methods*, Belmont, California: Wadsworth, 1973.

5. R. L. Eichhorn and T. W. Bice, *op. cit.*, p. 138.

6. S. Labovitz and R. Hagedorn, *Introduction to Social Research*, New York: McGraw-Hill, 1971, p. 29.

7. L. Kish, *Survey Sampling*, New York: John Wiley & Sons, 1965, p. 23.

8. *Ibid*.

9. *Ibid*.

10. *Ibid*., p. 24.

11. *Ibid*., p. 25.

12. Anthony Oreglia, *et al.*, *A Guide to the Development of Health Resource Inventories*, West Lafayette, Indiana: Health Planning Data Project, Purdue University, 1976.

13. J. H. Mueller, *et al.*, *op. cit.*, p. 55.

14. *Ibid*., p. 372.

15. National Center for Health Statistics, "Utilization of Short-Stay Hospitals: Summary of Nonmedical Statistics, United States-1972," *Vital and Health Statistics*, Series 13, no. 19, DHEW Publication No. (HRA) 75-1770, June 1975.

SECTION II

POPULATION, SOCIOECONOMIC AND HEALTH STATUS DATA

Chapter 3

The Need for Population, Socioeconomic and Health Status Data

A. POPULATION AND SOCIOECONOMIC DATA

Among the factors which health systems agencies (HSAs) and State Agencies will need to consider in deciding which health resources are needed in their health service areas is the size of the population. It will be evident throughout this Handbook that the availability of accurate estimates of the number of persons who reside in a particular health service area and estimates of their projected rate of growth are essential in the estimation of health services needs and the resources required to meet those needs. For example, an HSA will need to know how many persons reside in specified subareas of its health service area in order to calculate the required population to physician ratios for purposes of estimating and justifying the need for a new physician under the National Health Service Corps program (see Chapter 8 in Section III).

It is made quite explicit in P.L. 93-641 that HSAs and State Agencies will need to know the rate of population growth in their health service areas for purposes of institutional health services development. For example, under Title XVI of this law, federal funds for the construction of new inpatient medical facilities will be provided only "in areas which have experienced...recent population growth."[1]

In addition to determining the size of the population of the health service area, information is also needed concerning the composition of the population and its socioeconomic status. Information concerning the age structure of the population or its median family income, for example, can be very useful in the determination of needed health resources because of the correlation of these population characteristics with known variations in health services utilization. Such information will be required by HSAs and State Agencies in fulfilling several of their functions as described under P.L. 93-641. For example, as part of its review and comment authority, an HSA is required to consider "the special needs and circumstances of health maintenance organizations...."[2] In the process of reviewing an HMO application which projects an enrollment of members from a "medically underserved area," an HSA must decide if the applicant has accurately described the underserved area, taking into consideration 1) the ratio of primary care physicians to population; 2) the infant mortality rate; 3) the percentage of the population with incomes below the poverty level; and 4) the percentage of the population age 65 and over.[3] In deriving these measures, current estimates of population size will be required as denominator data in the first measure; data concerning the population's economic status will be required in the third measure; and data concerning the population's age structure will be required in the fourth measure.

Because of the need of HSAs and State Agencies to know how many persons reside in their health service areas and the demographic and socioeconomic characteristics of these persons, the first part of Chapter 4 in this section reviews in some detail the availability of these type of data from the Bureau of the Census. The latter part of Chapter 4 describes the various uses that an HSA and/or State Agency can make of these data.

B. HEALTH STATUS DATA

HSAs will have a number of uses for health status data, one of the most important of which will be the assessment of their own performance. As specified in P.L. 93-641, improvement in the health of their service area's residents is one of the purposes for establishing HSAs,[4] and the extent to which it can be demonstrated that the health of these residents has been improved will be a criterion in the review of HSAs to be conducted by the Secretary of HEW.[5] Such demonstrations can be made on the basis of health indicators which are derived from various types of health status data. Toward that end and others which pertain more directly to planning and resources development functions, each HSA is required under P.L. 93-641 to "assemble and analyze data concerning the status (and its determinants) of the health of the residents of its health service area."[6]

HSAs will be faced with the difficult task of defining measures of the health status of their populations for purposes of identifying and prioritizing their area's health problems. In general, such measures are referred to as health (status) indicators. Although the terms are sometimes used interchangeably, a special subset of health indicators which are based on multivariate components and have certain mathematical characteristics are called health (status) indices.[7]

1. Health Indicators

Indicators of the health status of a population are quantitative measures which purport to describe the health of that population. They are constructed from one or more numerical variables which are presumably related to the concept of health, there being no meaningful <u>direct</u> empirical measure thereof. The complexity of any health indicator is, therefore, a function of two things: 1) the complexity of its referent concept (i.e., health) and 2) the accuracy with which the indicator is expected to reflect this concept.

This can be illustrated as follows. If a concept such as health is defined in a very simple way, e.g., a non-death, a very simple indicator of health in a population is a mortality rate. Given accurate mortality data, such an indicator would provide a valid and reliable measure of health status as defined. It is clear, however, that simple survival is not equivalent to health; two populations which have the same death rates but divergent frequencies and degrees of illness can hardly be said to have the same level of health. Hence, some analysts have argued that levels of morbidity in a population are relevant dimensions of its health status. Each such addition to the definition of the concept of health increases its complexity. Since measures of a concept should reflect each relevant dimension of it, increasingly complex definitions of health require increasingly complex indicators.

The often-quoted definition of health adopted by the World Health Organization serves to demonstrate this point. When health is defined as "a state of complete physical, mental and social well-being, and not simply the absence of diseases and infirmity",[8] two measurement problems are introduced. On the one hand, a definition like this implies that health should be measured in "positive" terms rather than by an assessment of the relative frequency of one or more negative components (death, disease, disability, etc.). Work in this direction is nascent, and is hampered by the lack of an adequate theoretical conception of positive health.[9] On the other hand, the development of more sophisticated negative indices required by the complexity of the definition involves the slow process of empirical verification and is hampered by the known flaw that such indicators tend to "discriminate only among levels of poor health."[10]

In view of these issues it is emphasized that there are no absolute criteria for evaluating various measures of health status, although several authors have set forth elements of such criteria ranging from highly technical and mathematical stipulations to seemingly obvious (but often difficult) demands for linguistic clarity of purpose.[11] Essentially, in addition to reliability and validity, such features as the ability to reflect changes over time in significant aspects of the health of the living as well as in mortality, and susceptibility to the partitioning of components in order to link such components to the specific issues of concern would seem to be most important.

2. Health Indicators and Health Planning

When a health indicator is viewed from the perspective of health planning, certain additional issues become significant. First of all, in order to proceed with legislatively mandated activities, the planner must be prepared to [illegible]iate and use measures of health status in spite of their known limitations. Thus, in addition to necessary concern with the scientific and theoretical issues described above, the planner requires that indicators be 1) easily calculable and 2) susceptible to logical linkage not only to underlying health problems but also to one or more component(s) of the health services delivery system. Finally, the planner is, in the HSA context, constrained to use extant data where possible.

Ease of calculation is taken to include several dimensions. Obviously, the required mathematics should not presume highly sophisticated statistics or computer science techniques. The necessary data should be readily available on a continuous basis, amenable to straightforward translation into indicator components, and the actual derivation of the numbers should not require time or other resources to the extent that the time spent upon calculation outweighs the utility. Finally, the indicator must have a level of intuitive meaning that allows it to serve HSA community educative functions.

Related to this last dimension is the necessity that a measure be amenable to logical linkage with some relevant aspect(s) of the health services delivery system with which the agency is concerned. However accurate an indicator might be, its value to the health planner is limited unless it permits him to identify (at the minimum) an appropriate relationship between it and certain aspects of the health services delivery system. That is, the indicator should provide a basis for assuming that action which results in changes in the delivery system will be followed by an appropriate change in the indicator.

It must be stressed, of course, that the demonstration of <u>causal</u> relationships between an indicator and delivery system phenomena is not generally possible. However, since the planner cannot simply wait for the results of further research, he must demand, in lieu of demonstrable causal relationships, a clear and defensible logical connection. This allows the incorporation of an admittedly imperfect rationality into planning decisions and provides a basis for testing whether or not such planning decisions have the expected result.

These additional constraints on the utility of health indicators are formidable. The first decreases the range of applicable indicators to those which are easily calculable; the second restricts the planner to those which can be linked to specific types of health services. At the same time, since the indicators may be the basis for actions which will presumably affect the well-being of HSA residents, the need for validity and reliability increases. Because health is a concept of great complexity, a valid over-all indicator thereof must include measures of its numerous aspects. Such an indicator would not, in all probability, be easily calculable and because it would be multi-dimensional, linkages between the index and features of the delivery system would be difficult to demonstrate.

Partly as a result of this difficulty (which is really a demonstration of the fact that the state of the art is not highly advanced) health status indicators should be treated with considerable caution. In the discussion of health indicators in Chapter 5 of this section, simpler indicators are emphasized because linkages to relevant delivery system components are more clearly demonstrable. In addition, such indicators are more in keeping with the problem approach which is fundamental to health planning and resources development. Those

HSA's wishing to pursue an examination of indicators other than those discussed in Chapter 5 will find the quarterly publication _Bibliography on Health Indices_ a useful source of information and current research results. For information, or to be included on the mailing list contact:

Ms. Pennifer Erickson
Clearinghouse on Health Indices
Division of Analysis
National Center for Health Statistics, DHEW
Rockville, Maryland 20852

Chapter 4

The Sources and Uses of Population and Socioeconomic Data

A. SELECTION AND DEFINITION OF DATA ITEMS

The censuses conducted by the Census Bureau are the major sources of population and socioeconomic data.* Of all the demographic and socioeconomic data contained in the census of population, the following data items will be of most value to an HSA (the definition of the following data items are those utilized by the Census Bureau)[1]:

- total population of a geographic area recognized in census tabulations comprises all persons enumerated whose usual place of residence (i.e., usual place of abode, generally the place one lives and sleeps) at the time of the census was determined to be in that area.
- age is usually determined in completed years as of the time of enumeration from replies to a question on month and year of birth.
- sex
- race refers to the division of the population into white, Negro and several other racial categories. Respondents were asked to indicate their race by selecting one of several categories listed on the census questionnaire.

---

*Another important socio-demographic data source is the Mental Health Demographic Profile System which has been developed by the National Institute of Mental Health. These data are available for each community mental health center catchment area in the country and for each census tract, county, and minor civil division in the nation. The data included in the system cover: socioeconomic status (economic status, social status, educational status), ethnic composition, household composition and family structure, type and condition of housing and community instability. For information on how the data from this system may be obtained, see National Institute of Mental Health, Mental Health Demographic Profile System Description: Purpose, Contents and Samples of Uses. Series C, no. 11, DHEW Publication No. (ADM) 76-263, 1975. This document is available from the Superintendent of Documents, U.S. Government Printing Office, Washington D.C., 20402. The data in the Mental Health Demographic Profile System are derived from the dicennial census.

- *income* is the total monetary income ascertained for all persons 14 years of age and over for the preceding calender year, even if they had no income. Total income is the sum of the dollar amounts of money respondents reported receiving as wages or salary income, net nonfarm and farm self-employment income, and other income.

- *education* is defined as the number of school years completed as ascertained for persons 3 years of age and over, who were asked the highest grade or year or regular school they ever attended up to 6 or more years of college. The number tabulated in each category of years of school completed includes persons who report completing that grade or year plus those who attended but did not complete the next higher grade.

More detailed definitions of the above data items can be found in the *1970 Census Users' Guide*.[2] A rationale for selecting these particular data items is presented in the section which describes their availability from census publications.

B. DESCRIPTION OF DATA SOURCE

The <u>1970 Census of Population and Housing</u> is the principal source of population and socioeconomic data that are available to HSAs. Over a period of several weeks beginning April 1, 1970, each individual in the United States was enumerated according to their usual place of residence.*

1. Subject Content

Of all the questions asked in the 1970 census, only five were asked of all individuals. These "complete count" or "100-percent" items (i.e., relationship to household head, sex, race, age and marital status) permit a relatively accurate count of persons in each geographic area. All of the other questions asked in the 1970 census were asked of only a sample of the population.

> There was a 15-percent and a 5-percent sample in the 1970 census, and certain questions common to both samples resulted in a 20-percent sample. Whether a question was asked of everyone or of a sample of people depended in part on the size of the area for which statistics were to be tabulated and published. Information required for apportionment purposes and that which was to be tabulated for city blocks was collected on a 100-percent basis; that which was to be tabulated for larger areas, the smallest being a census tract, was asked on a 15- or 20-percent sample basis. The

---

*Taking into account the size of the population being enumerated, the United States census is probably as accurate as any census in the world. However, errors do occur. One of these errors is underenumeration. In the 1970 census, the Census Bureau has estimated that there was a net underenumeration of the U.S. total population of 5.3 million persons or 2.5 percent. The rate of underenumeration varied by race, sex and age. Among those with the highest rates of underenumeration were black males, with the rate for those aged 25-44 reaching as high as 17 to 19 percent. For a further discussion of this issue see, U.S. Bureau of the Census, <u>Current Population Reports</u>, Series P-23, No. 56, "Coverage of Population in the 1970 Census and Some Implications for Public Programs," U.S. Government Printing Office, Washington, D.C., 1975.

> 5-percent sample provides reliable data for all large counties, and States.[3]

Table 1 contains the lists of population and housing items asked of the entire population and the respective samples in both the 1960 and 1970 censuses.*

2. Geographic Coverage

Not all of the information collected in the census is made available for all the various geographic areas recognized by the Census Bureau. The size of the geographic area determines the extent of information available, and, in general, the larger the area, the greater the number of tabulations produced and published. The Census Bureau collects, tabulates and publishes its data for a myriad of overlapping federal, state and local political and administrative boundaries; however the most widely recognized and used boundaries include: region, division, state, county, standard metropolitan statistical area, urbanized area, place, minor civil division, census tract, block group and city block. The definitions of these areas, which follow, are relatively general; the "Census Users' Dictionary" in the 1970 Census User's Guide contains more precise and detailed definitions of these areas, as well as those not mentioned here.

> Divisions are areas composed of contiguous states, and regions are units composed of two or more geographic

*One of the more important general uses of census data is their comparability from decade to decade in order to chart statistical trends. Comparability of 1960 and 1970 census data for certain subject matter and geographic areas may be precluded by additions, deletions, or alterations of individual questions and/or boundaries of census geographic areas. Generally, there is greater comparability by subject content than be geographic area, especially for the smaller areas, such as census tracts and blocks. Redefinition of many census tract boundaries occurred between the 1960 and 1970 censuses. Users should be cautious of possible incompatabilities when comparing intercensus data. For a definition and discussion of the various terms appearing in Table 1 such as "race," "employment status," "tenure," etc., see the 1970 Census Users' Guide.

TABLE 1

Content and Coverage Comparison, 1960-1970

The sample percentages for population and housing items included in the 1970 census in comparison with the items in the 1960 census are shown below.

Population Items	1960	1970
Relationship to head of household	100%	100%
Color or race	100	100
Age (month and year of birth)	100	100
Sex	100	100
Marital status	100	100
State or country of birth	25	20
Years of school completed	25	20
Number of children ever born	25	20
Employment status	25	20
Hours worked last week	25	20
Weeks worked last year	25	20
Last year in which worked	25	20
Occupation, industry, and class of worker	25	20
Activity 5 years ago	-	20
Income last year:		
Wage and salary income	25	20
Self-employment income	25	[1]20
Other income	25	[2]20
Country of birth of parents	25	15
Mother tongue	25	15
Year moved into this house	25	15
Place of residence 5 years ago	25	[3]15
School or college enrollment (public or private)	25	15
Veteran status	25	15
Place of work	25	[4]15
Means of transportation to work	25	15
Mexican or Spanish origin or descent	-	5
Citizenship	-	5
Year of immigration	-	5
When married	25	[5]5
Vocational training completed	-	5
Presence and duration of disability	-	5
Occupation-industry 5 years ago	-	5

Housing Items	1960	1970
Number of units at this address	-%	[6]100%
Telephone	25	[7]100
Access to unit	100	100
Kitchen or cooking facilities	100	-
Complete kitchen facilities	-	100
Condition of housing unit	100	-
Rooms	100	100
Water supply	100	100
Flush toilet	100	100
Bathtub or shower	100	100
Basement	25	100
Tenure	100	100
Commercial establishment on property	100	100
Value	100	100
Contract rent	100	100
Vacancy status	100	100
Months vacant	25	100
Components of gross rent	25	20
Heating equipment	25	20
Year structure built	25	20
Number of units in structure and whether a trailer	20	20
Farm residence (acreage and sales of farm products)	25	20
Land used for farming	25	-
Source of water	20	15
Sewage disposal	20	15
Bathrooms	20	15
Air conditioning	5	15
Automobiles	20	15
Stories, elevator in structure	20	5
Fuel--heating, cooking, water heating	5	5
Bedrooms	5	5
Clothes washing machine	5	5
Clothes dryer	5	5
Dishwasher	-	5
Home food freezer	5	5
Television	5	5
Radio	5	5
Second home	-	5

[1] Single item in 1960; two-way separation in 1970 by farm and nonfarm income.
[2] Single item in 1960; three-way separation in 1970 by social security, public welfare, and all other receipts.
[3] This item is also in the 5-percent sample but limited to State of residence 5 years ago.
[4] Street address included in 1970.
[5] In 1960, whether married more than once and date of first marriage; in 1970, also includes whether first marriage ended by death of spouse.
[6] Collected primarily for coverage check purposes.
[7] Required on 100-percent for field followup purposes.

Source: U. Bureau of the Census, The 1970 Census and You. U.S. Government Printing Office, Washington, D.C., 1973, p. 3.

divisions. There are nine geographic divisions and four regions. See Figure 1.

- States, of course, are the major political units of the United States, and counties are the primary political and administrative divisions of the states.*

- Standard Metropolitan Statistical Areas (SMSA's) consist of a county or group of counties containing at least one city (or twin cities) with a population of 50,000 or more, plus adjacent counties which are metropolitan in character and are economically and socially integrated with the central city. The name of the SMSA is determined by the name of the central city. There is no limit to the number of adjacent counties included in an SMSA as long as they are integrated with the central city; nor is an SMSA limited to a single state: boundaries may cross state lines. There were 247 SMSA's recognized in the 1970 census.

- Urbanized Areas (UA's) contain a city (central city) or twin cities of 50,000 or more population plus the surrounding closely settled incorporated and unincorporated areas which meet certain criteria of population size and density. They differ from SMSA's chiefly in excluding the rural portions of counties comprising the SMSA's and excluding those places which are separated by rural territory from the densely populated fringe around the central city.

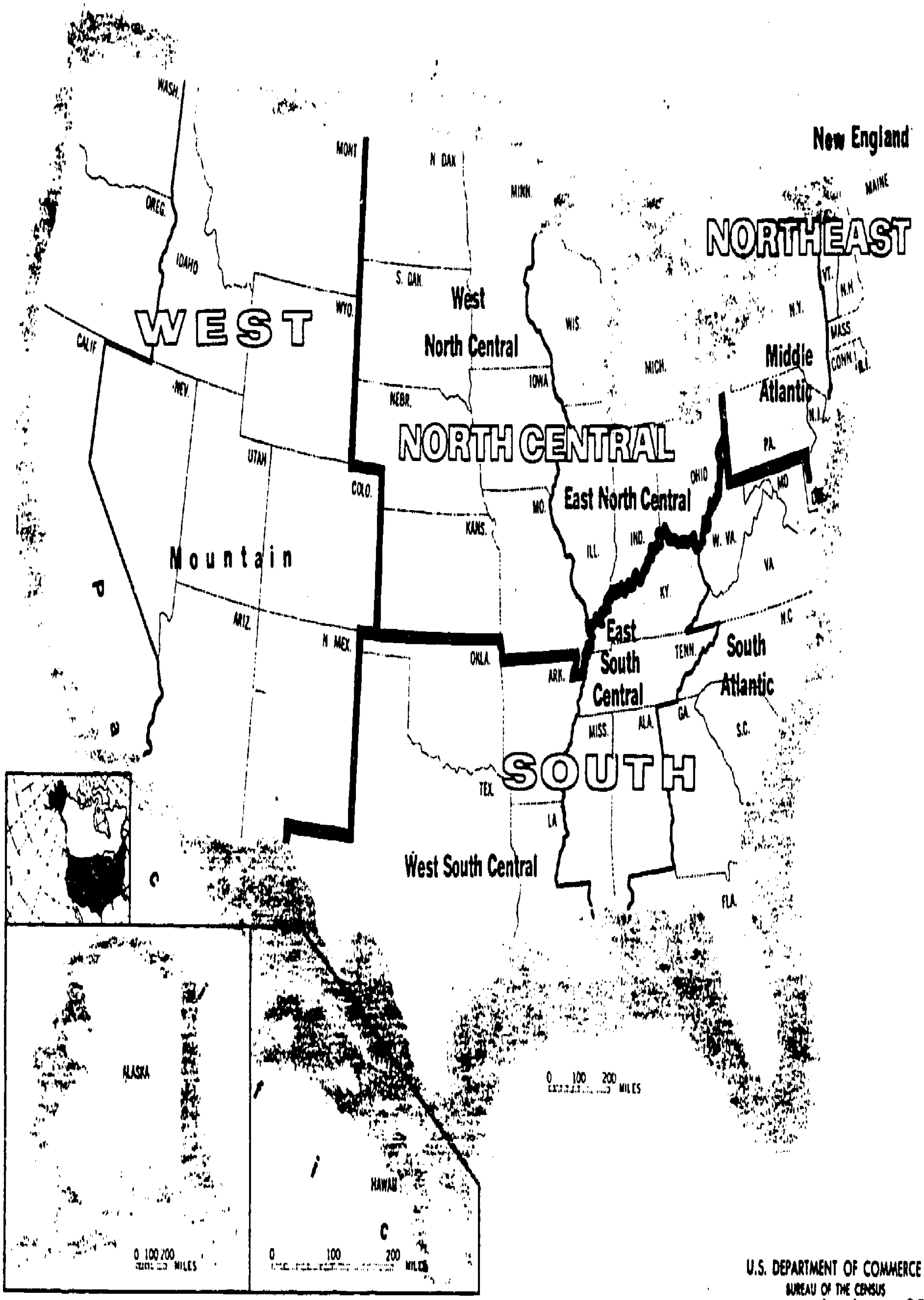
- Places (cities and other incorporated and unincorporated places) as used in the decennial population and housing census, refer to a concentration of population regardless of the existence of legally prescribed units, powers or functions. However, most of the places identified in the census are incorporated as cities, towns, villages or boroughs. All incorporated places regardless of size are delineated in census reports. For unincorporated places only those of 1,000 or more population are listed in the reports.

- Minor Civil Divisions (MCD's) are the primary political and administrative subdivisions of a county, for example, townships, precincts, magisterial districts, etc. MCD's vary considerably in geographic size and population. For those states in which MCD's are not suitable for presenting statistics, census county divisions (CCD's) are established by the Census Bureau.

---

*In Louisiana the primary political and administrative divisions are called parishes.

Figure 1. CENSUS REGIONS AND GEOGRAPHIC DIVISIONS OF THE UNITED STATES



Source: U.S. Bureau of the Census, 1970 Census Users' Guide, U.S. Government Printing Office, Washington, D.C., 1970, p.77.

73

- Census Tracts are small, relatively permanent areas into which large cities and adjacent areas have been divided for the purpose of showing comparable small area statistics. They are designed to be relatively homogeneous in population characteristics, economic status, and living conditions. The average census tract has a population of about 4,000 persons, although tracts having a population of as few as 1,000 persons are not unusual. All SMSA's recognized in the 1970 census are completely tracted. In addition, over 2,000 census tracts are recognized in non-SMSA cities and counties that contracted for them.

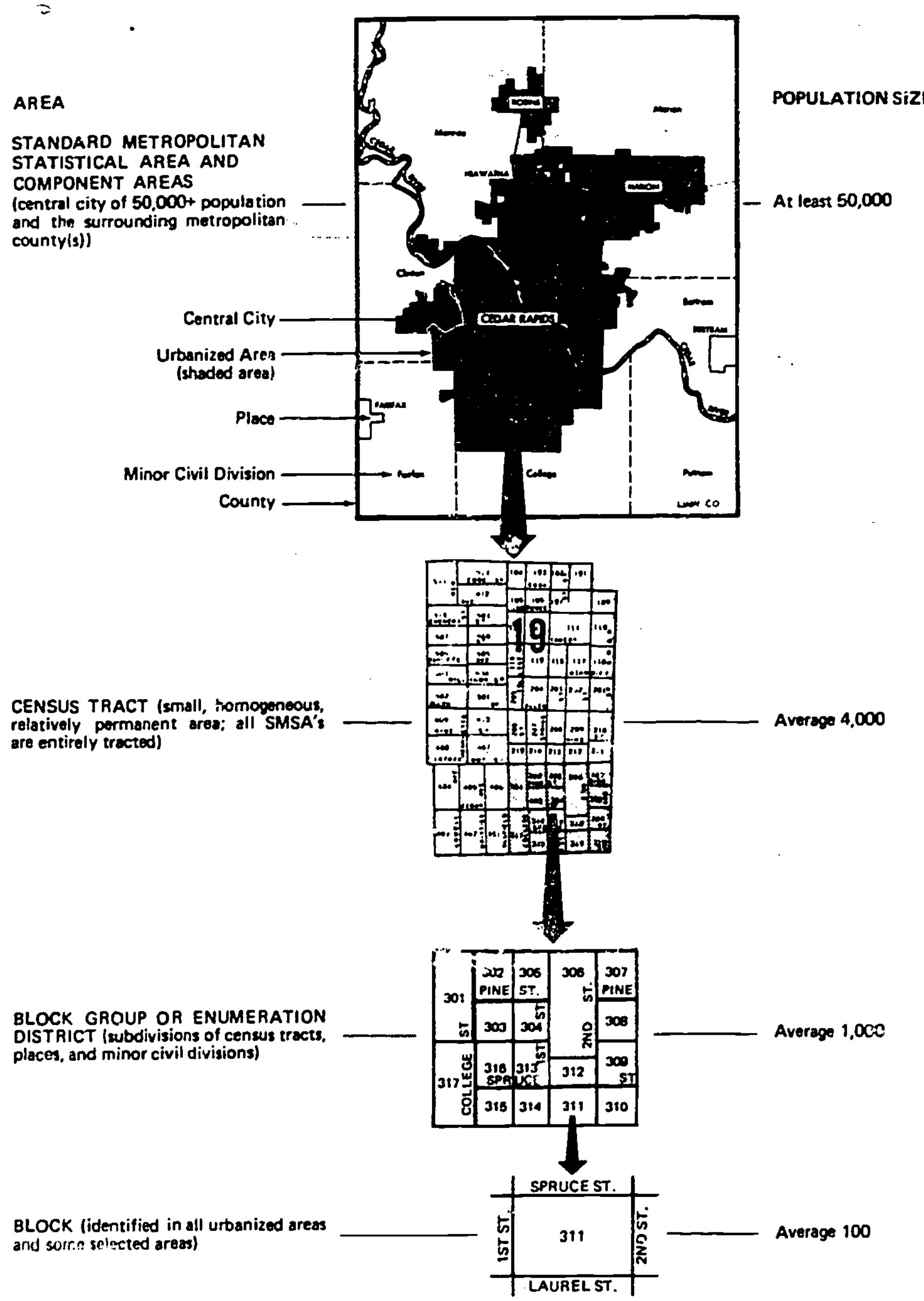
- Block Groups (BG's) are a combination of contiguous blocks containing between 200 and 600 households and having a combined average population of about 1,000. The term "block groups" corresponds to what was referred to in previous censuses as "enumeration districts." Block groups are always subdivisions of a census tract and there are, on the average, four blocks per census tract.

- Blocks are normally a well-defined rectangular piece of land, bounded by streets and roads, and contain an average population of 100. There are approximately ten blocks per block group and forty blocks per census tract. Blocks may not cross census tract boundaries, but may cross other boundaries such as city limits. Block data are published for all cities with populations of 50,000 or more. There are about 350 such cities. Block statistics are also published for approximately 900 cities and other areas that contracted for block statistics.

An illustration of the relative sizes of several of the above geographic areas is presented in Figure 2. Another important geographical distinction in the publication of 1970 census data includes the urban/rural distinction. For the 1970 census the urban population comprised all persons living in: 1) places of 2,500 population or more; or 2) the densely settled urban fringe, whether incorporated or unicorporated, of urbanized areas. Rural areas are those remaining areas not falling into one of the above categories.

Figure 2.

COMMON CENSUS GEOGRAPHIC AREAS



Source: U.S. Bureau of the Census, The 1970 Census and You. U.S. Government Printing Office, Washington, D.C., 1973, p. 5.

3. Data Products

The information contained on each of the individual census questionnaires was edited and processed onto a basic record tape (BRT). Because the BRT's contain information about each household, they are confidential and may be tabulated for statistical purposes only by Census Bureau employees.

Data summaries are prepared on computer tapes from the BRT's for each of the geographic areas: blocks, tracts, places, counties, etc. The tapes containing these area tabulations are called census summary tapes. Summary tapes in turn are used to prepare more limited sets of tabulations which appear in the printed reports. Some tabulations contained on the summary tapes and in the printed reports are also available on microfilm and microfiche.

- Printed Reports
  The printed reports are inexpensive, widely available data products of the 1970 census. Individual printed reports for each state can generally be purchased for under $5.00. These reports may be purchased directly from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402, or from U.S. Department of Commerce field offices. A summary of printed reports is presented in Table 2.

- Summary Tapes
  Summary computer tapes were released in several series, referred to as "counts." Table 3 summarizes available 1970 census summary tape products.

  Summary tapes by themselves are merely strings of numbers unintelligible to users without translation devices. With every tape file distributed to users, the Census Bureau supplies a copy of the appropriate technical documentation which describes the content of that particular tape file. Also, distinguishing among the many different areas for which data are presented on a summary tape reel requires a geographic identification tool called the Master Enumeration District List. It enables a data user to interpret the geographic codes which are part of each data record.

  Summary tapes can be purchased directly from the Census Bureau. The price for 1970 census summary tapes is $70 per reel. This amount covers the cost of the tape, technical documentation and handling charges. However, the

TABLE 2

Select Reports
1970 Census of Population and Housing

Series PC(1)-A (One per State)	Number of Inhabitants. Final official population counts are presented for States, counties, SMSA's, urbanized areas, minor civil divisions, census county divisions, all incorporated places, and unincorporated places of 1,000 inhabitants or more.
Series PC(1)-B (One per State)	General Population Characteristics. Statistics on age, sex, race, marital status, and relationship to head of household are presented for States, counties, SMSA's, urbanized areas, minor civil divisions, census county divisions, and places of 1,000 inhabitants or more.
Series PC(1)-C (One per State)	General Social and Economic Characteristics. These reports will focus on the population subjects collected on a sample basis. Each subject is shown for some or all of the following areas: States, counties, SMSA's, urbanized areas, and places of 2,500 inhabitants or more.
Series HC(1)-A (One per State)	General Characteristics for States, Cities, and Counties. Statistics on 100-percent housing subjects are presented for States, counties, SMSA's, urbanized areas, and places of 1,000 inhabitants or more
Series HC(1)-B (One per State)	Detailed Characteristics for States, Cities, and Counties. These reports focus on the housing subject collected on a sample basis. Each subject is shown for some or all of the following areas: States, counties, SMSA's, urbanized areas, and places of 2,500 inhabitants or more.
Series HC(3) (One per UA)	Volume III, Block Statistics. One report for each urbanized area showing data for individual blocks on selected 100-percent housing and population subjects. The series also includes reports for the communities outside urbanized areas which have contracted with the Census Bureau to provide block statistics from the 1970 census.
Series PHC(1) (One per SMSA)	Census Tract Reports. One report for each SMSA, showing data for most of the population and housing subjects included in the 1970 census. Some tables are based on the 100-percent data, others on the sample data.

Source: U.S. Bureau of the Census, The 1970 Census and You, U.S. Government Printing Office, Washington, D.C., 1973, p. 7.

TABL.

1970 Census Summ. apes

Summary Tape Series*	Areas Reported	Type of Data	No. of Tables and Data Cells for Each Area**
First Count	State, Congressional District, County, MCD or CCD, Place, and Block Group or Enumeration District	100 percent	54 tables containing 400 cells
Second Count	State, SMSA and Component Areas, County, MCD or CCD, Place, and Tract	100 percent	93 tables containing 3,500 cells
Third Count	Block	100 percent	36 tables containing 250 cells
Fourth Count	State, SMSA and Component Areas, County, MCD or CCD, Place, and Tract	100 percent and sample	327 tables containing 13,700 cells (more for State, County, and SMSA and Components)
Fifth Count	ZIP areas (3-digit areas nationwide; 5-digit areas only in SMSA's)	100 percent and sample	53 tables containing 900 cells
Sixth Count	State, SMSA, Metro. County, City of 50,000+, Central City	100 percent and sample	440 tables containing 260,000 data cells

*Referred to as "Counts" by the Census Bureau, e.g., First Count, Second Count, etc.

**A table, for example, might show the count of persons who are white, Negro, and other races. In this case, the table would have three cells.

Source: U.S. Bureau of the Census, The 1970 Census and You. U.S. Government Printing Office, Washington, D.C., 1973, p. 9.

cost of the tapes themselves represents only a small fraction of the total costs in computer equipment, operating time, programs, and personnel of actually working with the data and applying them to user's research and decision-making systems. Census data users may be able to lower these costs by cooperating with other users or by obtaining the services of <u>Summary Tape Processing Centers</u>.* These services would be especially important to those HSAs not possessing adequate computer facilities.

The decision to use either printed reports or summary tapes will depend upon several factors. However, probably the most important of these factors will be the subject content and geographic coverage needed by the HSA. Users should keep two points in mind: 1) data are displayed in considerably less detail in the printed reports than in the summary tapes; and 2) small-area data presented on tapes can be combined to create summaries for larger, census-recognized areas when the summaries are omitted from the tape. Furthermore, printed reports do not display data for block groups or zip code areas. Thus, summary tapes contain considerably more data than printed reports both in terms of more detailed subject content and more extensive geographic coverage.

- <u>Microfilm</u>
The contents of the 1st count summary tape are available from the Census Bureau on microfilm. The cost is about $8 per roll of 16mm microfilm with three rolls needed to contain the contents of a complete reel of summary tape. Microfilm frames are essentially an array of numbers organized into lines and columns, the interpretation of which requires documentation. Purchasers of microfilm will be provided with the necessary documentation. Users will also need to consult the Master Enumeration District List for interpreting the codes for political and statistical subdivisions of each state.

- <u>Microfiche</u>
All of the final printed reports issued from the 1970 census of population and housing are also available on

*These processing centers are public or private organizations which perform services on a profit or nonprofit basis for a selected group of data users or the general public. They are in no way "franchised" by the Bureau of the Census and the Bureau neither controls nor certifies their activities.

microfiche. Microfiche is a sheet of microfilm containing multiple microimages in a grid pattern. It contains from 58 to 70 pages of print reduced to approximately 1/20th of the original size. The price is about $.25 per card. Although more compact than printed matter, a special reader-printer (if copies are desired) is necessary in order to use it.

The Census Bureau has established a central office for handling requests for printed reports, summary tapes, microfilm, microfiche, and other data materials not discussed here (such as maps, public use samples, address coding guides, etc.). This office, The Central Users' Service, acts as liaison between users and the appropriate subject matter and processing divisions of the Bureau. The address is:

Central Users' Service
Bureau of the Census
Suitland, Maryland 20233

Those HSAs which anticipate regular and detailed use of census data should obtain the 1970 Census Users' Guide, Part 1, which is available from the Superintendent of Documents at a cost of $2.10.

4. The Availability of Specific Data Items

There are two important considerations in looking at the specific subject matter contained in the 1970 census. The first is the geographic or areal unit for whic specific data items are presented. The boundaries of most health service areas will include subareas which constitute an SMSA as well as subareas which lie outside of an SMSA. Part of the population served by HSAs, therefore, will reside within SMSAs and part will reside outside SMSAs. Whether or not the population served resides inside or outside an SMSA is an important consideration relative to the geographic detail in which population and socioeconomic data are presented in census materials. Population and socioeconomic data are presented for smaller areal sub-units within an SMSA than outside an SMSA. Within an SMSA most data items are presented for areal units as small as census tracts, and on some items, data are presented for areal units as small as blocks. Outside an SMSA the smallest areal sub-unit for whi data are presented varies considerably according to the specific data item.

The other important consideration is the medium in which the data are presented. In general, the specific data items are presented in greater detail in summary tapes than they are in printed reports.

Therefore, the following discussion of the subject and geographic detail of each of the specific items of population and socioeconomic data that are of interest to an HSA (i.e., total population counts and data on age, sex, race, income and education) will be presented in terms of the two important considerations of 1) whether or not the population at risk resides in an SMSA and 2) whether the data are being obtained from printed reports or summary tapes. Specific published reports (see Table 2) and summary tapes (see Table 3) will be indicated where appropriate.

In most instances, both within an SMSA and outside an SMSA, only the smallest areal unit for which data are presented will be named for each type of data product (printed reports and summary tapes). With few exceptions, notably zip code areas, data are available for all areal units above the smallest one specified, and in general, the presentation of data becomes increasingly more detailed as the size of the areal unit increases.

a. Total Population

Total population counts are needed by the HSA in determining the size of the population at risk. They are also needed as denominator data in the calculation of rates and ratios which are used to measure the population's health status, availability of health resources, and their patterns of utilization.

- Printed Reports

  Within SMSA: Total population counts are available for areal units down to the block level /Series HC(3)/.

  Outside SMSA: Total population counts are available for all incorporated places regardless of size, and for unincorporated places of 1,000 or more population /Series PC(1)-A/.

• Summary Tapes

> Within SMSA: Total population counts are available for all the areas included in the printed reports, plus 5-digit zip code areas within an SMSA (5th count).
>
> Outside SMSA: Total population counts are available for all the areas included in the printed reports, plus 3-digit zip code areas (5th count).

As can be seen from the above information for total population counts, summary tapes ... data for more areas than printed reports. Besides more e... geographic coverage of total population counts, summary tapes ... so provide more extensive breakdowns of these counts by age, sex, race, income and education and their crosstabulations.

b. Age and Sex

Age is one of the most important socio-demographic correlates of the need for health services and their utilization. It is a particularly important predictor of physician utilization primarily because of its close association with morbidity.

> The relationship between the volume of physician visits and age is best described by a U-shaped curve. Old people and the very young tend to use more services because the former have a higher prevalence of chronic diseases and the latter a higher prevalence of acute conditions.....ospital admission and/or discharge rates are lower for children than any other age group; the rate is highest in the reproductive years and then declines until age 65 and over where it peaks once again.[4]

Sex is also an important socio-demographic correlate of health services need and utilization because of its relationship to morbidity patterns. The need for health services and their utilization differs for males and females primarily because of women's needs for obstetrical care. "Beginning with the childbearing years, 15-44, and continuing through old age, females consume more physicians' services than males....Hospital admission and/or discharge rates are higher for females than males. If obstetrical admissions are excluded this gap narrows considerably."[5]

Printed Reports

Within SMSA: Population totals by age are available at the block level for the following age categories only: under 18, 18-61, and 62 and over /Series HC(3)/. Population totals by sex, however, are not available for blocks or block groups. Census tracts are the smallest areal unit for which population totals by sex are available. For census tracts, population totals by age and sex are presented in the categories shown below /Series PHC(1)/:

Male, all ages.......
Under 5 years.............
3 and 4 years...........
5 to 9 years..............
5 years.................
6 years.................
10 to 14 years............
14 years................
15 to 19 years............
15 years................
16 years................
17 years................
18 years................
19 years................
20 to 24 years............
20 years................
21 years................
25 to 34 years............
35 to 44 years............
45 to 54 years............
55 to 59 years..........
60 to 64 years.........
65 to 74 years..........
75 years and over.........
Female, all ages.....
Same as above

Outside SMSA: For places of 1,000-2,500 population, totals are presented by age (i.e., percent under 18 years and percent 65 years and over) and sex separately. For places of 2,500 to 10,000 population, data for the following age and sex categories are presented /Series PC(1)-B/

Male, all ages	....
Under 5 years	....
5-9 years	....
10-14 years	....
15-19 years	....
20-24 years	....
25-29 years	....
30-34 years	....
35-39 years	....
40-44 years	....
45-49 years	....
50-54 years	....
55-59 years	....
60-64 years	....
65-69 years	....
70-74 years	....
75 years and over	....
Under 18 years	....
21 years and over	....
65 years and over	....
Female, all ages	.....
Same as above	

Summary Tapes

Within SMSA: Age and sex data on the summary tapes are presented for blocks, block groups, and census tracts. Table 4 contains the age, sex, and race breakdowns presented in each of the first five counts of the summary tapes. As is customary for census data, the larger the areal unit the more detailed the subject matter. For example, in the 2nd count summary tape (census tracts) age totals are presented for each individual year of age (some 101 separate categories) for each sex and race group; whereas in the 3rd count summary tape (blocks), age totals are presented for only 21 categories with no breakdown for race. Because of its detailed age breakdown, the 2nd count summary tape provides the most flexibility in obtaining data for desired age, sex, and race categories. Computer programs can be written which combine the very detailed categories in the 2nd count into any number of age, sex, and race categories.

However, assuming that an HSA is using the census tract as its areal unit of analysis within an SMSA, the information contained on the 4th count summary tape would be more useful to an HSA than the 2nd count summary tape, even though the 2nd count provides more detail. This is because the 2nd count

Table 4

Breakdown of Age by Summary Tape

1st Count Block Groups	2nd Count Census Tracts	3rd Count Blocks	4th Count Census Tracts	5th Count Zip Code Areas
Male:	Total:		Total:	Total:
Under 5 years	Male:	Male:	Male:	Male:
5	Under 1 year	Under 5 years	Under 3	Under 3
6	1 year	5	3-4	3-4
7-9	2 years	6	5	5
10-13	3 years	7-9	6	6
14	.	10-13	7-9	7-9
15	.	14	10-13	10-13
16	.	15	14	14
17	to	16	15	15
18	.	17	16	16
19	.	18-19	17	17
20	.	20	18	18
21	99 years	21	19	19
22-24	100 years and over	22-24	20	20
25-34		25-34	21	21
35-44	Female:	35-44	22-24	22-24
45-54	Repeat Age	45-54	25-29	25-34
55-59		55-59	30-34	35-44
60-61	White:	60-61	35-39	45-54
62-64	Same	62-64	40-44	55-59
65-74		65-74	45-49	60-61
75 and over	Negro:	75 and over	50-54	62-64
	Same as Total		55-59	65-74
Female:		Female:	60-61	75 and over
Repeat Age		Repeat Age	62-64	
			65-69	Female:
			70-74	Repeat Age
			75 and over	
				White:
			Female:	Same as Total
			Repeat Age	
				Negro:
			White:	Same as Total
			Same as Total	
				Spanish Americans:
			Negro:	Same as Total
			Same as Total	
			Spanish Americans:	
			Same as Total	

summary tape only contains demographic information (i.e., "100 percent" data; see Table 1), whereas the 4th count summary tape contains both the 100 percent and sample data (e.g., income and education).

An important alternative to the census tract as the areal unit of analysis within an SMSA is the 5-digit zip code area. Its usefulness as an areal unit for aggregating and analyzing local health data has been demonstrated by the Northeast Ohio Regional Medical Program.[6] As indicated in Table 4, population totals by age, sex, and race are available for 5-digit zip code areas from the 5th count summary tapes.*

Outside SMSA: Population totals by age, sex, and race as shown in Table 4 are available for areal units as small as urban places of 1,000 or more population from the 2nd count summary tape. For urban areas of 2,500 population or more, population totals by age, sex, and race are available from the 4th count summary tape.

c. Race

Race is another important predictor of the volume of health services consumed and the site of physician visits. Whites consume more physician services than nonwhites and their rates of hospital discharges and admissions are greater. "More whites than nonwhites consult with physicians in the home, office, or by telephone, while more nonwhites see physicians in hospital clinics or emergency rooms."[7]

Depending upon the particular population the HSA is responsible for, a more specific breakdown on race than the usual white/nonwhite dichotomy may be required. HSAs located in the southwestern and mid-Atlantic sections of the United States, for example, will require population totals for Spanish American subgroups.

*Data are presented only for those 5-digit areas, or those portions of 5-digit areas, which fall entirely within an SMSA. Thus, if a 5-digit area crosses an SMSA line, the data summaries for that 5-digit area represent only that portion of the population residing within the SMSA.

Comparability of racial categories used in the census and in the various local health data sources will be an important consideration affecting the HSA's selection and use of census materials. The particular racial categories required will determine which of the census data products are most appropriate to obtain.

Printed Reports

Within SMSA: Population totals by race (white and Negro) can be obtained for blocks /Series HC(3)/. Population totals for Spanish Americans are not available at the block or block group level, but are available at the census tract level, but only for those census tracts containing 400 or more Spanish Americans. Population totals by race, by age and sex, are available at the census tract level for whites, Negroes, and Spanish Americans, but only for those census tracts containing 400 or more Negroes or Spanish Americans /Series PHC(1)/. The particular age and sex breakdowns for each race are the same as those presented above by sex. For census tracts not containing the 400 minimum, population totals are available only by race (white and Negro); age breakdowns by race are not available in such tracts.

Outside SMSA: Population totals by race (white, Negro and other) are available for places of 1,000 or more population /Series PC(1)-B/. Population totals by race (white, Negro, Indian, Japanese, Chinese, Filipino, all other) by sex are available for places of 2,500 to 10,000 population /Series PC(1)-B/.

Age and sex breakdowns by race (white and Negro) are available for places of 10,000 or more population. The age breakdowns are available in 5 year intervals beginning with under 5 years through 85 years and over. Age breakdowns are also available for single years under 21 years /Series PC(1)-B/.

Summary Tapes*

Within SMSA: Population totals by race (white, Negro and other) are available at the block level from 3rd count summary tapes. At the block group (1st count) population totals are available for racial categories including whites, Negroes, Indians, Orientals,** and others. The most complete categorization by race is included in 2nd count summary tapes at the census tract level. Population totals are available for racial categories including whites, Negroes, Indians, Japanese, Chinese, Filipinos, Hawaiians, Koreans, and others. Population totals for these same racial categories are also broken down by sex. Fifth count summary tapes provide population totals for 5-digit zip code areas within SMSA's for the following racial categories: whites, Negroes, Indians, Japanese, Chinese, Filipinos and others. Population totals for each central city of an SMSA for whites, Negroes, and Spanish Americans are provided in the 6th count summary tapes.

Age and sex breakdowns by race (white, Negro, Spanish American and others) are available at the census tract level only (see Table 4, 4th count). Thus, with the use of a 4th count summary tape an HSA will have the most complete and useful data on race for each census tract within an SMSA.

Outside SMSA: Population totals are available for urban places of 1,000 or more population for the racial categories discussed above in the 2nd count summary tape. For the racial categories discussed above in the 5th count, population totals are available for 3-digit zip code areas. The detailed breakdown of race by age and by sex available for each census tract within an SMSA from 4th count summary tapes is also available outside an SMSA for urban places of 2,500 or more population.

---

*All summary tape data by race are subject to stringent standards of data confidentiality. Certain summary tape tabulations, or parts thereof, by race, may be suppressed for those areal units where the population of specific racial groups is such that confidentiality might be violated if total counts were presented. For a more complete discussion of the issue of data confidentiality in the summary tapes, see the 1970 Census Users' Guide, Part II.

**Includes Japanese, Chinese, Filipinos, Hawaiians, and Koreans. In Alaska, Hawaiian and Korean are substituted for by Aleut and Eskimo, respectively.

d. Income

The utilization of a particular community's health services depends, in part, upon the income differentials among the community's population subgroups. In general, the higher the income, the greater the volume of total health services consumed, though new methods of financing medical care for the poor seem to be narrowing the gaps somewhat. For example, hospital admission rates, once determined to a large extent by income, are now approaching equality for all income groups as a result of the growth of health insurance and financing programs for the poor.[8] The utilization of other types of health care services, however, remains strongly determined by income. High income groups use more specialists, and more preventive and dental services.

Unlike the demographic data on age, sex and race, data on income are seldom recorded by the providers of health services. Thus, income data are rarely available from local health data sources, and, therefore, one of the most important contributions of census data to an HSA's health information system is the great wealth of socioeconomic information that they provide. This information can be linked to locally generated health services utilization data through a technique known as ADMATCH, which will be discussed later in this chapter.

Data on income are quite extensively presented in the census reports. Not only are the total 1969 income figures presented for all families and unrelated individuals, but figures are also available for types of income, ratio of family income to poverty level, etc. Which of these various types of income data will be most useful will depend upon each HSA's specific needs. The following discussion will focus only upon the presentation of total counts of families by income group.

·Printed Reports

Within SMSA: Total family counts by income are available at the census tract level. The particular income breakdowns used in the census tract printed reports /Series PHC(1)/ follows:

Number of Families by Income in 1969

All families.......
Less than $1,000..............
$1,000 to 1,999...............
$2,000 to 2,999...............
$3,000 to 3,999...............
$4,000 to 4,999...............
$5,000 to 5,999...............
$6,000 to 6,999...............
$7,000 to 7,999...............
$8,000 to 8,999...............
$9,000 to 9,999...............
$10,000 to 11,999.............
$12,000 to 14,999.............
$15,000 to 24,999.............
$25,000 to 49,999.............
$50,000 or more...............
Median income.................
Mean income...................

Total counts of families by income group by race (white, Negro and Spanish American) are also available at the census tract level, but only for those census tracts with 400 or more Negroes or Spanish Americans [Series PHC(1)]. As indicated below, the income breakdowns by race, while not as complete as those used for all families, are nevertheless relatively detailed.

Number of Families by Income in 1969 by Race
(Negro and Spanish Americans)

Negro:

All families.......
Less than $1,000............
$1,000 to 1,999.............
$2,000 to 2,999.............
$3,000 to 3,999.............
$4,000 to 4,999.............
$5,000 to 5,999.............
$6,000 to 6,999.............
$7,000 to 7,999.............
$8,000 to 8,999.............
$9,000 to 9,999.............
$10,000 or more.............
Median income...............

Spanish Americans:

Same income breakdown as above

Outside SMSA: Total counts of families by income are available in urban places of 2,500 or more population /Series PC(1)-C/. The income breakdowns are the same as those presented above for census tracts.

Total counts of families by income by race (white, Negro, and Spanish American) are available for urban places of 10,000 or more population with 400 or more Negro or Spanish American population /Series PC(1)-C/.

Summary Tapes

Within SMSA: Total counts of families by income are available from 4th, 5th, and 6th count summary tapes. For census tract and 5-digit zip code areas the income breakdowns are essentially the same as those which appear in the census tract printed reports.

Total counts of families by income by race (white, Negro, and Spanish American) are available at the census tract level (4th count). The breakdowns of income by race are the same as those presented in the printed reports.

Outside SMSA: Total counts of families by income are also available from 4th, 5th and 6th count summary tapes. The smallest areal unit of data presentation is urban places of 2,500 or more population (4th count). The income categories presented are the same as those presented in the printed reports.

Total counts of families by income by race (white, Negro, and Spanish American) are presented for urban places of 2,500 or more population (4th count).

e. Education

Education is another important social determinant of overall health services utilization, especially preventive health services. The better educated consume more physician and dental services. Hospital admission rates are higher and average lengths of stay are shorter for the better educated.[9]

Like income, data on education are also extensively presented in census reports. However, from the standpoint of health planning, the usual need for education data is for population totals of the adult population 25 years of age or older by level of educational attainment.

Printed Reports

Within SMSA: Population totals by education are available at the census tract level /Series PHC (1)/. The breakdown by education presented in the printed census tract reports is as follows:

Years of School Completed

Persons, 25 Years Old and Over	....
No school years completed	....
Elementary 1 to 4 years	....
5 to 7 years	....
8 years	....
High School 1 to 3 years	....
4 years	....
College 1 to 3 years	....
4 years or more	....
Median school years completed	....
Percent high school graduates	....

Population totals by education by race (white, Negro, and Spanish American) are also available at the census tract level, but only for those census tracts with 400 or more Negroes or Spanish Americans /Series PHC(1)/. The specific education breakdowns by race are the same as above.

Outside SMSA: Population totals by education ar available for urban places of 2,500 or more population /Series PC(1)-C/. The education breakdowns are the same as above, except that they do not include "percent high school graduates." Total counts by education by race (white, Negro, and Spanish American) by sex are available for urban places of 10,000 or more population /Series PC(1)-C/. The education breakdowns are the same as above.

Summary Tapes

Within SMSA: Population totals by education are available from 4th, 5th and 6th count summary tapes. For each census tract, (4th count) the education breakdowns are those presented below:

Population 25 Years Old and Over
by Years of School Completed and Sex

Male:
No school years completed (includes
nursery and kindergarten)
Elementary 1 to 4 years............
5 to 6 years............
7 years............
8 years............
High School 1 to 3 years...........
4 years...........
College 1 to 3 years...........
4 years...........
5 years or more...

Female:
Same education breakdowns as above

For each 5-digit zip code area within an SMSA (5th count) the education breakdowns are those presented below:

Population 25 Years Old and Over
by Years of School Completed and Age

25-44 years old:
No school years completed (includes
nursery and kindergarten)
Elementary 1 to 7 years............
8 years............
High School 1 to 3 years...........
4 years...........
College 1 to 3 years...........
4 years or more...

45-54 years old:
Repeat Years of School Completed

55 years old and over:
Repeat Years of School Completed

Compared with the breakdowns presented in printed census tract reports /Series PHC(1)/, the education breakdowns presented in 4th count summary tapes are more detailed. Also, and perhaps more important, population totals by education contained in 4th count summary tapes are also broken down by sex, which is not done at the census tract level in printed reports. Furthermore, the breakdowns indicated above are also available from the 4th count summary tape by race (white, Negro, and Spanish American).

Outside SMSA: Population totals by education by sex and race are available from the 4th count summary tape for places of 2,500 or more population in the same formats as those presented above for census tracts.

The particular age, sex, race, income and education categories needed by an HSA will determine which of the above census data products would be most suitable to obtain. Of all the various reports and tapes that are available, however, the 4th count summary tape will probably satisfy the needs of most HSAs for socio-demographic data from the 1970 census.

Updates of certain of the data items described above are available from selected series of the Census Bureau's Current Population Reports. For example, yearly population estimates (i.e., total population counts) are available for selected areas in each state from Series P-25 and P-26.

C. DATA ANALYSIS AND USE

One of the most important uses of census data is in describing the size and composition of the population within an HSA's planning area. Knowing how many persons reside in a planning area and something about their demographic characteristics (such as age, sex and race) and their socioeconomic characteristics (such as income and education) is, of course, absolutely essential for health planning, resources development and project review. Other important uses of census data include their use as denominator data in the calculation of rates, their use in survey sampling, and their use in conjunction with local health data sources that have been geocoded through the application of the ADMATCH program.

1. Denominator Data

Census data are importantly used as denominator data in the calculation of various ratios and rates. These ratios and rates will be used by an HSA in assessing the health status of its community, the health resources that are available in the community and the community's need for and utilization of these resources. In assessing the health status of its community, several types of mortality, natality and morbidity rates can be used by an HSA. Certain mortality and natality rates are also needed by an HSA in making population estimates and projections. Examples of pertinent health status and demographic information that can be obtained by combining numerator data from a community's vital records with denominator data from the census are presented in Table 5.[10]

An HSA will also need certain types of ratios which require census data in order to assess the availability in its planning area of various types of health resources. For example, the availability of physicians may be assessed on the basis of a ratio, the number of physicians per 100,000 persons per physician.

Table 5

Examples of Pertinent Health and Demographic Information to be Derive
Numerator Data from Vital Records with Denominator Data from th

Type of Statistic	Numerator Information Available from Most State and Local Registration Systems	Denominator Informati Available from Census Data
Demographic (population composition and distribution)	1. Births (by characteristics of parents) 2. Deaths (by age, sex, race, etc.)	1. Age groups 2. Sex 3. Race 4. In and out migrat
Natality	1. Births a. Legitimate or illegitimate b. Attendant c. Institution d. Plurality e. Detailed characteristics of child and parents	1. Marital status of females 2. Age groups 3. Sex 4. Race 5. Education 6. Socioeconomic characteristics
Mortality	1. All deaths a. Causes by age group, sex, and race b. County and city allocations c. Occupation and industry 2. Maternal deaths 3. Infant deaths 4. Fetal deaths	1. Age groups 2. Sex 3. Race 4. Occupation and industry 5. Marital status
Morbidity	1. Reported disease cases a. Cancer b. Venereal disease c. Tuberculosis d. Contagious diseases	1. Age groups 2. Sex 3. Race 4. Marital status 5. Occupation
Social	1. Marriages (age, race, previous marital status) 2. Divorces (age, race, number times married)	1. Marital status 2. Sex 3. Education

Table 5

f Pertinent Health and Demographic Information to be Derived by Combining rator Data from Vital Records with Denominator Data from the Census

Numerator Information Available from Most State and Local Registration Systems	Denominator Information Available from Census Data	Health and Population Information Basic to Local and State Health Program Planning and Administration
1. Births (by characteristics of parents) 2. Deaths (by age, sex, race, etc.)	1. Age groups 2. Sex 3. Race 4. In and out migrations	1. Population projections for State and by local areas 2. Population distribution variations by detailed characteristics and by detailed geographic areas
1. Births a. Legitimate or illegitimate b. Attendant c. Institution d. Plurality e. Detailed characteristics of child and parents	1. Marital status of females 2. Age groups 3. Sex 4. Race 5. Education 6. Socioeconomic characteristics	1. Illegitimacy rates 2. Fertility rates 3. Age/race specific birth rates 4. Birth rates by socio-economic strata 5. Birth registration completeness tests
1. All deaths a. Causes by age group, sex, and race b. County and city allocations c. Occupation and industry 2. Maternal deaths 3. Infant deaths 4. Fetal deaths	1. Age groups 2. Sex 3. Race 4. Occupation and industry 5. Marital status	1. Age-adjusted death rates 2. Age/race/sex specific death rates 3. Cause specific death rates 4. Education specific perinatal death rates 5. Occupation specific death rates 6. Life tables 7. Marital status/cause specific death rates 8. Infant deaths by socio-economic strata
1. Reported disease cases a. Cancer b. Venereal disease c. Tuberculosis d. Contagious diseases	1. Age groups 2. Sex 3. Race 4. Marital status 5. Occupation	1. Age-adjusted morbidity rates 2. Specific morbidity rates by age/race/sex 3. Population at risk 4. Prevalence rates
1. Marriages (age, race, previous marital status) 2. Divorces (age, race, number times married)	1. Marital status 2. Sex 3. Education	1. Specific marriage rates by age and race 2. Specific divorce rates by age and race 3. Relative risk rates for marriages and divorces 4. Marriage and divorce rates by socioeconomic strata

In assessing the community's need for and utilization of health resources, an HSA will again need rates which require census data in their calculation. For example, a community's "need" for health services may be assessed on the basis of a disability indicator, which is conventionally presented in the form of a rate of the number of disability days per _person_ per unit of time. A community's health services utilization may be assessed on the basis of hospital and physician utilization indicators, which are also conventionally presented in the form of rates, e.g., the number of hospital admissions per 1,000 _persons_ and the number of physician visits per _person_ per unit of time respectively.

The use of census data as denominator data in the calculation of these various rates and ratios is relatively simple and straightforward. A specific illustration of the use of census data in the denominator of hospital utilization rates within Central HSA is presented below.*

Estimating Central HSA's Hospital Utilization

Problem: the staff of Central HSA is interested in the current (1975) rate of utilization of Central HSA's community hospitals and in the pattern of utilization since 1970.

Available Data Sources: the 1970 through 1975 editions of the American Hospital Association's _Guide to the Health Care Field_; the _1970 Census of Population: Number of Inhabitants_, Series PC(1)-A; and 1971 through 1975 population estimates of counties from _Current Population Reports_, Series P-26.

Solution: Step 1: From the AHA's _Guide to the Health Care Field_ list and sum the number of admissions

*Other illustrations of the calculation of certain rates can be found in the _Guide to Data for Health Systems Planners_. This guide was prepared by the Center for Census Use Studies, Bureau of the Census, and is available from the Bureau of Health Planning and Resources Development, HRA, DHEW.

to each community hospital in Central HSA for each year from 1970 to 1975.

Hospital	Number of Admissions (000's) 1970	1971	1972	1973	1974	1975
#1	18.0	→				
#2	3.5	→				
#3	6.9	→				
#4	4.0	→				
#5	6.7	→				
↓						
#38	2.3	→				
Total	260.5	278.5	291.6	320.1	345.3	361.5

Step 2: From the 1970 census data and the population estimates for 1971 through 1975 list and sum the number of residents in each county in Central HSA.

County	Population (000's) 1970	1971	1972	1973	1974	1975
#1	138.5	→				
#2	30.8	→				
#3	54.5	→				
#4	35.1	→				
#5	53.9	→				
↓						
#30	20.9	→				
Total	2083.9	2088.0	2097.6	2109.2	2133.1	2151.4

Step 3: Calculate the number of hospital admissions per 1,000 residents in Central HSA for each year by dividing the total number of admissions by the total number of residents and multiplying by 1,000.

example (1970):

$$\frac{260.5}{2,083.9} \times 1,000 = 125.0$$ hospital admissions per 1,000 persons

Step 4: After having completed the calculations in Step 3 for each year from 1970 to 1975, the resulting rates can be displayed as follows:

Hospital Utilization in Central HSA
1970-1975

Year	Hospital Admissions per 1,000 Residents
1970	125.0
1971	132.1
1972	139.0
1973	151.8
1974	161.9
1975	168.0

2. Sampling Applications

Among the sampling applications that can be made of census data are: 1) the use of census data as a check on the representativeness of survey samples; and 2) the use of census data in obtaining stratified survey samples.*

*For further information concerning the sampling applications of census materials, see Margaret Gurney, Sampling Applications of the 1970 Census Publications, Maps, and Public Use Summary Files. U.S. Bureau of the Census (Technical Report No. 27), U.S. Government Printing Office, Washington, D.C., May, 1972.

a. Representativeness of Survey Samples

Census data can be used as a check on the representativeness of the sample of persons included in a local health survey, or for that matter, any source of data that purports to provide information about the population in a given area based upon data that pertain only to a sample of that population. An HSA should be assured of the representativeness of survey samples and other data sources before attempting to generalize their results to the entire targeted population.

Comparing the distributions of various socio-demographic variables in both the sample and the population will provide evidence of whether or not the sample is representative. An example of the factors that could be used in checking the representativeness of a particular sample using census data as the criteria is presented in Table 6. Table 6 compares data from a national health survey conducted by the National Opinion Research Center and the Center for Health Administration Studies* with the then current estimates of the U.S. population.[11] A comparison analogous to that presented in Table 6 can be made using data from a local health survey along with appropriate census data.

b. Obtaining a Stratified Survey Sample

In most instances a simple random sample (i.e., a selection process which insures that each person in the population has an equal chance of being included in the sample) will satisfy an HSA's need for a representative survey sample. However, such samples are often difficult and expensive to obtain. It is more likely that an HSA will utilize stratified sampling procedures. Such stratification may also be required by study objectives. For example, if an 'SA wanted to assess its community's need for geriatric health services and institutions for the aged, it might want a sample that includes a greater proportion of those persons aged 65 and over than exists within the overall population. In most communities those aged 65 and over

*For information concerning CHAS publications, write to: Center for Health Administration Studies, University of Chicago, 5720 Woodlawn Avenue, Chicago, Illinois 60637.

Table 6

Comparison of Demographic Characteristics of NORC Sample, February, 1964, with Census Estimates of the U.S. Population

Characteristic of Population	Percent		
	NORC	Census [1]	
Sex			
Male	49	49[2]	(March, 1964)
Female	51	51	
Age			
0-4	10	11[3]	(July, 1963)
5-14	21	21	
15-19	9	8	
20-24	6	6	
25-34	12	12	
35-44	13	13	
45-54	11	11	
55-64	8	9	
65-74	6	6	
75 and over	3	3	
NA	1	-	
Region			
Northeast	20	25[4]	(March, 1964)
North Central	29	28	
South	33	30	
West	18	16	
Residence			
Urban	67	71[5]	(April, 1963)
Rural non-farm	23	22	
Rural farm	10	7	
Color			
White	85	88[6]	(March, 1964)
Non-white	15	12	
Marital Status (14 years and over)			
Single	23	23[7]	(March, 1964)
Married	65	64	
Widowed, divorced, separated	12	12	
School years completed			
0-4	5	7[8]	(March, 1964)
5-8	27	27	
9-11	19	18	
12	30	30	
13-15	10	9	
16 and over	9	9	

Characteristic of Population	Percent		
	NORC	Census [1]	
Labor force participation (14 years and over)			
Yes	52	54 [9]	(March, 1964)
No	48	46	
Major occupation group (14 years and over)			
Professional	12	13 [10]	(March, 1964)
Clerical	15	16	
Sales	6	7	
Craftsmen	15	12	
Operatives	16	19	
Service	12	13	
Non-farm laborer	8	5	
Farmers and managers	3	3	
Farm workers and foremen	2	2	
Family income before taxes			
Under $1,000	4	8 [11]	(1963)
1,000-1,999	9	10	
2,000-3,499	14	13	
3,500-4,999	13	13	
5,000-7,499	25	23	
7,500-9,999	15	15	
10,000-12,499	10	8	
12,500 and over	10	9	
Number of persons in household			
1	16	16 [12]	(March, 1964)
2	26	27	
3	18	17	
4	17	17	
5	11	11	
6	6	6	
7 and over	6	6	

Notes—Table 6

[1]Unless otherwise indicated these estimates exclude (or have been adjusted to exclude): 1) the institutionalized population as defined in U.S. Census of Population: 1960, Inmates of Institutions, Final Report PC(2)-8A; 2) members of the armed forces except those living off post or on post with their families.

[2]Current Population Reports, P-20, no. 142, derived from Table B.

[3]Ibid., P-25, no. 276. This estimate excludes all members of the armed forces.

[4]Ibid., P-20, no. 142, derived from Table A.

[5]Current Population Reports, P-27, no. 34, and U.S. Census of Population: 1960, vol. I.

[6]Current Population Reports, P-20, no. 142.

[7]Ibid., P-20, no. 135.

[8]Ibid., P-20, no. 138.

[9]Ibid., P-20, no. 141, derived from Table 8.

[10]Ibid., P-20, no. 141, derived from Table 9.

[11]Ibid., P-60, no. 43, Table 1. Based on income distribution for families and unrelated individuals.

[12]Ibid., P-20, no. 139. Based on household size distribution including heads and all related persons.

constitute about 10 percent of the total population. Hence, in a random sample of 1500 persons, about 150 will be aged 65 and over. If the aged of a community are the focus of a special study, a larger number than this will be required for detailed analysis, so a proportion greater than 10 percent may be warranted in the survey sample.

Census data can be used by an HSA to identify those areas within its community which have large concentrations of persons aged 65 and over. For example, for communities which are SMSA's, an HSA would be able to obtain census data on age (i.e., percent over age 61) for geographic areas as small as the city block. It would be possible to divide the community into two strata, one stratum containing all the city blocks with the proportion of persons aged 62 and over at or below the 10 percent norm, and the other stratum containing all the blocks with the proportion of persons aged 62 and over greater than the 10 percent norm. Through random sampling within a disproportionate number of these latter blocks, an HSA could increase the probability of obtaining an overall sample with a proportion of older persons greater than that contained in the community as a whole.

A particular community may be divided into any number of strata based upon a single criterion or on a combination of two or more criteria. A simple random sample can then be taken from each stratum, and the subsamples are then joined to form the total sample. The joining of disproportionate subsamples is not always a simple process, however.

3. The ADMATCH Program and the Uses of Geocoded Data

Many types of health data are being made available, at least for urban areas, by advances in the development of health data systems. Vital statistics, i.e., mortality and natality data, have always been and continue to be the most easily accessible health related information. Information describing the use of health services in geographically defined populations is becoming increasingly more accessible to health planners from family health surveys, hospital discharge abstracting systems, ambulatory care studies and the like. Improvements in the collection of health manpower and facility data are providing the necessary information regarding community health resources. To make more efficient use of the data being derived from these health data components, HSAs will need a method for interrelating these data and for keying them to characteristics of defined populations.

The Bureau of the Census has developed several computerized techniques, which are designed to be efficient, easy-to-use methods of interrelating data derived from two or more separate record files.* One such technique, called ADMATCH, is an address matching system capable of geocoding computer-readable records containing street addresses. With ADMATCH, local records are matched to a geographic reference file. Any geographic codes (e.g., zip codes, census tract, etc.) which are on the reference file will be attached to each individual local record. The process of attaching geographic codes is known as <u>geocoding</u>.

Prior to the development of ADMATCH, geocoding was done manually by a clerk. The clerk would read the address on each record, find a corresponding address in some master street index (such as a post-office zip code directory, a census tract coding guide, etc.), and then transfer the appropriate geocode onto the record. For large numbers of records, manual geocoding can be expensive, time consuming, and often inaccurate.

Geocoding is not a new process. The possibility of using the computerized geocoding process available with ADMATCH, however, is an innovation. By replacing the clerk and thus eliminating most of the manual operation, ADMATCH is intended to provide a faster, and potentially more accurate, system for geocoding records from a variety of sources.

---

*For a description of these techniques, see the <u>Guide to Data for Health Systems Planners</u>, <u>op</u>. <u>cit</u>., Appendix I.

Experimental uses of ADMATCH by the Census Use Study demonstrated its appropriateness as a method for interrelating the health data contained in diverse health data components.[12] It was further demonstrated that with the use of 1970 census data, ADMATCH provided the method needed for keying community health-related data to the socio-demographic and economic characteristics of targeted populations.

The use of ADMATCH will provide HSAs with a geocoded local health data file. While the uses of these geocoded health data files are quite varied, it is important to recognize that ADMATCH provides nothing more than a geocoding function. This function is of no small consequence however, especially in light of the widespread interest in small-area data analysis. Aggregation and statistical analyses of the resultant geocoded data are not part of the ADMATCH program.

To some extent the uses of geocoded health data will be dependent upon the particular subject matter contained in the local health data file. The uses of geocoded birth and death records, for example, will differ in some ways from the uses of geocoded ambulatory encounter or hospital abstract data. There are, however, some common uses that can be made of local health data that have been geocoded.

a. The Identification of Populations-at-Risk and the Location of Health Facilities

One of the most important characteristics of geocoded data is their aggregation to small geographic areas. For communities constituting an SMSA, it is possible to aggregate geocoded health data to census tracts or blocks.* If the geocoded health data were hospital discharge abstracts, for example, the number of persons in each census tract who were hospitalized during a given period of time could be

*The user is not restricted to these areas only. It is possible to aggregate geocoded data to other special areas, such as the health districts within a community, by including the appropriate codes in the GBF before ADMATCHing.

determined. If the data were death records, the number of deaths in each census tract could be determined.

There are, of course, many more examples depending upon the type of geocoded data. The important point, however, is that geocoded health data can be aggregated to small areas in order to locate population subgroups with special health problems. Having certain health data pinpointed to small areas in a community is an indispensable aid to rational health planning and resources development. For example, an HSA could identify those sub-areas within its service area most in need of prenatal health care by determining from geocoded birth records the number of mothers within each census tract who had little or no prenatal care. By locating specific neighborhoods with high percentages of women who had had inadequate prenatal care, decisions concerning the appropriate placement of new prenatal care clinics within the community could then be made.[13] Without the geocoded birth file, these decisions would have to be made using other, more dubious criteria. Thus, the aggregation of geocoded health data to special areas within a community can facilitate the identification of particular populations-at-risk, which in turn, can aid health planners in making a frequently encountered decision: where to locate a particular health services facility.

b. The Development of Neighborhood Health Profiles

More than one geocoded local data file can be used to pinpoint geographically those neighborhoods within an urban area where there is a significant health risk. Several local data files can be geocoded, and the information contained in each aggregated to the same geographic sub-areas within the community. Health profiles based upon several health indicators (e.g., birth rates, death rates, hospital admission rates, disability rates, physician utilization rates, etc.) can then be created for each neighborhood. The geocoded data would supply the necessary numerator data for the calculation of these rates for each special area. Through a continual updating process these geocoded files would serve as the foundation of a community health information system which could be used to monitor the health services needs and utilization of each neighborhood.

c. Geocoded Data and Computer Graphics

The display of various types of geocoded health data on maps can be especially useful in identifying the service areas of particular health care facilities.* A specific illustration of the use of a DIME file, hospital emergency room records and computer graphics in determining a hospital's service area is presented in Figures 3 and 4.[14] The emergency room services of the seven major hospitals serving the inner city of Buffalo were evaluated to determine their adequacy, efficiency and patterns of utilization. Basic to both the analysis and presentation of the results were computer-drawn maps presenting the percentage of outpatients seen at each hospital from the various geographic areas. Figure 3 illustrates the customary determination of a hospital service area. The shading represents the proportion of Community Hospital's patients who reside in each census tract. By comparison, Figure 4 illustrates the proportion of all patients receiving outpatient care in each census tract who utilize Community Hospital.

On the basis of Figure 3, it might be assumed that access to Community Hospital's emergency room does not constitute a major problem, in that most of its patients reside in census tracts contiguous to that hospital. However, on the basis of Figure 4, it has been shown that, in fact, large percentages of patients in certain census tracts have to travel relatively long distances to utilize Community Hospital's emergency room.

---

*Grid coordinates for the nodes in the DIME file provide the capability for mapping either census data or local health data; see U.S. Bureau of the Census, <u>Census Use Study: Computer Mapping</u>, Report No. 2, Washington, D.C., 1969.

Figure 3. Proportion of Community Hospital's Outpatients Who Reside in Each Census Tract

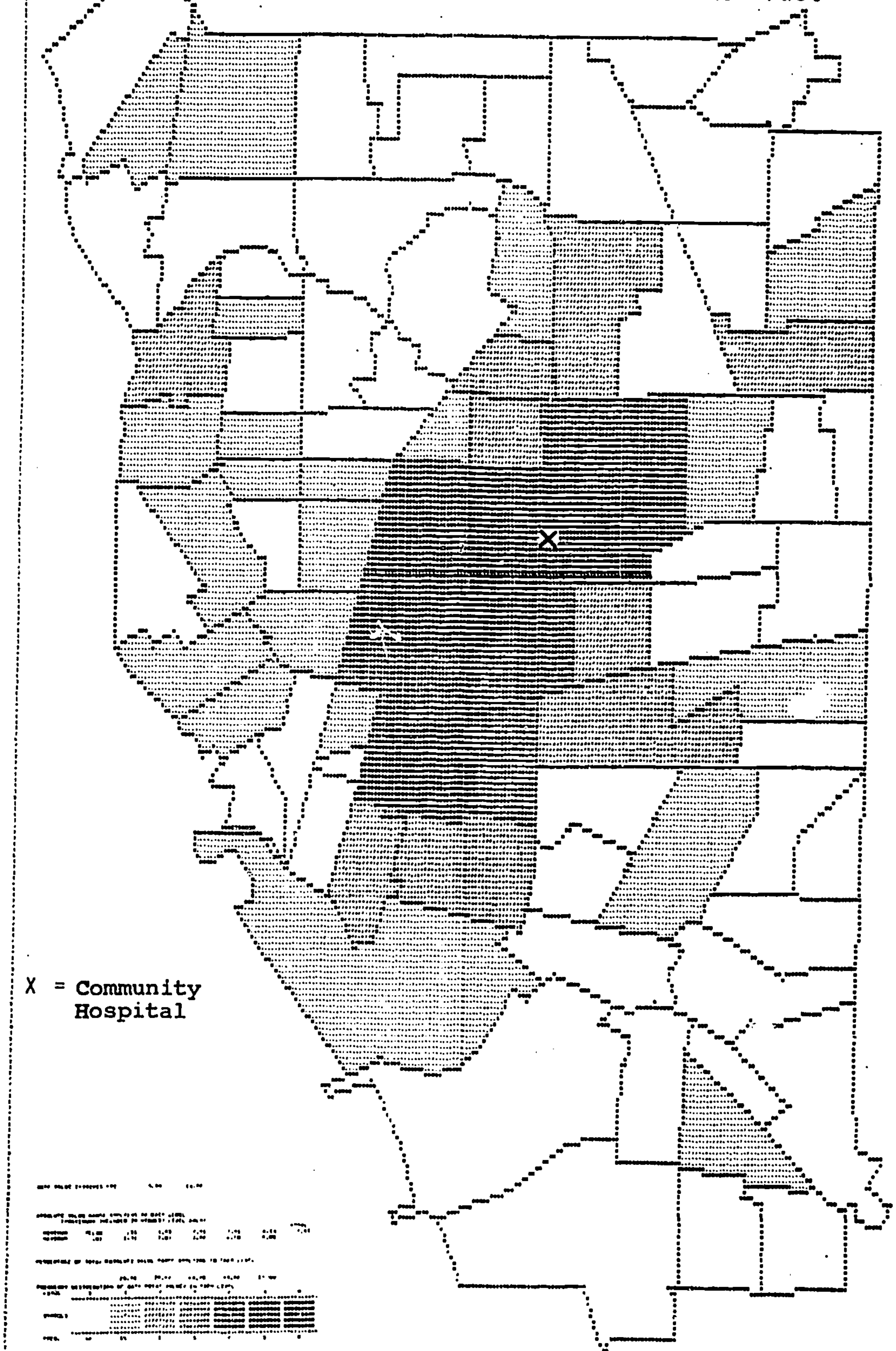
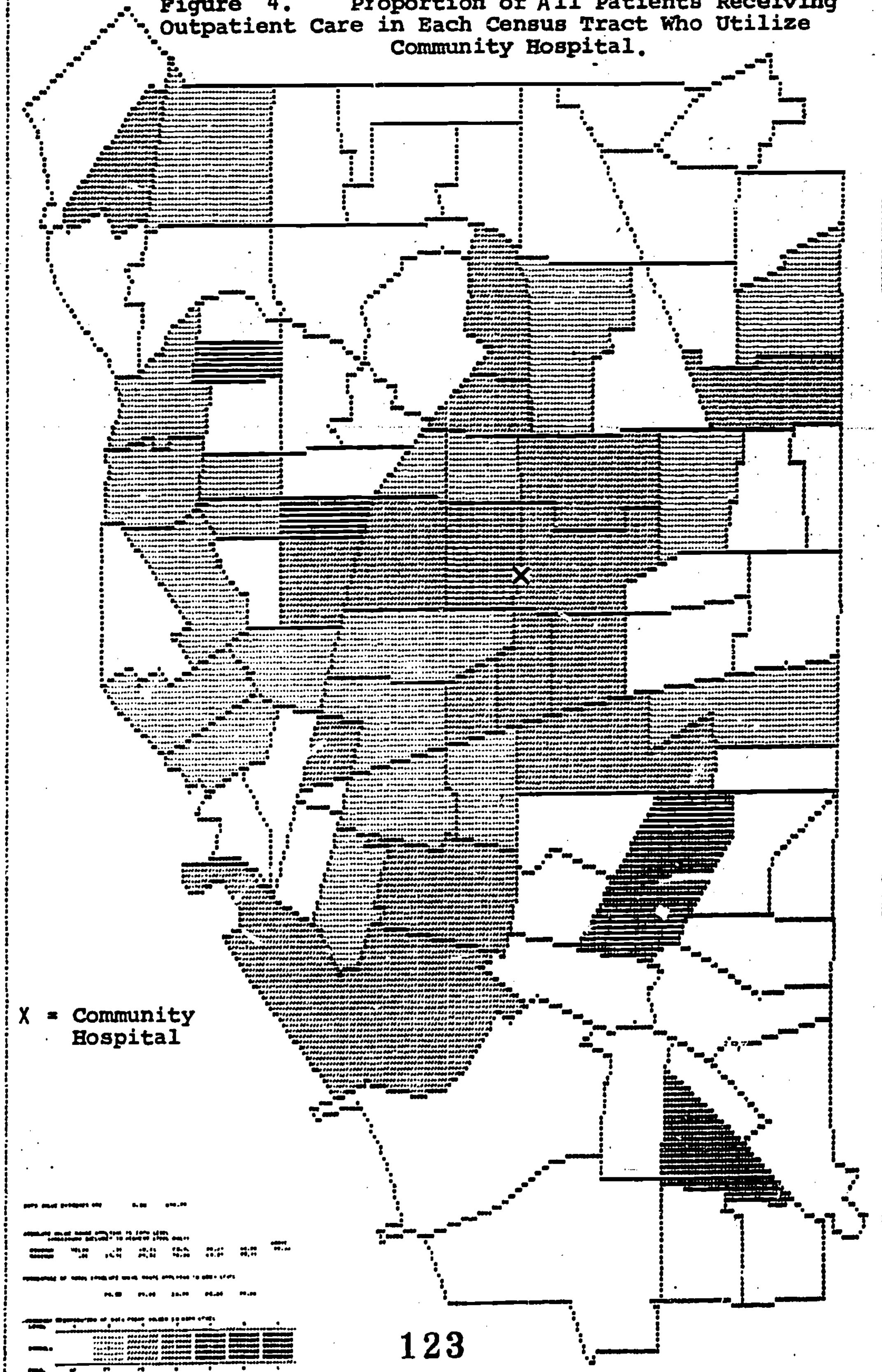


Figure 4. Proportion of All Patients Receiving Outpatient Care in Each Census Tract Who Utilize Community Hospital.



d. The Linkage of Geocoded Health Data with Census Data

A local health data file, if geocoded, can be used in-and-of itself to supply health planners with valuable decision-making information. The analysis of this information would be enhanced, however, if it were linked with census data. While local health data sources contain detailed information on health, they usually do not contain the related socioeconomic information. A geocoded hospital discharge file, for example, can be used to indicate the residential location of hospital users and non-users. It will not, however, provide information about their socioeconomic characteristics. Inferences concerning the characteristics of these population subgroups (the users and non-users), such as their level of education, family income, occupation, employment status, mobility and standard of housing, can be made with the use of 1970 census data.

e. A Case Study of the ADMATCH Program

The staff of the Health Planning Data Project at Purdue University applied the ADMATCH program to a specific health problem in a large metropolitan area in order to assess its effectiveness as a method for interrelating the health, demographic and socioeconomic information of a targeted population. The program was evaluated for its effectiveness in linking information within small geographic areas. The ADMATCH technique was also evaluated in comparison with the technique it was intended to replace, namely manual geocoding. A further discussion of the actual application of the ADMATCH program in this study, including an assessment of its effectiveness relative to manual geocoding is presented in Appendix A.

The first step in the application of the ADMATCH program was the selection of a local data file. Several factors were involved in the selection process. A large midwestern city was selected as the geographic area from which the "local" data file would be derived. This city, which will be referred to as "Central City", had a 1970 population of approximately 750,000 persons.

The next consideration was the character of the required local data file. A local data file was needed that 1) contained local street addresses; 2) was easily accessible; 3) was relatively large and up-to-date; 4) provided information on health-related conditions with a broad distribution throughout Central City; and 5) provided information relative to a specific health problem within the community. In view of these criteria, the vital records file, i.e., birth and infant death certificates, for Central City was selected as the local data file (see Appendix B).

Vital records for the year 1970 were selected as a subfile for analysis. The information contained in these records could be related to data from the decennial census, which also were collected in 1970. The 1970 birth and infant death file for Central City contained approximately 16,000 individual records, each of which contained a street address. The vital records file for Central City was made available to the Purdue staff, with the appropriate confidentiality stipulations, by the State Board of Health. The vital records file contained information on events with a broad distribution throughout the targeted population, i.e., the incidence of childbirth and infant mortality. By relating the 1970 birth and infant death experience of the targeted population (derived from the birth and death certificates) to its demographic and socioeconomic characteristics (derived from 1970 census data), information concerning the social and medical factors associated with infant mortality was obtained.

Among the specific uses made of the geocoded birth file was the identification of medically underserved populations, which in turn resulted in the identification of appropriate sites for future health care facilities. A health planning agency in Central City wanted to pinpoint those areas in Central City containing large proportions of women who had not received adequate prenatal care. This information was needed in order to select the most appropriate locations for future prenatal care clinics.

Two items of information relative to the utilization of prenatal care were included on the birth certificates. These were 1) the trimester in which prenatal care began, and 2) the total number of prenatal care visits. Because of the restrictions on confidentiality, this information could not have been made publicly available on an individual household basis. However, through the process of geocoding, the information was aggregated to the census tract level, and thus, made available for health planning purposes without violating the restrictions on confidentiality. Only aggregated data were made available to the staff of the health planning agency in Central City.

An index of inadequate prenatal care was derived, and it was possible with the geocoded data to identify those areas within Central City which contained significant proportions of women who had not received adequate prenatal care. Along with other information, these data were then used in determining the most appropriate locations for the placement of prenatal care clinics.

Another use of the geocoded birth and death file for 1970 was made possible through linkage of the geocoded file with 1970 census data. This linkage permitted an examination of the interrelationships between various social and medical factors associated with infant mortality. In 1971, the Central City health planning agency had as one of its objectives a 25 percent reduction in infant mortality within two years. The accomplishment of this objective required an understanding of the effects of many factors contributing to infant morbidity and mortality.

Using the geocoded birth and infant death file linked with census data, the Purdue staff conducted an in-depth study of the problems of infant morbidity, as reflected by the number of premature births, and infant mortality in Central City. The purpose of this study was to identify the type and magnitude of effects of several socioeconomic and health-related variables upon prematurity and fetal and infant mortality. Initially, all of the data from the census and vital records used in this study were aggregated to the census tract level. The variables in this study included:

- race refers to the percentage of all infants born to Negro mothers
- median family income
- education refers to the percentage of all persons twenty-five years of age or older who were high school graduates
- age refers to the percentage of all mothers who were below twenty years of age
- marital status refers to the percentage of all infants born illegitimately, i.e., out-of-wedlock
- inadequate prenatal care refers to the percentage of mothers who did not have a minimum of five total visits and/or whose first visit occurred after the first trimester
- prematurity refers to the percentage of all infants with birth weights of less than 2501 grams
- perinatal mortality rate refers to fetal deaths plus deaths occurring during the first week after birth per 1,000 pregnancies (i.e., live births plus fetal deaths)
- postperinatal mortality rate refers to deaths within the remainder of the first year per 1,000 live births

Information on race, age, marital status, prenatal care and birth weight were derived from the geocoded birth records. Information on income and education was derived from 1970 census data. Information concerning perinatal and postperinatal mortality was derived from the geocoded infant and fetal death records. All of the vital record data were aggregated to the census tract level through the use of the ADMATCH program. There were 183 census tracts in Central City.

The bivariate correlations* between the above variables are shown in Table 7. Further statistical analyses of these variables were made through the path analytic technique, and certain inferences were drawn concerning the causal relationships among these variables.[15] There existed, however, a potentially serious problem in analyzing data that had been aggregated to the census tract level, and the fact that these analyses and their resultant causal inferences were made on the basis of aggregated (census tract) data turned out to be an important consideration regarding the planning and resources development decisions of Central City's health planning agency. In the academic literature this problem is labeled the "ecological fallacy," and an HSA that uses the ADMATCH technique should be aware of the problem of the ecological fallacy if it intends to make planning and resources development decisions based upon inferences drawn from geocoded and aggregated data.

The issue of the ecological fallacy is discussed in Appendix C. An example of its effect upon the study under consideration is demonstrated below.

---

*Correlation coefficients are measures of association which are interpreted as indicating the degree or strength of a relationship between variables. While these can refer to a variety of types of relationship, this discussion focuses on those of a bivariate linear nature. Typically, the coefficients are derived from sample data and thus represent estimates of analogous population coefficients. As a result, they can be interpreted as actual measures of association in the sample or as estimates of the analogous associations in the total population. In the latter case, interpretation can include the concepts of confidence intervals, statistical significance of differences, etc.

By convention, statisticians have emphasized measures of association which take unity as an upper limit and either zero or -1.0 as the lower limit. The measures achieve the upper limit (or the lower when it is -1.0) only when the relationship between the two variables is perfect, and they take the value zero when there is no linear relationship between the variables. Thus, a coefficient of .8 represents a relatively strong positive relationship, while one of -.2 suggests a weak negative relationship. Various coefficients are calculated in various ways and each need not be dealt with here. In all cases the calculation involves a summary of the amount of variation in one variable that is accounted for by variation in another variable.

Table 7

Bivariate Correlation Coefficients

	X1	X2	X3	X4	X5	X6	X7
X1	1.00	-.43	-.32	.55	.88	.66	.45
X2		1.00	.85	-.75	-.57	-.66	-.34
X3			1.00	-.79	-.50	-.67	-.32
X4				1.00	.72	.78	.41
X5					1.00	.77	.46
X6						1.00	.49
X7							1.00
X8							
X9							
Mean	21.74	10361.80	50.08	21.62	14.96	43.67	6.70
Standard Deviation	32.84	3315.60	21.34	12.44	14.98	17.85	3.94

X1 percent nonwhite births
X2 median family income
X3 percent adults with high school education
X4 percent mothers under age 20
X5 percent illegimate births

X6 percent births with adequate prenatal c
X7 percent premature b
X8 perinatal mortality
X9 postperinatal morta rate

Table 7

Bivariate Correlation Coefficients

X2	X3	X4	X5	X6	X7	X8	X9
-.43	-.32	.55	.88	.66	.45	.24	.15
1.00	.85	-.75	-.57	-.66	-.34	-.22	-.21
	1.00	-.79	-.50	-.67	-.32	-.21	-.21
		1.00	.72	.78	.41	.33	.11
			1.00	.77	.46	.28	.18
				1.00	.49	.25	.30
					1.00	.20	.25
						1.00	-.06
							1.00
10361.80	50.08	21.62	14.96	43.67	6.70	27.57	13.64
3315.60	21.34	12.44	14.98	17.85	3.94	24.97	19.11

nt nonwhite births
n family income
nt adults with high
l education
nt mothers under age 20
nt illegimate births

X6 percent births with inadequate prenatal care
X7 percent premature births
X8 perinatal mortality rate
X9 postperinatal mortality rate

f. Inferences Drawn from Aggregated Data and the Ecological Fallacy

Among the causal relationships of most interest to Central City's health planning agency was the relationship between the delivery of health services to pregnant women and the health status of their newborn infants. Knowing the extent of this relationship would have important implications regarding the agency's decision to support the development of prenatal care clinics in areas of Central City where prior analyses had revealed large concentrations of pregnant women who had not received adequate prenatal care. Purdue staff, therefore, conducted a more thorough analysis of the relationship between these two factors.

From the original cohort of 15,798 births, which occurred in 1970 in Central City, 300 multiple births and 1,557 single births with missing data on the birth certificates were excluded from analyses. The resulting sample consisted of 13,941 birth certificates.

The independent variable was the mother's prenatal care, and the dependent variable was the infant's birth weight. On the basis of the 13,941 individual birth certificates, individual correlations (fourfold-point correlations)* were computed for the following <u>individual</u> attributes:

- the adequacy of the mother's prenatal care, with "inadequacy" being defined as care which began later than the first trimester of pregnancy or included less than five total visits during the period of pregnancy, and
- the infant's birth weight, with weights of less than 5 lbs. 9 ozs. (2501 grams) defined as premature.

The data from the individual certificates were then aggregated to the census tract and township levels. This was accomplished through the

*For the calculation of fourfold-point correlation (also referred to as "PHI"), see N. M. Downie and R. W. Heath, <u>Basic Statistical Methods</u>, third edition, Harper and Row, New York, 1970, pp. 113-114.

use of the ADMATCH program as previously described. There were a total of 183 census tracts and nine townships located in Central City.

Ecological correlations (Pearsonian product-moment correlations)* were computed on both census tract and township aggregated data for the following group attributes:

- the percent of mothers in each census tract and township with inadequate prenatal care, and
- the percent of births in each census tract and township with premature birth weights.

A comparison of the individual and ecological correlations between inadequate prenatal care and prematurity are summarized below:

	Correlation Coefficients
Individual Correlation	.09
Ecological Correlations:	
Census Tract	.45
Township	.76

Although all of the above correlations, which are significant at the .01 level, indicate a positive relationship between inadequate prenatal care and prematurity, or conversely, adequate prenatal care and maturity, it can be seen that significant differences exist between the individual and either of the ecological correlations.

As indicated by the individual correlation (.09) there is virtually no relationship between prenatal care and birth weight. On the other hand, a moderate relationship between these two factors is indicated by the ecological correlation by census tracts (.45), and a relatively strong relationship is indicated by the ecological correlation by townships (.76). Obviously, depending upon which unit of data analysis

*For the calculation of Pearsonian product-moment coefficients see Herbert M. Blalock, Jr., Social Statistics, second edition, McGraw-Hill, New York, 1972, pp. 376-378.

is utilized, (i.e., individual, or aggregated by census tract or by township) the data can be either mildly suggestive or strongly supportive of a policy decision regarding the development of prenatal care clinics.

With improvement in the health status of its area's newborn infants as its goal, Central City's health planning agency might be very supportive of the development of prenatal care clinics on the basis of the ecological correlations. On the strength of the individual correlations, however, strong support for such development would seem less warranted.

The strength of the relationship between prenatal care and birth weight as reflected in the size of the coefficients in the above correlations is affected by the size of the sub-areas or sub-units to which the data were aggregated. The size of the coefficient for townships, which are larger in size than census tracts, is larger than the coefficient for census tracts.

The results of this study are consistent with those in the literature which show that the size of the correlation coefficient depends to a marked degree upon the size of the sub-areas used in the analyses, and that there is a positive relationship between the size of the coefficients and the average size of the areas for which they are determined.[16] The mathematical relationships between associational measures derived from varying levels of data analysis are quite technical. They have already been adequately summarized elsewhere, and will not be discussed here.[17] The fact that the coefficients of ecological correlations increase in size relative to the size of the sub-areas for which they are calculated is especially important to note, given the tendency of many health planning agencies to collect and analyze data which are aggregated for areas which are even larger than census tracts and townships, e.g., counties.

In summary, the central issue of this analysis is that ecological correlations are not always equal to their corresponding individual correlations. These findings point out that a good deal of caution must be exercised in making inferences about the characteristics of individuals on the basis of aggregated data. It is also apparent from these findings that health planning and resources development decisions should be made on the basis of individual versus ecological correlations whenever relationships between the attributes of individuals are of primary interest. This is not to say, of course, that aggregated data and the ecological correlations that are derived from them have no use in health planning and resources development. As has been shown earlier in this chapter and as will be shown in later chapters, aggregated health-related data have many important uses in health planning.

Appendix A

A Description and Assessment of an Actual Application of the ADMATCH Program

The local data file for this study consisted of approximately 16,000 birth and death certificates in a large midwestern community ("Central City"). Much of the data contained on the birth certificates (see Appendix B) had already been abstracted, keypunched and transferred to magnetic tapes by the State Board of Health. The mother's street address from the birth certificate was not included, however, on the magnetic tape file. The statistical reporting procedures of the State Board of Health did not require the identification of vital events by household residence, and thus, street addresses were not included as part of the computerized data bank on fertility.

For reasons of confidentiality, individual street addresses are customarily omitted from records contained in computerized data storage systems. The absence of street addresses on the computerized vital records file for Central City was believed, therefore, to be typical of most computerized data files in the country. This fact points to one of the major obstacles in using the ADMATCH technique, i.e., the requirement that local data files contain street addresses which are in a computer-readable format and stored on appropriate media (magnetic tapes or punched cards).

The use of the birth and death certificate file in the ADMATCH program required, therefore, that the addresses from these certificates be obtained and manually transcribed from the original certificates onto coding sheets for keypunching onto computer cards. The preparation of the local data file for computer manipulation proved to be a rather cumbersome and time consuming exercise. Most of the birth certificates, for example, were filled out in longhand rather than being typed. Due to illegibility, the addresses on many of the certificates were difficult to read and interpret. In approximately 10 percent of the cases, judgements were required to distinguish between addresses such as Eastridge Dr. and East Ridge Dr., or to determine whether S.E. Street meant S. East Street or Southeast Street. Reference to Central City's street guide was of little help in these cases because of the many possible legitimate variations in address interpretation.

A computer program package on ADMATCH was obtained from the Bureau of the Census. The geographic base file used was an address coding guide (ACG). A 98 percent acceptance level was used in the matcher program, i.e., the matcher program was run with a match attempted only on the street name and house number range. It is possible to use more stringent acceptance levels, for example, matching on primary and secondary street types and street directions.

Of the 15,777 birth records entered into the ADMATCH program, 14,364 (91 percent) were successfully geocoded. Some 579 records were rejected because the house numbers for the given street names did not fall within the accepted number range for those streets in the ACG. Examination of the reject file revealed that over 90 percent of the records rejected on the basis of invalid street names were due to misspellings. For example, it was found that forty rejected records contained the street name TALBOT, which should have been spelled TALBOTT. Misspelled street names also accounted for a significant portion of the records rejected, erroneously, on the basis of invalid house number. Matcher programs can be written to accept variations in the spelling of a street. Subsequent uses of the matcher program could include instructions to change, for example, the street name TALBOT to TALBOTT before attempting to match.

In order to evaluate the effectiveness of the ADMATCH program, the local data file was also manually geocoded. A printout of the local data file (birth and death certificate file) complete with addresses, and a printout of the ACG were used in the manual geocoding exercise. Each address on the local data file was located in the ACG printout and the appropriate census tract number for that address was written beside it. Of the 16,423 birth and death records used in the manual geocoding process, 16,123 (98 percent)were successfully geocoded. Of the 300 unmatched addresses, approximately 275 were eventually geocoded with the use of a metropolitan area map containing census tract numbers and boundaries.

Comparing manual geocoding with ADMATCH on the basis of match rates, manual geocoding proved, in this instance, to be more successful. Manual geocoding also required fewer manhours of labor and was less costly than the computerized ADMATCH approach. Much of the extra time and expense required for ADMATCH, however, was due to technical problems in trying to convert the ADMATCH program, which was written in IBM System/360 assembler language, for use on Purdue's CDC 6500 computer. Also, this was the first attempt on the part of the Purdue staff to use ADMATCH, and because of its unfamiliarity to the staff, a substantial amount of time was required to learn and master the relatively intricate workings of the program so that it could be used correctly. Now that experience has been gained in its use, subsequent uses of the ADMATCH technique should require considerably less computer and programming time.

Appendix B

Birth, Fetal and Death Certificates

TYPE OR PRINT PLAINLY WITH UNFADING INK THIS IS A PERMANENT RECORD

Below for State Office Use

A.______
B.______
C.______
D.______
E.______
F.______
G.______
H.______
I.______
J.______
1.______
2.______
3.______
4.______
5.______
6.______
7.______
8.______

Local No.______

INDIANA STATE BOARD OF HEALTH
CERTIFICATE OF LIVE BIRTH

SAMPLE COPY

CHILD

CHILD—NAME FIRST MIDDLE LAST
1.

SEX
3.

THIS BIRTH— SINGLE, TWIN, TRIPLET, ETC. (SPECIFY)
4a.

IF NOT SINGLE BIRTH—BORN FIRST, SE... THIRD, ETC. (SPECIFY)
4b.

CITY, TOWN, OR LOCATION OF BIRTH
5b.

INSIDE CITY LIMITS (SPECIFY YES OR NO)
5c.

HOSPITAL—NAME
5d.

MOTHER

MOTHER—MAIDEN NAME FIRST MIDDLE LAST
6a.

AGE (AT TIME OF THIS BIRTH)
6b.

RESIDENCE—STATE
7a.

COUNTY
7b.

CITY, TOWN, OR LOCATION
7c.

INSIDE CITY (SPECIFY YES OR NO)
7d.

STREET AND NUMBER OR R.F.D. ZIP
7f.

FATHER

FATHER—NAME FIRST MIDDLE LAST
8a.

AGE (AT TIME OF THIS BIRTH)
8b.

INFORMANT—SIGNATURE
9a.

CERTIFIER

I CERTIFY THAT THE ABOVE NAMED CHILD WAS BORN ALIVE AT THE PLACE AND TIME AND ON THE DATE STATED ABOVE.
10a. SIGNATURE

DATE SIGNED
10b.

PHYSICIAN'S NAME (TYPE OR PRINT)
10d.

PHYSICIAN CODE NO.

MAILING ADDRESS

HEALTH OFFICER—SIGNATURE
11a.

SAMPLE COPY

CONFIDENTIAL INFORMATION FOR MEDICAL AND HEALTH

FATHER ITEM 12-13

RACE—FATHER: WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY)
12.

EDUCATION—SPECIFY HIGHEST GRADE COMPLETED: ELEMENTARY (0, 1, 2, 3, 4, . . . OR 8) | HIGH SCHOOL (1, 2, 3, OR 4) | COLLEGE (1, 2, 3, 4, OR 5+)
13.

ARE NOW
14a.

MOTHER

RACE—MOTHER: WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY)
15.

EDUCATION—SPECIFY HIGHEST GRADE COMPLETED: ELEMENTARY (0, 1, 2, 3, 4, . . . OR 8) | HIGH SCHOOL (1, 2, 3, OR 4) | COLLEGE (1, 2, 3, 4, OR 5+)
16.

DATE MONTH
17a.

DATE LAST NORMAL MENSES BEGAN MONTH DAY YEAR
18.

MONTH OF PREGNANCY PRENATAL CARE BEGAN FIRST, SECOND, THIRD, ETC. (SPECIFY)
19a.

PRENATAL VISITS TOTAL NUMBER (IF NONE, SO STATE)
19b.

WAS A SEROLOGICAL TEST FOR SYPHILIS MADE ON THIS MOTHER? (a) DURING PREGNANCY?______ (STATE YES OR NO)

(b) AT DELIVERY?______ (STATE YES OR NO) (c) IF NO TEST WAS MADE, STATE REASON______

DO NOT RECORD RESULT OF BLOOD TEST ON THIS CERTIFICATE

COMPLICATIONS RELATED TO PREGNANCY (DESCRIBE OR WRITE "NONE")
24.

COMPLICATIONS NOT RELATED TO PREGNANCY (DESCRIBE OR WRITE "NONE")
25.

CONGENITAL MALFORMATIONS
27. YES____ NO____

COMPLICATIONS OF LABOR (DESCRIBE OR WRITE "NONE")
26.

S.B.H.—113-1

INDIANA STATE BOARD OF HEALTH
CERTIFICATE OF LIVE BIRTH

Local No. .... 113 ....

SAMPLE COPY

CHILD	CHILD—NAME FIRST MIDDLE LAST 1.			DATE OF BIRTH (MONTH, DAY, YEAR) 2a.	HOUR 2b. M.
	SEX 3.	THIS BIRTH— SINGLE, TWIN, TRIPLET, ETC. (SPECIFY) 4a.	IF NOT SINGLE BIRTH—BORN FIRST, SECOND, THIRD, ETC. (SPECIFY) 4b.	COUNTY OF BIRTH 5a.	
	CITY, TOWN, OR LOCATION OF BIRTH 5b.	INSIDE CITY LIMITS (SPECIFY YES OR NO) 5c.	HOSPITAL—NAME (IF NOT IN HOSPITAL, GIVE STREET AND NUMBER) 5d.		
MOTHER	MOTHER—MAIDEN NAME FIRST MIDDLE LAST 6a.		AGE (AT TIME OF THIS BIRTH) 6b.	STATE OF BIRTH (IF NOT IN U.S.A., NAME COUNTRY) 6c.	
	RESIDENCE—STATE 7a.	COUNTY 7b.	CITY, TOWN, OR LOCATION 7c.	INSIDE CITY LIMITS (SPECIFY YES OR NO) 7d.	TOWNSHIP 7e.
	STREET AND NUMBER OR R.F.D. ZIP 7f.				
FATHER	FATHER—NAME FIRST MIDDLE LAST 8a.		AGE (AT TIME OF THIS BIRTH) 8b.	STATE OF BIRTH (IF NOT IN U.S.A., NAME COUNTRY) 8c.	
	INFORMANT—SIGNATURE 9a.			RELATION TO CHILD 9b.	
CERTIFIER	I CERTIFY THAT THE ABOVE NAMED CHILD WAS BORN ALIVE AT THE PLACE AND TIME AND ON THE DATE STATED ABOVE. 10a. SIGNATURE		DATE SIGNED (MONTH, DAY, YEAR) 10b.	ATTENDANT—M.D., D.O., MIDWIFE, OTHER (SPECIFY) 10c.	
	PHYSICIAN'S NAME (TYPE OR PRINT) 10d.	PHYSICIAN CODE NO.	MAILING ADDRESS (STREET OR R.F.D. NO., CITY OR TOWN, STATE, ZIP)		
	HEALTH OFFICER—SIGNATURE 11a.			DATE RECEIVED BY LOCAL HEALTH OFFICER MONTH DAY YEAR 11b.	

SAMPLE COPY

CONFIDENTIAL INFORMATION FOR MEDICAL AND HEALTH USE ONLY

FATHER ITEM 12-13	RACE—FATHER: WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY) 12.	EDUCATION—SPECIFY HIGHEST GRADE COMPLETED: ELEMENTARY (0, 1, 2, 3, 4, . . . OR 8) 13.	HIGH SCHOOL (1, 2, 3, OR 4)	COLLEGE (1, 2, 3, 4, OR 5+)	PREVIOUS DELIVERIES—HOW MANY OTHER CHILDREN ARE NOW LIVING 14a.	WERE BORN ALIVE—NOW DEAD 14b.	WERE BORN DEAD (FETAL DEATH AT ANY TIME AFTER CONCEPTION) 14c.
MOTHER	RACE—MOTHER: WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY) 15.	EDUCATION—SPECIFY HIGHEST GRADE COMPLETED: ELEMENTARY (0, 1, 2, 3, 4, . . . OR 8) 16.	HIGH SCHOOL (1, 2, 3, OR 4)	COLLEGE (1, 2, 3, 4, OR 5+)	DATE OF LAST LIVE BIRTH MONTH DAY YEAR 17a.		DATE OF LAST FETAL DEATH MONTH DAY YEAR 17b.
	DATE LAST NORMAL MENSES BEGAN MONTH DAY YEAR 18.	MONTH OF PREGNANCY PRENATAL CARE BEGAN FIRST, SECOND, THIRD, ETC. (SPECIFY) 19a.		PRENATAL VISITS TOTAL NUMBER (IF NONE, SO STATE) 19b.	LEGITIMATE (SPECIFY YES OR NO) 20.		BIRTH WEIGHT ____ LB ____ OZ 21.

WAS A SEROLOGICAL TEST FOR SYPHILIS MADE ON THIS MOTHER? (a) DURING PREGNANCY? ____________ (STATE YES OR NO) DATE OR TRIMESTER ____________

(b) AT DELIVERY? ____________ (STATE YES OR NO) (c) IF NO TEST WAS MADE, STATE REASON ____________

WERE PRECAUTIONS TAKEN AGAINST OPHTHALMIA NEONATORUM? ________

DO NOT RECORD RESULT OF BLOOD TEST ON THIS CERTIFICATE

COMPLICATIONS RELATED TO PREGNANCY (DESCRIBE OR WRITE "NONE")

24.

COMPLICATIONS NOT RELATED TO PREGNANCY (DESCRIBE OR WRITE "NONE")

25.

CONGENITAL MALFORMATIONS OR ANOMALIES OF CHILD

27. YES ______ NO ______ IF YES FILE FORM SDFW 1232

COMPLICATIONS OF LABOR (DESCRIBE OR WRITE "NONE")

26.

S.B.H.—113-1

125

TYPE OR PRINT PLAINLY WITH UNFADING INK
THIS IS A PERMANENT RECORD

Below for State Office Use

A.______
B.______
C.______
D.______
E.______
F.______
G.______
H.______
I.______
J.______
1.______
2.______
3.______
4.______
5.______
6.______
7.______
8.______

Local No.______

INDIANA STATE BOARD OF HEALTH
CERTIFICATE OF FETAL DEATH

TYPE, OR PRINT IN PERMANENT INK
SEE HANDBOOK FOR INSTRUCTIONS

FETUS
FETUS—NAME FIRST MIDDLE LAST
1.
SEX
3.
THIS DELIVERY—SINGLE, TWIN, TRIPLET, ETC. (SPECIFY)
4a.
IF NOT SINGLE DELIVERY BORN F... THIRD, ETC. (SPECIFY)
4b.
CITY, TOWN, OR LOCATION OF DELIVERY
5b.
INSIDE CITY LIMITS (SPECIFY YES OR NO)
5c.
HOSPITAL—NAME
5d.

MOTHER
MOTHER—MAIDEN NAME FIRST MIDDLE LAST
6a.
RESIDENCE—STATE
7a.
COUNTY
7b.
CITY, TOWN, OR LOCATION
7c.

FATHER
FATHER—NAME FIRST MIDDLE
8a.

SAMPLE COPY

9. PART I. FETAL DEATH WAS CAUSED BY: (ENTER ONLY ONE CAUSE PER LINE F...

CAUSE
FETAL OR MATERNAL CONDITION DIRECTLY CAUSING FETAL DEATH
IMMEDIATE CAUSE
(a)
FETAL AND/OR MATERNAL CONDITIONS, IF ANY, GIVING RISE TO THE IMMEDIATE CAUSE (A), STATING THE UNDERLYING CAUSE LAST
DUE TO, OR AS A CONSEQUENCE OF
(b)
DUE TO, OR AS A CONSEQUENCE OF
(c)
PART II. OTHER SIGNIFICANT CONDITIONS OF FETUS OR MOTHER; CONDITIONS CONTRIBUTING TO FETAL DEATH BUT NOT RELATED TO CAUSE GIVEN IN PART I (A)
FETUS DIED BEFORE LABOR, DU... OR DELIVERY, UNKNOWN (SP...
10.

CERTIFIER
I CERTIFY THAT THIS DELIVERY OCCURRED ON THE DATE STATED ABOVE AND THE FETUS WAS BORN DEAD
12a. SIGNATURE
DATE SIGNED (MONT...
12b.
CERTIFIER—MAILING ADDRESS (STREET OR R.F.D. NO., CITY OR TOWN, STATE, ZIP)
12d.

BURIAL
BURIAL, CREMATION, OR REMOVAL (SPECIFY)
14a.
CEMETERY OR CREMATORY—NAME
14b.
DATE (MONTH, DAY, YEAR)
14d.
FUNERAL HOME—NAME AND ADDRESS (STREET OR R.F.D. NO., C...
15a.
FUNERAL DIRECTOR—SIGNATURE
15b.
HEALTH OFFICER—SIGNATURE
16a.

SAMPLE COPY

CONFIDENTIAL INFORMATION FOR MEDICAL AND H...

FATHER
RACE—FATHER
WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY)
17.
EDUCATION—SPECIFY HIGHEST GRADE COMPLETED
ELEMENTARY (0, 1, 2, 3, 4, ... OR 8)
18.
HIGH SCHOOL (1, 2, 3, OR 4)
COLLEGE (1, 2, 3, 4, OR 5+)
AGE
19a.

MOTHER
RACE—MOTHER
WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY)
20.
EDUCATION—SPECIFY HIGHEST GRADE COMPLETED
ELEMENTARY (0, 1, 2, 3, 4, ... OR 8)
21.
HIGH SCHOOL (1, 2, 3, OR 4)
COLLEGE (1, 2, 3, 4, OR 5+)
DAT...
22a.
DATE LAST NORMAL MENSES BEGAN (MONTH, DAY, YEAR)
23.
MONTH OF PREGNANCY PRENATAL CARE BEGAN 1ST, 2ND, ETC. (SPECIFY)
24a.
PRENATAL VISITS TO... IF NONE, SO STATE
24b.

WAS A SEROLOGICAL TEST FOR SYPHILLIS MADE ON THIS MOTHER? (a) DURING PREGNANCY?______ (STAT...
27.
COMPLICATIONS NOT RELATED TO PREGNANCY (DESCRIBE OR WRITE "NONE")
28.
CONGENITAL MAL...
29.
COMPLICATIONS OF LABOR (DESCRIBE OR WRITE "NONE")
30.

SBH 6-24-15
REV. 1968

INDIANA STATE BOARD OF HEALTH

CERTIFICATE OF FETAL DEATH

State File No.

Local No.

TYPE, OR PRINT IN PERMANENT INK SEE HANDBOOK FOR INSTRUCTIONS

FETUS

FETUS—NAME FIRST MIDDLE LAST 1.

DATE OF DELIVERY (MONTH, DAY, YEAR) 2a. HOUR 2b. M.

SEX 3.

THIS DELIVERY—SINGLE, TWIN, TRIPLET, ETC. (SPECIFY) 4a.

IF NOT SINGLE DELIVERY BORN FIRST, SECOND, THIRD, ETC. (SPECIFY) 4b.

COUNTY OF DELIVERY 5a.

CITY, TOWN, OR LOCATION OF DELIVERY 5b.

INSIDE CITY LIMITS (SPECIFY YES OR NO) 5c.

HOSPITAL—NAME (IF NOT HOSPITAL, GIVE STREET AND NUMBER) 5d.

MOTHER

MOTHER—MAIDEN NAME FIRST MIDDLE LAST 6a.

AGE (AT TIME OF THIS DELIVERY) 6b.

STATE OF BIRTH (IF NOT IN U.S.A., NAME COUNTRY) 6c.

RESIDENCE—STATE 7a.

COUNTY 7b.

CITY, TOWN, OR LOCATION 7c.

INSIDE CITY LIMITS (SPECIFY YES OR NO) 7d.

STREET AND NUMBER 7e.

FATHER

FATHER—NAME FIRST MIDDLE LAST 8a.

AGE (AT TIME OF THIS DELIVERY) 8b.

STATE OF BIRTH (IF NOT IN U.S.A., NAME COUNTRY) 8c.

CAUSE

9. PART I. FETAL DEATH WAS CAUSED BY: [ENTER ONLY ONE CAUSE PER LINE FOR (a), (b), AND (c)] SPECIFY FETAL OR MATERNAL

FETAL OR MATERNAL CONDITION DIRECTLY CAUSING FETAL DEATH — IMMEDIATE CAUSE (a)

FETAL AND/OR MATERNAL CONDITIONS, IF ANY, GIVING RISE TO THE IMMEDIATE CAUSE (A), STATING THE UNDERLYING CAUSE LAST — DUE TO, OR AS A CONSEQUENCE OF (b)

DUE TO, OR AS A CONSEQUENCE OF (c)

PART II. OTHER SIGNIFICANT CONDITIONS OF FETUS OR MOTHER: CONDITIONS CONTRIBUTING TO FETAL DEATH BUT NOT RELATED TO CAUSE GIVEN IN PART I (A)

FETUS DIED BEFORE LABOR, DURING LABOR OR DELIVERY, UNKNOWN (SPECIFY) 10.

AUTOPSY (SPECIFY YES OR NO) 11a.

IF YES, WERE FINDINGS CONSIDERED IN DETERMINING CAUSE OF DEATH 11b.

CERTIFIER

I CERTIFY THAT THIS DELIVERY OCCURRED ON THE DATE STATED ABOVE AND THE FETUS WAS BORN DEAD

12a. SIGNATURE

DATE SIGNED (MONTH, DAY, YEAR) 12b.

ATTENDANT—M.D., D.O., MIDWIFE, OTHER (SPECIFY) 12c.

CERTIFIER—MAILING ADDRESS (STREET OR R.F.D. NO., CITY OR TOWN, STATE, ZIP) 12d.

AUTHORIZED OFFICIAL (IF DELIVERY NOT ATTENDED BY PHYSICIAN) 13. SIGNATURE

BURIAL

BURIAL, CREMATION, OR REMOVAL (SPECIFY) 14a.

CEMETERY OR CREMATORY—NAME 14b.

LOCATION (CITY OR TOWN, STATE) 14c.

DATE (MONTH, DAY, YEAR) 14d.

FUNERAL HOME—NAME AND ADDRESS (STREET OR R.F.D. NO., CITY OR TOWN, STATE, ZIP) 15a.

FUNERAL HOME No.

FUNERAL DIRECTOR—SIGNATURE 15b.

HEALTH OFFICER—SIGNATURE 16a.

DATE RECEIVED BY LOCAL REGISTRAR (MONTH, DAY, YEAR) 16b.

SAMPLE COPY

CONFIDENTIAL INFORMATION FOR MEDICAL AND HEALTH USE ONLY

FATHER

RACE—FATHER WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY) 17.

EDUCATION—SPECIFY HIGHEST GRADE COMPLETED: ELEMENTARY (0, 1, 2, 3, 4, ... OR 8) HIGH SCHOOL (1, 2, 3, OR 4) COLLEGE (1, 2, 3, 4, OR 5+) 18.

PREVIOUS DELIVERIES—HOW MANY OTHER CHILDREN: ARE NOW LIVING 19a. WERE BORN ALIVE—NOW DEAD 19b. WERE BORN DEAD (FETAL DEATH AT ANY TIME AFTER CONCEPTION) 19c.

MOTHER

RACE—MOTHER WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY) 20.

EDUCATION—SPECIFY HIGHEST GRADE COMPLETED: ELEMENTARY (0, 1, 2, 3, 4, ... OR 8) HIGH SCHOOL (1, 2, 3, OR 4) COLLEGE (1, 2, 3, 4, OR 5+) 21.

DATE OF LAST LIVE BIRTH MONTH DAY YEAR 22a.

DATE OF LAST FETAL DEATH MONTH DAY YEAR 22b.

DATE LAST NORMAL MENSES BEGAN (MONTH, DAY, YEAR) 23.

MONTH OF PREGNANCY PRENATAL CARE BEGAN (1ST, 2ND, ETC. (SPECIFY) 24a.

PRENATAL VISITS TOT. NO. (IF NONE, SO STATE) 24b.

LEGITIMATE (SPECIFY YES OR NO) 25.

BIRTH WEIGHT 26.

WAS A SEROLOGICAL TEST FOR SYPHILIS MADE ON THIS MOTHER? (a) DURING PREGNANCY? DATE OR TRIMESTER (STATE YES OR NO)

27.

CONGENITAL MALFORMATIONS OR ANOMALIES OF FETUS (DESCRIBE OR WRITE "NONE") 29.

COMPLICATIONS NOT RELATED TO PREGNANCY (DESCRIBE OR WRITE "NONE") 28.

COMPLICATIONS OF LABOR (DESCRIBE OR WRITE "NONE") 30.

SBH 6-24-15
REV. 1968

142

TYPE OR PRINT PLAINLY WITH UNFADING INK

THIS IS A PERMANENT RECORD

Below for State Office Use

A______
B______
C______
D______
E______
F______
G______
H______
I______
J______

1______
2______
3______
4______
5______
6______
7______
8______

Disposition Permit Issued / /

Provisional Certificate ☐ Yes ☐ No

EMBALMER'S NAME..........

LICENSE No..........

FUNERAL DIRECTOR'S LICENSE No..........

INDIANA STATE BOARD OF HEALTH

DIVISION OF VITAL RECORDS

MEDICAL CERTIFICATE OF DEATH

Local No..........

PERMANENT INK SEE HANDBOOK FOR INSTRUCTIONS

SAMPLE COPY

DECEASED	DECEASED—NAME FIRST MIDDLE LAST				SEX
	RACE WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY)	AGE—LAST BIRTHDAY (YEARS) 5a.	UNDER 1 YEAR MOS. DAYS	UNDER 1 DAY HOURS MIN.	DATE OF (MONTH...
	CITY, TOWN, OR LOCATION OF DEATH 7b.	INSIDE CITY LIMITS (SPECIFY YES OR NO) 7c.	HOSPITAL OR OTHER INSTITUTION— 7d.		
	STATE OF BIRTH (IF NOT IN U.S.A., NAME COUNTRY)	CITIZEN OF WHAT COUNTRY 9.	MARRIED, NEVER MARRIED, WIDOWED, DIVORCED (SPECIFY) 10.		
USUAL RESIDENCE WHERE DECEASED LIVED. IF DEATH OCCURRED IN INSTITUTION, GIVE RESIDENCE BEFORE ADMISSION.	SOCIAL SECURITY NUMBER 12.	USUAL OCCUPATION (GIVE KIND OF WORK DONE DURING MOST OF WORKING LIFE, EVEN IF RETIRED) 13a.	KIND OF ... 13b.		
	RESIDENCE—STATE 14a.	COUNTY 14b.	CITY, TOWN OR LOCATION 14c.	INSIDE CITY LIMITS (SPECIFY YES OR...) 14d.	
	STREET AND NUMBER 14f.				
PARENTS	FATHER—NAME FIRST MIDDLE LAST 15.		MOTHER—MAIDEN NAME 16.		
	INFORMANT—NAME 17a.	RELATIONSHIP 17b.	MAILING ADDRESS 17c.		
CAUSE	PART I. DEATH WAS CAUSED BY: (ENTER ONLY ONE CAUSE PER LINE FOR (a), (b), A...				
	18. IMMEDIATE CAUSE (a)				
	CONDITIONS, IF ANY, WHICH GAVE RISE TO IMMEDIATE CAUSE (A), STATING THE UNDERLYING CAUSE LAST	DUE TO, OR AS A CONSEQUENCE OF: (b)			
		DUE TO, OR AS A CONSEQUENCE OF: (c)			
	PART II. OTHER SIGNIFICANT CONDITIONS CONDITIONS CONTRIBUTING TO DEATH BUT NOT RELATED TO CA... GIVEN IN PART I (A)				
	DEATH OCCURRED (HOUR) 20a. M.	THE DECEDENT WAS PRONOUNCED DEAD MONTH DAY YEAR HOUR 20b. M.	DATE... 21a.		
CERTIFIER	CERTIFIER—NAME (TYPE OR PRINT) 22a.	SIGNATURE 22b.			
	MAILING ADDRESS—CERTIFIER STREET OR R.F.D. NO 23.	CITY OR TOWN			
BURIAL	BURIAL, CREMATION, REMOVAL (SPECIFY) 24a.	CEMETERY, CREMATORY, FUNERAL HOME 24b.	LOCATION CI... 24c.		
	DATE (MONTH, DAY, YEAR) 24d.	FUNERAL HOME—NAME AND ADDRESS (STREET OR R.F... 25a.			
	FUNERAL DIRECTOR—SIGNATURE 25b.	HEALTH OFFICER—SIGNATURE 26a.			

SAMPLE COPY

SBH 6-24-2

INDIANA STATE BOARD OF HEALTH

DIVISION OF VITAL RECORDS

MEDICAL CERTIFICATE OF DEATH

Local No.................... State No....................

PERMANENT INK SEE HANDBOOK FOR INSTRUCTIONS

DECEASED

1. DECEASED—NAME FIRST MIDDLE LAST
2. SEX
3. DATE OF DEATH (MONTH, DAY, YEAR)
4. RACE WHITE, NEGRO, AMERICAN INDIAN, ETC. (SPECIFY)
5a. AGE—LAST BIRTHDAY (YEARS)
5b. UNDER 1 YEAR MOS. DAYS
5c. UNDER 1 DAY HOURS MIN.
6. DATE OF BIRTH (MONTH, DAY, YEAR)
7a. COUNTY OF DEATH
7b. CITY, TOWN, OR LOCATION OF DEATH
7c. INSIDE CITY LIMITS (SPECIFY YES OR NO)
7d. HOSPITAL OR OTHER INSTITUTION—NAME (IF NOT IN EITHER, GIVE STREET AND NUMBER)
8. STATE OF BIRTH (IF NOT IN U.S.A., NAME COUNTRY)
9. CITIZEN OF WHAT COUNTRY
10. MARRIED, NEVER MARRIED, WIDOWED, DIVORCED (SPECIFY)
11. SURVIVING SPOUSE (IF WIFE, GIVE MAIDEN NAME)
12. SOCIAL SECURITY NUMBER
13a. USUAL OCCUPATION (GIVE KIND OF WORK DONE DURING MOST OF WORKING LIFE, EVEN IF RETIRED)
13b. KIND OF BUSINESS OR INDUSTRY

USUAL RESIDENCE WHERE DECEASED LIVED. IF DEATH OCCURRED IN INSTITUTION, GIVE RESIDENCE BEFORE ADMISSION.

14a. RESIDENCE—STATE
14b. COUNTY
14c. CITY, TOWN OR LOCATION
14d. INSIDE CITY LIMITS (SPECIFY YES OR NO)
14e. TOWNSHIP
14f. STREET AND NUMBER
14g. IS RESIDENCE ON A FARM? YES ☐ NO ☐

PARENTS

15. FATHER—NAME FIRST MIDDLE LAST
16. MOTHER—MAIDEN NAME FIRST MIDDLE LAST
17a. INFORMANT—NAME
17b. RELATIONSHIP
17c. MAILING ADDRESS (STREET OR R.F.D. NO., CITY OR TOWN, STATE, ZIP)

CAUSE

PART I. DEATH WAS CAUSED BY: (ENTER ONLY ONE CAUSE PER LINE FOR (a), (b), AND (c)) — APPROXIMATE INTERVAL BETWEEN ONSET AND DEATH

18. IMMEDIATE CAUSE

(a)

CONDITIONS, IF ANY, WHICH GAVE RISE TO IMMEDIATE CAUSE (A), STATING THE UNDERLYING CAUSE LAST

DUE TO, OR AS A CONSEQUENCE OF:

(b)

DUE TO, OR AS A CONSEQUENCE OF:

(c)

PART II. OTHER SIGNIFICANT CONDITIONS CONDITIONS CONTRIBUTING TO DEATH BUT NOT RELATED TO CAUSE GIVEN IN PART I (A)

19a. AUTOPSY (YES OR NO)
19b. IF YES WERE FINDINGS CONSIDERED IN DETERMINING CAUSE OF DEATH

20a. DEATH OCCURRED (HOUR) M.
20b. THE DECEDENT WAS PRONOUNCED DEAD MONTH DAY YEAR HOUR M.
21a. DATE SIGNED (MONTH, DAY, YEAR)

CERTIFIER

22a. CERTIFIER—NAME (TYPE OR PRINT)
22b. SIGNATURE (DEGREE OR TITLE)
23. MAILING ADDRESS—CERTIFIER STREET OR R.F.D. NO CITY OR TOWN STATE ZIP

BURIAL

24a. BURIAL, CREMATION, REMOVAL (SPECIFY)
24b. CEMETERY, CREMATORY, FUNERAL HOME
24c. LOCATION CITY OR TOWN STATE
FUNERAL HOME NUMBER
24d. DATE (MONTH, DAY, YEAR)
25a. FUNERAL HOME—NAME AND ADDRESS (STREET OR R.F.D. NO., CITY OR TOWN, STATE, ZIP)
25b. FUNERAL DIRECTOR—SIGNATURE
26a. HEALTH OFFICER—SIGNATURE
26b. DATE RECEIVED BY LOCAL HEALTH OFFICER

SAMPLE COPY

SAMPLE COPY

EMBALMER'S NAME ....

LICENSE No. ....

FUNERAL DIRECTOR'S LICENSE No. ....

SBH 6-24-2

Appendix C

The Ecological Fallacy

In a 1950 article, W. S. Robinson differentiated two types of correlations, _individual_ and _ecological_. According to Robinson,

> In an individual correlation the variables are descriptive properties of individuals, such as height, income, eye color, or race, and not descriptive statistical constants such as rates or means. In an ecological correlation the statistical object is a group of persons. The correlation between the percentage of the population which is Negro and the percentage which is illiterate for the 48 states...is an ecological correlation....The variables are percentages, descriptive properties of groups, and not descriptive properties of individuals.[1]

As stated by Robinson, however, in studies which use ecological correlations, the purpose is to discover something about the behavior of individuals, and this is where the problem arises. Robinson demonstrated that ecological correlations between rates or averages for areas need not, and often do not, accurately represent the characteristics of the individuals within the areas. He states:

> While it is theoretically possible for the two (ecological and individual correlations) to be equal, the conditions under which this can happen are far removed from those ordinarily encountered in data. From a practical standpoint, therefore, the only reasonable assumption is that an ecological correlation is almost certainly not equal to its corresponding individual correlation.[2]

Explanations of the behavior of individuals made on the basis of ecological correlations, therefore, may be in error. Selvin labeled this inferential problem the _ecological fallacy_,[3] which may occur whenever ecological correlations are used in place of individual correlations in explaining the behavior of individuals.

To illustrate that ecological correlations cannot validly be used as substitutes for individual correlations, Robinson points out that the ecological (linear) correlation of race and illiteracy for the nine geographic divisions of the United States in 1930 is .946, whereas the individual (fourfold-point) correlation of race and illiteracy for the United States as a whole is only .203. The results are the same when the ecological areas are states or counties. For example, the ecological correlation on a state rather than a divisional basis is .773, still considerably higher than the individual correlation of .203.[4]

From another example, Robinson shows that associations found at the individual level may differ in sign as well as magnitude from those based on corresponding group data.[5] Findings similar to those of Robinson are reported in two other studies. Goodman found an individual correlation of .29 between race and occupation for employed females in Chicago in 1940, whereas the ecological correlation between the same variables from sixteen community areas, was .93.[6] Using data from a cross-national study incorporating 48 subpopulations in seven countries, Bice and Kalimo conclude that "the inferences drawn about the causal relations among the ecological variables cannot be generalized to apply to the individual persons in the corresponding population without paying attention to a possible causal fallacy in such generalizations."[7]

Chapter 5

The Sources and Uses of Health Status Data

Part I. Mortality Data

A. SELECTION AND DEFINITION OF DATA ITEMS

Information regarding mortality has traditionally been considered the most basic indicator of the health (or more precisely the lack thereof) of a population. Typically, mortality is expressed in rates, i.e., the number of deaths occurring in a specified period of time in a defined population, per population unit. Because of the agreement to express mortality in terms of such rates, only two basic data items are required: a numerator, defined as the number of deaths occurring in the population of interest during the specified time period, and a denominator, defined as the total number of members of that population. By convention the denominator is taken as the population at the mid-point of the specified time interval. These two figures represent a ratio which is sometimes multiplied by a constant for purposes of clarity. For example, the estimated resident population of the United States as of January 1, 1975 was 212.3 million persons and from July 1974 to July 1975, 1,935,000 deaths were reported to have occurred in the nation. When these data are expressed as a rate,

$$\frac{1,935,000}{212,302,000} \times 1,000$$

the result, known as the crude death rate, equals 9.1 deaths per 1,000 populatiom in this time period.

In addition to the two basic data items (the numerator and denominator which form the ratio), the HSA will require various control data in order to specify more detailed rates. These include such items as age, race, sex, geographic location and cause of death. As will be discussed below, the more detailed rates which can be constructed given such information have greater applicability to the health planning situation.

The control data which are typically available include the following information regarding each death:

- sex
- race
- age
- county of death
- hospital name
- state of birth
- citizenship
- marital status
- occupation
- residence
- cause of death (ICDA - code)
- autopsy
- and other information as relevant for certain specified types of deaths.

B. DESCRIPTION OF DATA SOURCES

Mortality data are derived from the U.S. Standard Death Certificate (see Appendix B of Chapter 4 in this section) which is completed by the attending physician or medical examiner for each death. Copies of the data forms are provided to the federal government and are published as Monthly Vital Statistics Report [1] and annually in Vital Statistics of the United States.[2]

The data are also maintained by state boards of health and/or local offices. Because published data are often aggregated to levels which will have little value except as points of comparison, the numerator figures for specific rates will in many cases have to be extracted by the HSA from these state level data files. The ease with which this can be done will vary depending upon the data storage and retrieval facilities and capacity of the state in question. For the most part, required denominator data are easily obtained from published Census reports.

C. DATA ANALYSIS AND USE

Because of the breadth of mortality data, and their general currency and availability, it is sometimes argued that mortality rates continue to be the health indicators most useful to health planners. On the other hand, these data clearly emphasize a narrow (if critical) negative dimension of health, and linkages between mortality rates and specific components of the health services delivery system are often difficult to establish. In addition, the reliability and validity of the cause of death data in mortality files is a subject of some debate. This issue is one aspect of a more general set of questions regarding the accuracy of clinical diagnosis. Research by Sanders,[3] Emerson,[4] Swartoot and Webster,[5] among others has indicated considerable discrepency between the cause of death shown on clinical certificates of death and that derived from autopsy findings.

In general, however, mortality rates will be useful to the health planner if they are very specific and if their use is primarily directed to the identification of health problem areas. For example, crude death rates (as described above) have often been used to make general comparisons of the health of several populations, or to examine trends over time. It is obviously not unreasonable to argue that populations which have lower death rates are "better off" on this critical dimension of health.

For the local area health planner, however, this kind of information will have little utility. First of all, crude death rates show relatively little variance in the contemporary United States. Second, even if an HSA or sub-region could be shown to have an inordinately high crude death rate, such information would not necessarily lead the planner toward action that could reasonably be expected to reduce the rate. With more specific rates, the planner is in an increasingly better position to assume that they are related to some part of the health care delivery system. For example, the calculation of death rates for age-specific sub-populations provides a first step in directing attention to a demonstrable issue: if infant mortality rates* are inordinately high the possibility arises that obstetrical care is less than adequate. It must be emphasized, however, that an assumption about such a connection is still extremely tenuous. It may well be that in those instances where infant mortality is relatively high, nutrition, sanitation, housing or other factors are far more important causal contributions than any feature of health care delivery. Nonetheless, the more specific rate serves to narrow the planner's focus in a useful way.

With yet further specificity and careful control of population variables some of these alternative explanations might be rejected. For example, if neonatal mortality rates** are calculated for several areas known to have similar demographic characteristics, (age, income, racial composition, etc.) a high rate in one of the regions is not likely due to socioeconomic differences.*** In such a case, it is logical to examine differences in the level of the <u>availability</u> of health resources of the appropriate type; e.g., pre-natal care, obstetrical services, etc. Where it can be demonstrated that areas with higher neonatal death rates are those in which relevant health care is less available, the assumption that the rate might be lowered by the application of appropriate resources is strengthened. On the other hand, if areas where the higher rates occur have an equivalent level of available resources, the planner can avoid the error of allocating additional resources when such action might serve no useful purpose. In this case, an examination of health service <u>utilization</u> patterns might be valuable.

---

*Unlike other mortality rates, the various rates applicable to children under one year of age are expressed as number of infant deaths per 1,000 live births. Infant deaths are defined as deaths of children under one year of age (not including fetal deaths).

**Neonatal deaths are those occurring within the first 28 days following birth.

***To say that these differences are not likely due to socioeconomic differences is not to say that it is impossible. See the discussion of the ecological fallacy in Chapter 4 of this section.

To summarize this brief discussion of mortality data, it should be noted that such information represents crude, but potentially useful data to the health planner. This potential utility depends upon the specificity of the rate (e.g., cause-specific death rates to the county level with controls for age, race and income). Even at this level of detail, it will be rare that mortality rates tie neatly to service availability. Rather, the rates will more often provide a first step in identifying a problem and indicating areas where more detailed analysis might provide the requisite information on which to base fruitful planning decisions. A concise and useful description of various mortality rates, calculation techniques and data sources is available to the HSA's in the Guide to Data for Health Systems Planners.[6] Using the techniques described there, HSA's will be able to compare local rates with those of the U.S. as a whole or with other relevant standards.

In addition, mortality data can be used to construct a more general "Unnecessary Death Index" (UDI) which, while incorporating unfortunate terminology, is useful as a means of expressing the degree to which the death rate in any given area differs from an analogous rate in some referent population.[7] UDI is defined simply as the difference between an expected death rate and the actual death rate in the population of concern, i.e.,

$$UDI = DR_A - DR_E$$

where the two death rates can be either crude or specific, but must be analogous.

The definition of an expected death rate is arbitrary, with such possibilities as the use of national or regional average rates being obvious examples. Within a problem identification context, it may be more valuable to use a rate which represents some desirable standard, such a standard will result in the calculation of UDI values which are generally positive. The HSA can then set as objectives the lowering of such values and monitor over time the success of efforts directed thereto.

The UDI has another very practical value. It expresses the relative frequency of deaths in the area of concern in an intuitively clear way. As a result, it will be useful in the community educative functions of the HSA, and may serve as a means of supporting resource allocation decisions.

A. SELECTION AND DEFINITION OF DATA ITEMS

It is, of course, entirely possible that a group of populations might have comparable death rates (even to very specific levels) while considerable variation in the frequency and/or nature of disease exists. Such an event would obviously indicate differences in the levels of health of the several populations—differences that could not be identified through the use of mortality data. Information describing the occurrence of disease and/or infirmity is termed "morbidity data" and is useful in examining this dimension of health status.

The total scope of morbidity is immensely broad, in that the term is often taken to include information related to any condition which is a departure from a state of physical or mental well-being. Thus, the presence of absence of various diseases, impairments, malformations, whether they are acute or chronic, the specific condition, the impact of the condition in terms of disability, and various other features, are within the class of information known as morbidity data. However, in this chapter, discussion is, for the most part, limited to information regarding the presence or absence of specified disease conditions. Disability data are dealt with in Part 3 of this chapter.

Morbidity information as defined here is typically expressed in terms of <u>incidence</u> and/or <u>prevalence</u> (of a specific disease, class of diseases, or disease in general). The former term refers to the number of new cases reported during a given time period, while the latter is the sum of reported cases existing at some specified time. For example, assume that within a defined population there were 1000 reported cases of disease X as of some point in time (t_1), and that during the single year following t_1, 100 new cases of the disease were reported. For the year following t_1 the incidence of disease X is 100, while the prevalence would be 1000 at t_1 and 1100 (1000 + 100) at t_2, assuming no cures or deaths, which would obviously be included in actual calculations. Such figures can refer to individual diseases, groups or classes of diseases, or a sum of all reported cases of reportable diseases.

It is possible to express these figures in terms of rates per unit population in much the same fashion as mortality rates are defined. Several differences must be borne in mind, however. The number of reported cases of a disease may well include successive occurrences in the same person, such that an incidence rate of 10 per 1000 population might result from ten individuals contracting the disease or the less likely possibility of one individual contracting the disease ten times. In addition, the type of data collected regarding each case depends upon the disease in question. Such control variables as age, race, sex and geographic location are routinely collected in the cases of some diseases but not others (see Figures 1, 2 and 3). As a result, the extent to which preferred specificity can be achieved will vary.

B. DESCRIPTION OF DATA SOURCES

Reliable extant data describing morbidity at the local level are not abundant. The major sources of such data are disease registries. Generally these are maintained at the state level and include up to fifty reportable diseases. For some diseases, an epidemiological report form is completed for each case and ultimately stored in original form, on microfilm or some machine-readable medium. For diseases such as chicken pox the number of cases is simply reported, while for venereal disease, very detailed information is often collected but issues of confidentiality may make the utilization thereof difficult. Typically, published reports of the disease registry data are available from state health departments, and summary statistics are provided to, and published by the Center for Disease Control (CDC) on a weekly basis.[8]

Figures 1, 2 and 3 are reproductions of forms utilized by physicians for the reporting of two classes of disease in one midwestern state. Although there is variation from state to state, these exemplify the type of data which are collected. In most cases summary statistics are annually published.

Apart from generally maintained reportable disease registries, some states and local areas, as well as individual hospitals "have or are actively developing, registries for specific disease categories".[9] The current development of cancer registries is a case in point. Such local area registries may be (or become) important sources of data for HSAs for several reasons. First, because of their local nature these records may be of particular utility in examining problems which are of unique significance in the local community. Second, with a much smaller areal base, the data are likely to be more timely and accessible. Third, because programs to collect such data are in many cases still in the planning stages, HSA's may have the opportunity to participate in the development of data collection forms and techniques, thereby insuring that the information is gathered in a manner consistent with the data needs of the agency.

Figure 1

Side 1

PHYSICIAN'S WEEKLY REPORT OF DISEASES BY NUMBER OF CASES

DISEASE	NO.	DISEASE	NO.
Chickenpox		Pertussis	
Conjunctivitis		Pneumonia	
Helminthic Infestations		Rheumatic Fever	
Histoplasmosis		Rocky Mountain Spotted Fever	
Impetigo		Rubella (German Measles)	
Infectious Mononucleosis		Scabies	
Influenza		Scarlet Fever	
Lymphocytic Choriomeningitis		Streptococcal Infections	
Measles (Rubeola)		Tinea Capitis (Ringworm of Scalp)	
Mumps (Infectious Parotitis)			
Pediculosis (Lousiness)			

SBH 2-18-102
4-63 100M

DATE______________19___ ________________________ M.D.

Side 2

PHYSICIAN'S WEEKLY REPORT OF DISEASES BY NUMBER OF CASES

INSTRUCTIONS TO PHYSICIANS—HCD 3 provides: "It shall be the duty of all physicians to report within 24 hours from the time of diagnosis all diseases or infestations listed below in this regulation* to the local health officer. All cases shall be excluded from schools or public gatherings as specified by the local health officer. The report may be made by number of cases of each disease on the official report card prescribed and furnished by The [redacted] State Board of Health or may be made by telephone or other means of communication. Such reports need not identify the patient, age, sex, color, date of onset and name of physician unless specifically requested by the local health officer."

*See diseases listed on front of card.

Completed cards are to be forwarded to the local health officer in whose jurisdiction the patient resides. Government stamped envelopes are furnished for this purpose.

INSTRUCTIONS TO HEALTH OFFICERS: Local Health Officer will total cases reported to him on this card and report on form SBH 2-18-50 (Health Officer's Weekly Report). Local health officer **should not** forward this form SBH 2-18-102 to [redacted] State Board of Health.

Figure 2

Side 1

PHYSICIAN'S REPORT OF COMMUNICABLE DISEASE BY NAME OF CASE

DISEASE OR SUSPECTED DISEASE		DATE OF ONSET		DATE OF REPORT
NAME OF PATIENT		AGE	MALE ☐ FEMALE ☐	WHITE ☐ NON-WHITE ☐
PATIENT'S STREET ADDRESS		CITY		COUNTY
PHYSICIAN TREATING THIS CASE	OFFICE ADDRESS			CITY
IF HOSPITALIZED—NAME AND CITY OF HOSPITAL				
REMARKS: (INCLUDE ANY AVAILABLE LABORATORY REPORTS)				
NAME OF PERSON COMPLETING REPORT, IF OTHER THAN PHYSICIAN NAMED ABOVE				ADDRESS

SBH 23-014
7/75 10M

Side 2

PHYSICIAN'S REPORT OF COMMUNICABLE DISEASE BY NAME OF CASE

INSTRUCTIONS TO PHYSICIAN: HCD 2 provides: "It shall be the duty of all superintendents, responsible officers or duly appointed representatives of each hospital in this state and all physicians to report within 24 hours all cases and suspected cases of the diseases listed below in this regulation to the local health officer in whose jurisdiction the patient is at the time of diagnosis.

Animal Bites	Infectious Hepatitis	Smallpox
Anthrax	Malaria	Tetanus
Aseptic Meningitis	Meningococcic Meningitis	Trachoma
Botulism	Ophthalmia Neonatorum, Gonococcal	Trichinosis
Brucellosis	Poliomyelitis	Tularemia
Cholera	Psittacosis	Typhoid Fever
Diphtheria	Q Fever	Typhus-Endemic
Dysentery, Amebic and Bacillary	Rabies in Man	Typhus-Epidemic
Encephalitis, Acute Infectious	Salmonellosis, Other than Typhoid Fever	Yellow Fever

Other communicable diseases which are not prevalent in [redacted] but are dangerous to the public health when they do occur and are not mentioned elsewhere in these regulations shall be reported in a like manner. Venereal diseases shall be reported in accordance with [redacted] State Board of Health Regulation HVD 2."

Venereal Disease and Tuberculosis are reported on special report forms provided by the State Board of Health, and are available from your local health department.

INSTRUCTIONS: Forward completed card to local health officer.

INSTRUCTIONS TO LOCAL HEALTH OFFICER: After noting case on SBH-23-020 (Health Officer's Weekly Report) forward this card to [redacted] State Board of Health.

Figure 3: An Example Venereal Disease Report Form

STATE BOARD OF HEALTH	CONFIDENTIAL VENEREAL DISEASE REPORT	DATE

PATIENT'S NAME: (5-24)	AGE (25-26)	RACE (27)	SEX (28)	MARITAL STATUS (29)	RESIDE

ADDRESS:	CITY:	COUNTY:

SYPHILIS (34-35)	OTHER VENEREAL DISEASES
☐ (10) PRIMARY (INITIAL CHANCRE PRESENT)	☐ (10) GONORRHEA (☐ LAB
☐ (20) SECONDARY (LESIONS OF SKIN OR MUCOSA)	☐ (20) PROPHYLACTIC TREATMENT
☐ (30) EARLY LATENT (LESS THAN 1 YEAR)	☐ (30) CHANCROID
☐ (40) EARLY LATENT (1 TO 2 YEARS)	☐ (40) GRANULOMA INGUINALE
☐ (45) LATE LATENT (OVER 2 YEARS)	☐ (50) LYMPHOGRANULOMA VENERE
☐ (50) CARDIOVASCULAR	
☐ (60) NEUROSYPHILIS	TREATMENT STATUS (
☐ (70) OTHER LATE ____________________	☐ (1) PATIENT HAD PRIOR TREAT INFECTION.
☐ (80) CONGENITAL	
☐ (90) PROPHYLACTIC TREATMENT FOR SYPHILIS	☐ (2) PATIENT IS UNDER OR HAS TREATMENT.

SIGNATURE OF PHYSICIAN

ADDRESS

Figure 3: An Example Venereal Disease Report Form

ARD OF HEALTH | CONFIDENTIAL VENEREAL DISEASE REPORT | DATE OF REPORT (1-4)

-24) | AGE (25-26) | RACE (27) | SEX (28) | MARITAL STATUS (29) | RESIDENCE CODE (30-33)

CITY: | COUNTY:

YPHILIS (34-35)

(INITIAL CHANCRE PRESENT)
Y (LESIONS OF SKIN OR MUCOSA)
TENT (LESS THAN 1 YEAR)
TENT (1 TO 2 YEARS)
ENT (OVER 2 YEARS)
SCULAR
HILIS
TE________________________
AL
CTIC TREATMENT FOR SYPHILIS

OTHER VENEREAL DISEASES (36-37)

☐ (10) GONORRHEA (☐ LAB. ☐ CLINICAL)
☐ (20) PROPHYLACTIC TREATMENT FOR GONORRHEA
☐ (30) CHANCROID
☐ (40) GRANULOMA INGUINALE
☐ (50) LYMPHOGRANULOMA VENEREUM

TREATMENT STATUS (38)

☐ (1) PATIENT HAD PRIOR TREATMENT FOR THIS INFECTION.
☐ (2) PATIENT IS UNDER OR HAS COMPLETED TREATMENT.

ICIAN

There are other sources of morbidity data, primarily survey information, which may be useful to some HSA's. For example, a National Health Interview Survey (NHIS) is conducted on an ongoing basis by the National Center for Health Statistics.[10] It involves continuous sampling and interviewing of the civilian population regarding health issues. Included in the interview is the collection of information about morbid conditions (date of onset, condition, restriction of activity, treatment, etc.) with reference to three time periods: the two week, three month, and one year periods preceding the interview. It should be noted that morbidity, in this survey, is taken to include "active or progressive diseases, impairments, injuries or congenital malformations."[11]

The sampling technique utilized in this survey permits direct analysis only at the national or regional level, but will provide points of comparison if the HSA has local area data available. Moreover, it may be possible to synthetically derive estimates from NHIS data that are applicable at the state level. Research in this area is continuing with some current effort directed at assessing the utility of reweighted means as estimates.[12]

Another source of morbidity data is the National Health Examination Survey.[13] Also based on a national sample, this survey is of particular interest because it includes a clinical examination of respondents. As a result, it yields data regarding conditions and other health-related phenomena which may not be reported in interviews. Constraints on the use of these data are comparable to those described above with reference to the health interview survey data.

A further source of potentially useful morbidity information includes a variety of publicly accessible survey data files from several sources. Although there would appear to be no exhaustive catalog describing these resources, many are available from the National Technical Information Service,* NCHS, or other research organizations.

*National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22151

A few examples will suffice to indicate the nature of this data resource and its potential utility for any given HSA.

- The Experimental Health Services Delivery Systems' Household Survey

  This survey was implemented in 17 EHSDS communities over the period 1972-1974. It gathered information about the health services utilization, disability, etc., of each member of the sampled families. The core portions of the data from all 17 of these surveys are available on a single computer tape through NTIS. To the extent that an HSA has comparable boundaries or even a similar demographic structure, one or more of these files of survey data may be useful.

- Other Local Area Surveys

  Depending upon location, some HSAs may find that other predecessor health planning organizations, e.g., CHP agencies, RMPs or university or governmental research groups have gathered data relevant to morbidity at some time in the past. While the issue of timeliness will be critical, it is possible that such information may be of value to the HSA.

- Other National Surveys

  In addition to the national survey research carried out directly by NCHS, such organizations as the Center for Health Administration Studies (CHAS)* and the National Opinion Research Center (NORC) have conducted surveys related to health issues.[14] In some cases these may provide detail of specific interest to certain HSAs.

---

*Center for Hospital Administration Studies, University of Chicago, 5720 Woodlawn Avenue, Chicago, IL 60637.

C. DATA ANALYSIS AND USE

In general, it is contended here that morbidity information, especially rates of the reported occurrence of disease, will be most useful to the planner as suggestive, rather than definitive data. It is not considered likely that such data can provide the kind of information directly useful in the planning process. Rather, in conjunction with other data, morbidity information is likely to be most useful in directing the health planner to those areas where more detailed research will be most valuable. The reasons for this limited utility can be expressed in two major issues relating to the disease registry type of information regarding morbid conditions. The first deals with the accuracy of the data, while the second refers to the complex relationship between available morbidity data and the status of health services provision in any given geographic area.

Regarding the former issue, it is well known that morbidity data derived from registries are understatements of the actual presence of disease in the population of interest. Some diseases are not reportable, so any summation represents only the reported cases of some subset of disease entities. Furthermore, it is axiomatic that some portion of the actual occurrences of reportable diseases will go unreported for any one of a variety of reasons. This can be demonstrated as follows: assume that ten persons have been infected by reportable disease X. Six of these persons visit a physician or other provider of health care during the course of the disease—four cases are thus not available for reporting. Of the six persons seen by physicians, perhaps one case remains undiagnosed and hence unreported. It is further possible that one or more of the diagnosed cases are not reported—physicians sometimes treat certain patients or diseases in confidence. Each of these potential barriers to complete reportage will have varying levels of impact depending upon the disease. In a case such as influenza, fewer doctor contacts will result; in cases such as venereal disease, non-reporting may be significant. In any event, the diagnosed and reported cases of a disease will underestimate its actual presence, and the degree and pattern of underestimation will be unknown.

The second issue in the use of morbidity data derives from what has been termed "a 'paradoxical effect' whereby better health services result in a greater prevalence"[15] of measured morbidity. That is, the measured frequency of disease will in part be a function of the availability and utilization of medical services. This effect operates in two ways. On the one hand, as the example provided above indicates, diseases can only be reported if there is a physician contact. As the proportionate number of physicians increases, the number of physician contacts will probably increase, as will the number of reports of disease. On the other hand, as medical care becomes more available it presumably will contribute in some way to average longevity, permitting the survival of individuals who are subject (perhaps especially so) to various morbid conditions.

A concrete example may serve to clarify these issues. Suppose a health planner has assembled morbidity data for three regions, calculated the prevalence rates of several diseases for the regions, and has found that the rates tend to be marginally higher in one of the areas. The planner examines the demographic characteristics of each region, and finding them similar, rejects the hypothesis that the higher rates are due to differences in the age, sex or economic structures. He then examines the availability of services in the three regions and finds that there are marginally fewer doctors per person in the same area where there are marginally higher rates of reported disease. At this point a planner might conclude that the region with higher rates has lower health status and proceed with actions directed at providing more and better health services—for example, by directing efforts towards attracting physicians, or by favorably viewing proposals for the construction of an outpatient department in the region's hospital. If such action results in increased service availability, the measured levels of morbidity in the region of interest would increase (due to an increase in physician contacts) and, relative to the other areas originally examined, this area would appear to be getting worse, rather than better.

This example demonstrates that simplistic action/evaluation programs based on disease registry data would be particularly problematic. Low morbidity rates based on such data may well imply not a state of health, but a lack of physicians (or both, or neither) and hence planning resource allocation on such a basis could be hazardous.

This is not to argue that morbidity data of this type are not of interest to the health planner. Combined with other information, they represent a significant component of health status, and they have certain clear uses in and of themselves. For example, some of the reportable infectious diseases have for some time been the object of extensive immunization programs. Monitoring the reports of such diseases (controlling for population variables and the number of physicians submitting reports) will provide the HSA with indicators of the current success of immunization practices. Similarly, certain diseases which are associated with ongoing prevention/diagnosis/treatment programs such as venereal disease and tuberculosis are susceptible to similar monitoring (if careful data controls are maintained) for the purpose of assessing the programs. Sudden changes in the reported rates of disease are also useful data—especially regarding disease related to environmental and/or public health issues such as salmonellosis,[16] hepatitis, etc.

As was the case regarding mortality data, the utility of morbidity information for planning (as opposed to purely informational) purposes will depend upon their completeness and specificity. Unlike mortality information, however, morbidity data will not typically be complete and will rarely be amenable to the type of specific rate construction that is desirable.

Part 3. Disability Data

A. SELECTION AND DEFINITION OF DATA ITEMS

In addition to incidence and prevalence rates of various diseases, morbidity information is sometimes measured in terms of the extent to which functional disability results from morbid conditions. The result is a measure of the impact rather than the presence/absence of such conditions. Disability information is sometimes expressed in terms of occurrence rates in a population, e.g., percent experiencing disability. More frequently, the degree of disability is assessed in such terms as the number of disability days per unit population per unit time, e.g., 6.0 bed disability days per person per year. In such cases, disability days are usually classified according to type, such as restriction of usual activity, work/school loss, bed-disability, etc.). When summed over the population base, which is usually a sample, disability data provide statistics describing one measure of functional impairment in that population.

In this part, three types of disability data will be discussed: occurrence, number of bed-days and number of restricted-activity days. This is an admittedly limited emphasis but should adequately exemplify the potential uses of such data. NCHS definitions of the relevant terms follow:[17]

- disability: "a general term used to describe any temporary or long-term reduction of a person's activity as a result of an acute or chronic condition."
- bed-disability day: "a day...on which a person stays in bed for all or most of the day because of a specific illness or injury. All or most of the day is defined as more than half of the daylight hours. All hospital days for inpatients are considered to be days of bed disability even if the patient was not actually in bed at the hospital."
- restricted-activity day: "a day...on which a person cuts down on his usual activities for the whole of that day because of an illness or injury. The term 'usual activities' for any day means the things that the person would ordinarily do on that day. For children under school age,

usual activities depend on whatever the usual pattern is for the child's day, which will in turn be affected by the age of the child, weather conditions and so forth. For retired or elderly persons, usual activities might consist of almost no activity, but cutting down on even a small amount for as much as a day would constitute restricted activity. On Sundays or holidays usual actvities are the things the person usually does on such days-going to church, playing golf, visiting friends or relatives, or staying at home and listening to the radio, reading, looking at television and so forth. Persons who have permanently reduced their usual activities because of a chronic condition might not report any restricted activity days during a two-week period. Therefore, absence of restricted activity days does not imply normal health.

Restricted activity does not imply complete inactivity but does imply only the minimum of usual activities. A special nap for an hour after lunch does not constitute cutting down on usual activities, nor does the elimination of a heavy chore like cleaning ashes out of the furnace or hanging out the wash. If a farmer or housewife carries on only the minimum of a day's chores, however, this is a day of restricted activity.

A day spent in bed or a day home from work or school because of illness or injury is, of course, a restricted activity day.

Combined with standard control data, such as were described in previous parts (age, race, sex, income, geographic location, etc.), disability data can be used to calculate:

- number of persons experiencing disability, for various sub-groups and total population
- number of bed-disability days per person per year, for various sub-groups and total population
- number of restricted-activity days per person per year for various sub-groups and total population

When aggregated to the total health service area, figures such as these can be compared to referent population figures as one indicator of overall health conditions. When determined for specified geographic or population sub-groups, the figures can contribute to efforts to assess the relative health services needs of such sub-groups.

B. DESCRIPTION OF DATA SOURCES

Disability data are generally derived from interview surveys (see Chapter 15 in Section IV). As a result, there are not a large number of sources of previously collected data which are directly relevant to any given local area.* Furthermore, because they are typically derived from samples, such data that are available are subject to the normal constraints in generalizing to a specified target population.

National and regional disability data are available from two sources. Foremost is the information derived from NCHS's National Health Interview Survey (NHIS) and published as Series 10 of Vital and Health Statistics. These reports include detailed tabulations of both the occurrence and degree (number of days) of various types of disability, broken down by a variety of socio-demographic variables. Unfortunately, the nature of the sampling procedure in this survey limits analysis to the national and regional levels. Hence, the direct applicability of these data at the sub-regional level is limited to their value as comparative referent points for those localities which have comparable data available at the local level. An indirect use of these data for deriving synthetic estimates applicable at the state level is discussed below.

The second source of disability data includes other national surveys—such as that administered by the Center for Health Administration Studies (CHAS) in 1970. Like the NHIS data, these are not designed for analysis to the level of small geographic sub-units and will, therefore, be most useful as referent data. The great detail in such data files is unusual, however, and their utility to the local

---

*For certain specific types of "disability days" local data may be available. For example, data on the number of hospital and nursing home patient days may be available to use in estimating "bed-disability days"; school and work absentee records may be available to use in estimating "restricted-activity days"; welfare records pertaining to the payments to persons who are blind and disabled may be available to use in estimating the number of persons experiencing disability, etc.

area planner, even for straightforward informational purposes, should not be ignored.

A second general type of detailed survey data regarding disability includes previously conducted local or community surveys such as those completed by EHSDS, CHPs or other research organizations. This data source will be useful only under certain conditions:

- the data must be reasonably current
- the survey population must be fundamentally analogous to the population of the HSA
- the sampling procedures used must be consistent with the data analysis objectives of the HSA
- the data collection techniques (question wording, definitions of variables, etc.) must be consistent with those of interest to the HSA

It is unlikely that there will be a direct fit between any previously collected data file and these criteria of utility. Hence, the HSA will be faced with the question of whether or not such data have a <u>degree</u> of applicability that make them valuable. For example, data which were gathered in the preceding two or three years are no doubt worthy of examination (if the other conditions are met), while a data file which is ten years old is probably not applicable to the current situation. There is no arbitrary time limit to data utility—the decision will have to be made within the context of the objectives of the HSA. Similarly, it is unlikely that an HSA will have precisely the same boundaries as some previous health planning or management organization. Some HSA's may, however, have a population whose demographic characteristics are very similar to such predecessor organizations. While direct estimation of population characteristics from such a data file is impossible, it may be valuable for an HSA to examine disability data which refer to a <u>similar</u> population. Again, the utility will depend on the HSA's objectives. The same questions will hold for other issues of data comparability.

It must be acknowledged that these sources of disability data are not as broadly based or directly useful as the mortality and morbidity data discussed previously. It may be that a given HSA will simply not have access to "extant" disability data which are consistent with its research objectives. In such cases, primary data collection by way of a family health survey may be necessary (see Chapter 15 in Section IV).

C. DATA ANALYSIS AND USE

The techniques of analysis and potential uses of disability data are fundamentally limited by the data sources. The national survey data are extremely detailed regarding the nature and degree of disability, but are not directly applicable to local area sub-units. Only the community household survey combines these attributes in a way that permits detailed analysis of disability information as opposed to more general descriptive uses.

National survey data may be used in two ways. The available estimates are in themselves valuable reference points for comparison if the HSA has comparable disability data of a local nature. In addition, it is possible to adjust the regional (or national) estimates according to the demographic structure of a sub-unit or state, providing "synthetic estimates" of these rates for states, or other units. These estimates may not be used for comparative purposes with the national data from which they are derived because such comparisons would be tautological.

Research directed at the examination of synthetic estimates is continuing and final conclusions as to their reliability and validity cannot be drawn. However, two techniques for the derivation of such estimates have been described in the literature. Regardless of which technique is used, certain features of these synthetic estimates must be borne in mind. First, they are not probability estimates and, therefore, it is not appropriate to estimate variation with standard probability-based statistical techniques.[18] Second, such estimates are not unbiased. Third, the estimates are based on the assumption that certain characteristics which are related to health variable at the national level are related to the same variables in the same way at the geographic level of interest.

The first of the two techniques to be described here was developed by NCHS in response to known demand for sub-national estimates of statistics derived from the National Health Interview Survey.[19] The estimation procedure begins with the known relationship between such socio-demographic characteristics as race, age, income, etc., and such health indicators as disability. Therefore, the estimate is essentially the national average, adjusted to the characteristics of the population of interest based on the following model:

$$\bar{x} = \sum_{\alpha=1}^{k} (P\alpha \; \bar{x}\alpha)$$

where:

$\bar{x}$ = the estimate to be derived.

$P\alpha$ = the proportion of the population of concern (e.g., a state) who are members of cell α. The latter is the group defined by the socioeconomic dimension(s) of concern, e.g., age, race, sex, etc. (source: Census data).

$\bar{x}\alpha$ = the national estimate of $\bar{x}$ for the members of the α cell in the national sample. (source: National Health Interview Survey data).

k = the number of α cells utilized.

A brief example of the use of this model follows: suppose an HSA wishes to estimate the average number of bed-disability days experienced annually by residents of a state using the NCHS procedures. This would be achieved by 1) selecting the α groups of interest, presumably those in which the local population has notably higher or lower representation than the nation, and recording the number of disability days per person per year experienced by the members of each selected α group in the national sample ($\bar{x}\alpha$); 2) recording the proportion of the state population who are members of analogous α groups ($P\alpha$); and 3) summing the products of these figures. If only two α groups (such as two racial categories) were selected, the estimation would proceed as follows:

	white	all other
Bed-disability days per person per year:	5.9 ($\bar{x}\alpha_1$)	7.6 ($\bar{x}\alpha_2$)
Proportions of state population in each group:	.65 ($P\alpha_1$)	.35 ($P\alpha_2$)
State estimates for each group:	.65(5.9)	35(7.6)
State estimate:	6.50 ($\Sigma P\alpha\bar{x}\alpha$)	

This synthetic state estimate thus represents the number of bed-days per person per year that would have been experienced by a population whose racial composition was that described, if two race groups were the only two groups of concern and the members of the two racial groups in the state experienced disability at the same rate as their counterparts in the national sample.

It should be noted that this simple example understates one clear difficulty in the use of this technique. According to NCHS, "The α variables were limited to what are believed to be the seven key ones in health experiences." These seven variables are color, sex, age, (four groups) residence, family income, family size and industry of head of family. These variables would provide 384 cross-classification cells, but are collapsed into 78 cells for which "reliable estimates" are availab. Bed-disability days data are available from published reports for e α cells defined by the following cross classifications.[20]

- Place of Residence by Sex by Age
- Family Income by Sex by Age
- Color by Sex by Age

Before embarking on synthetic estimation procedures, the HSA must, therefore, insure that the requisite cell data are available from NCHS and that analogous local population data are available from census reports. The NCHS cross classifications vary depending upon the disability variable of concern (e.g., bed-days, restricted-activity days, etc.). The census data are generally available in detail greater than that utilized by NCHS (see Chapter 4 in this section).

A variant procedure for deriving comparable estimates at either the state or SMSA level has been described by Schneider.[21] In this instance, the estimates are derived with mathematical techniques drawn from regression. The results, however, are analogous to the weighted averages described above. The two techniques differ in procedure but very little in terms of results.

The Schneider technique is, like that of NCHS, based on the known relationship between disability information and certain demographic characteristics of a population. However, when applied, this procedure requires less data manipulation. This is because the regression coefficients required are based on bivariate correlations which can be calculated from easily available published tables. As was indicated previously, the NCHS procedure would require crosstabulation at a level of detail not common in published tables. In this regard the Schneider procedure is more broadly applicable.

On the other hand, the Schneider procedure requires the availability of computer facilities and the necessary understanding of correlation and regression to compute beta weights associated with each demographic variable of interest. Once the weights have been determined, "computing the estimates for a specific population is a very simple task that can easily be done with pencil and paper. The value of each demographic variable for the target population is multiplied by the corresponding beta weight. These terms are added up to obtain the estimate."[22] The beta weights are calculated by computer from a correlation matrix derived from published tables. A more detailed description of the latter calculations is available in Schneider's article as well as most statistics text books.

It should be emphasized that like the NCHS technique, the regression estimation procedure includes no method of statistically assessing the accuracy of the estimates, and is by definition insensitive to idiosyncratic differences in a specified population or changes occurring therein.

In spite of their limitations, estimates derived from national survey data are probably better than no estimates, and the adjustments undoubtedly make them more applicable to local areas than the national estimates would be in the aggregate form. They will be useful to health planners in either one of two situations. In the absence of local area estimates derived from other surveys, synthetic estimates can provide crude indicators of disability in a specified population (and hence a benchmark which adds meaning to comparisons of the relative availability of services). When local data are also available, the synthetic estimates provide a means whereby actual local conditions can be compared with conditions which would obtain if the population were similar to the nation as a whole. In this way, an assessment of relative health status can be obtained.

Part 4. Composite Indicators of Health Status

It is sometimes useful to construct measures of health status which combine estimates of more than one health related phenomena.[23] Such composite indicators reflect the impact of a number of dimensions of the concept of health. As such, they provide a means of summarizing health status in a population. However, composite indicators are less sensitive to changes in a subset of the phenomena under examination than are more specific health status indicators, and they are more difficult to relate to specific components of the health services delivery system. Each composite indicator discussed in this part is based on one or more of the variables and data sources described earlier. Definitions and sources are therefore not repeated.

Sullivan's Life Table Approach:[24] This index represents an attempt to combine mortality and morbidity data in a single figure. The index value is an estimate of the number of years of life free of disability that would be experienced by members of a given age cohort if current mortality and disability rates obtain throughout their lives. Because mortality and morbidity conditions are unlikely to remain static over the survival span of any cohort, the index should not be construed as a valuable predictive tool. However, by providing estimates based on current health phenomena they allow the health planner to evaluate the projected impact of current health conditions in the population of concern and to compare those estimates with referent populations.

In essence the index is composed of two data items: 1) a current estimate of life expectancy and 2) a current estimate of the number of disability days experienced by each relevant sub-population. After a necessary conversion to annual units, the index is expressed as

LE - DE

where,

LE = life expectancy in years

DE = disability expectancy in years

For example, if the male population has an average life expectancy at birth of 66.8 years, and a disability expectancy of 5.2 years, the index value for the sub-population equals 61.6 years.

If populations with large differences in conventional life expectancy are compared, these differences will account for a large proportion of differences in index values. However, the rates of disability are significant contributors which will take on greater importance if populations are defined in ways that limit variation in mortality rates.

An important issue in the utility of this index is data availability. The planner will need abridged life tables for the population groups of concern as well as age-specific disability rates for the defined groups. At the national level, life tables are available broken down by sex and race. It is possible that state level tables might be available from insurance companies or alternatively, they might be constructed from state-specific mortality data.

As we have indicated earlier, it is disability information which is more difficult to obtain. Although the use of previously described synthetic estimation procedures to obtain disability rates is not entirely precluded, it would be problematic because the national estimates are presumably valuable points of comparison. They cannot, therefore, be adjusted for local conditions and simultaneously compared to local conditions. The resulting comparisons would be uninterpretable. Hence, the planner will be able to utilize Sullivan's approach only to the extent that 1) local disability data are available or 2) he has no interest in comparative analysis.

<u>Q-Index</u>: In an ongoing attempt to develop a guide for resource allocation based on the relative impact of various diseases, the Division of Indian Health, has developed a composite index called "Q". Because of the ongoing nature of this work, several versions of the index exist. The following discussion draws heavily from Donabedian.[25]

The Q-Index can be calculated for a specific disease, a group of diseases or a summation over all disease categories. It takes the form:

$$Q = (M)(D)(P) + \frac{C}{N}(CON) + \frac{A}{N}(CON) + \frac{B}{N}(CON)$$

where,

M = the ratio of the specified mortality rate in the target population to that in the referent population

D = crude mortality rate (per 100,000 per year) in the target population

P = average years of life lost due to each death

C = days of restricted activity during the year in the target population

A = number of inpatient days during the year in the target population

B = number of hospital outpatient visits during the year in the target population

N = number of persons in the target population

CON = a conversion factor to translate days or visits into years (since mortality losses are expressed in years)

For example, to convert $\frac{C}{N}$ (restricted-activity days per person per year) to years of restricted-activity per 100,000 population CON = $\frac{100,000}{365} = 274$. If an outpatient visit was assumed to equal the loss of 1/3 day of activity, that factor would equal $\frac{100,000}{365} \times \frac{1}{3} = 91$

Of these factors, only A and B have not been discussed in previous parts. Estimates of both factors can be obtained from extant hospital utilization data, when B is defined so that outpatient visits include only those to hospital outpatient facilities.[26] There is no fundamental reason why factor B could not be expanded to include ambulatory care visits in other settings if the necessary data were available (see Chapter 11 in Section IV). In such cases, the conversion factor would be empirically derived from the mean length of each visit.

Like Sullivan's approach, the Q-Index provides an estimate of the total losses due to mortality and certain types of morbidity in a target population (factors D, P, C, A and B). It should be noted that the index value obtained is substantially affected by the losses due to mortality incorporated in factors D and P. In fact, for 11 specified disease categories the rankings of total loss that are obtained from mortality data alone are strikingly comparable to those obtained by use of Q. According to Donabedian "it is clear that (in the total Q-index) losses from mortality are by far the dominant factor, with losses from restricted activity a poor second. Losses from hospital stay or ambulatory care are negligible by comparison."[27]

In addition to estimating losses due to mortality and morbidity, the Q-index includes a factor which reflects (presumably) the amenability of disease-specific mortalities to influence by health action. Briefly, factor M represents the degree to which a mortality rate in the target population exceeds that of the referent population. This difference is assumed to reflect the effects of some set of ameliorating influences, including health services, which exist in the referent population. Even with careful selection of a referent population with comparable socio-demographic characteristics, it would remain virtually impossible to identify the specific conditions or components of the health services system which are contributing to the lower rate. Nevertheless, the very fact that such "better" rates are known to be achievable provides the planner with a meaningful objective, and careful comparison of conditions in the two populations might provide indications of the health action areas where positive results are possible.

A number of other quite complex weighted average estimation procedures have been used to assess health status. These range from fairly simple additive models to immensely complicated formulae which weight, and then sum, as many as fifty health related variables. Such efforts have two major problems. The selection of component parts is as yet unguided by a clear conceptual understanding of the nature of health. Hence, one such summary score includes various death rates, health expenditures per capita, physicians per 100,000 population and the ratio of medical care costs to the consumer price index, while another includes suicide rates, illegitimacy rates, etc. In the present state of development, the selection of component variables seems to have little rhyme or reason. The various weights assigned to each variable are plagued by the same problem—how can they be determined in some non-arbitrary manner?

The second major problem refers to the utility of such composite measures. Even if one were confident that the ultimate summary score was a valid and reliable indicator of health, the very number of its constituent components makes linkage to specific health services delivery components impossible. Thus, this type of indicator would not likely be useful to HSAs in making resources development type decisions.

FOOTNOTES - SECTION II

Chapter 3

1. P.L. 93-641, Section 1601 (3).
2. *Ibid.*, Section 1532 (c) (8).
3. *Federal Register*, October 18, 1974, p. 37317.
4. P.L. 93-641, Section 1513 (a) (1).
5. *Ibid.*, Section 1535 (c) (7) (A).
6. *Ibid.*, Section 1513 (b) (1) (A).
7. T. W. Bice, "Comments on Health Indicators: Methodological Perspectives." Paper presented at the Seminar on Health Indicators, Columbia University, August 1973, Revised June 1975.
8. World [illegible] Organizations, "Constitution", In *The First Ten Years o[illegible] e World Health Organization*, Geneva, W. H. O. 1958.
9. Bice, *op. cit.*, p. 10.
10. *Ibid.*, p. 11.
11. cf. D. F. Sullivan, *Conceptual Problems in Developing an Index of Health*, NCHS Series 2, No. 17, Washington, D.C., 1966; J. M. Moriyama, "Problems in the Measurement of Health Status" in E. B. Sheldon and W. E. Moore (eds), *Indicators of Social Change*, New York, Russell Sage,1968; S. B. Goldsmith, "The Status of Health Status Indicators," *Health Services Report*, 87, 1972.

Chapter 4

1. U.S. Bureau of the Census, *1[illegible] Census Users' Guide*. U.S. Government Printing Office, Washington, D.C., 1970, pp. 91-109.
2. *Ibid.*
3. *Ibid.*, p. 21.
4. Lu Ann Aday, *The Utilization of Health Services: Indices and Correlates - A Research Bibliography*. DHEW Publication No. (HSM) 73-3003, 1972, p. 17.
5. *Ibid.*, pp. 18-19.

6. Barry Decker, Hospital Discharges and Days of Care in Relation to Medical Need, Northeast Ohio Regional Medical Program, Cleveland, Ohio.

7. Lu Ann Aday, op. cit., p. 20.

8. Ibid., pp. 23-24.

9. Ibid., pp. 19-20.

10. The source of the information contained in Table 5 is, with slight editing, the Kansas State Board of Health.

11. Ronald Andersen and Odin W. Anderson, A Decade of Health Services, University of Chicago Press, Chicago, 1967, pp. 180-181.

12. U.S. Bureau of the Census, Census Use Study: Health Information System, Report No. 7, Washington, D.C., 1969.

13. U.S. Bureau of the Census, Census Use Study: Health Information System II, Report No. 12, Washington D.C., 1971.

14. Harry A. Sultz, Frank J. Rens, and Donald A. Brothers, "A CHIP System Based on Computer Graphics," paper presented at the 99th annual meeting of the APHA, Minneapolis, Minnesota, October 12, 1971.

15. Anthony Oreglia, "A Study of Social and Medical Factors Affecting Infant Mortality." Unpublished Master's Thesis, Purdue University, August 1974.

16. H. M. Blalock, Jr., Causal Inferences in Nonexperimental Research. University of North Carolina Press, Chapel Hill, 1964, Chapter IV; and C. E. Gehlke and K. Biehl, "Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material." Journal of the American Statistical Association, 29, March, 1934, pp. 169-170.

17. Otis D. Duncan, et al., Statistical Geography: Problems in Analyzing Areal Data. The Free Press of Glencoe, Illinois, 1961, pp. 60-80.

Chapter 4 (Appendix C)

1. W. S. Robinson, "Ecological Correlations and the Behavior of Individuals." American Sociological Review, 15, June, 1950, p. 351.

2. Ibid., p. 357.

3. Hanan Selvin, "Durkheim's Suicide and Problems of Empirical Research." American Journal of Sociology, 63, May, 1958.

4. W. S. Robinson, _op_. _cit_., p. 353.

5. _Ibid_., p. 354.

6. Leo A. Goodman, "Some Alternatives to Ecological Correlation." _American Journal of Sociology_, 64, May, 1959, p. 615.

7. Thomas W. Bice and Esko Kalimo, "Causal Models and Fallacies in Cross-National Epidemiological Research." Paper presented at the Sixth International Scientific Meeting of the International Epidemiological Association, Primosten, Yugoslavia, August 28-September 4, 1971.

Chapter 5

1. National Center for Health Statistics, _Monthly Vital Statistics Report_, Public Health Service, DHEW, (monthly publication).

2. NCHS, _Vital Statistics of the United States_, Public Health Service, DHEW (published annually).

3. B. S. Sanders, "Completeness and Reliability of Diagnosis in Therapeutic Practice," _Journal of Health and Human Behavior_, Vol 5, 1964, pp. 84-94.

4. H. Emerson, "Reliability of Statements of Cause of Death from Clinical and Pathological Viewpoints", _American Journal of Public Health_, Vol. 6, 1916, pp. 680-685.

5. H. O. Swartoot and R. G. Webster, "To What Degree are Mortality Statistics Dependable?" _American Journal of Public Health_, Vol. 30, 1940, pp. 811-815.

6. _Guide to Data for Health Systems Planners_, DHEW Publication No. (HRA) 76-14502, 1976.

7. L. Guralnick and A. Jackson, "An Index of Unnecessary Deaths," _Public Health Reports_, Vol. 82, 1967, pp. 180-182.

8. Center for Disease Control, _Morbidity and Mortality Weekly Reports_, Atlanta, Georgia.

9. _Guide to Data for Health Systems Planners_, _op_. _cit_., p. 58.

10. For information regarding survey methodology see National Center for Health Statistics, _Vital and Health Statistics_, Series 2.

11. _Ibid_., Series 10, No. 90.

12. NCHS, _Synthetic State Estimates of Disability_, DHEW, 1968.

13. NCHS, _Vital and Health Statistics_, Series 11.

14. cf. R. Andersen and O. W. Anderson, A Decade of Health Services Research, Chicago, University of Chicago, 1967, and R. Andersen et al., Health Services Use, DHEW Publication No. (HSM) 73-3004, Washington, D.C., 1972.

15. B. S. Sanders, "Measuring Community Health Levels," American Journal of Public Health, Vol. 54, 1964, pp. 1063-1070.

16. cf. E. V. Morse and M. A. Duncan, "Salmonellosis: An Environmental Health Problem Affecting Animals and Man," Journal of the American Veterinary Medical Association, Vol. 165, 1974, pp. 1015-1019.

17. NCHS, "Current Estimates from the Health Interview Survey," Vital and Health Statistics, Series 10, No. 95, Rockville, Maryland, 1973.

18. M. E. Gonzalez, "Use and Evaluation of Synthetic Estimates." Paper presented at th 133 annual American Statistical Association meeting, New York City, 1973, p. 1.

19. National Center for Health Statistics, Synthetic State Estimates of Disability, op. cit.

20. NCHS, Vital and Health Statistics, Series 10, Nos. 4, 24, 47 and 67.

21. R. E. Schneider, "A Procedure for Estimating Health Characteristics of Target Populations Using Available Data," Unpublished paper; Minneapolis, Minnesota, 1970.

22. Ibid., p. 15.

23. cf. C. L. Chiang "An Index of Health: Mathematical Models" Vital and Health Statistics, Series 2, No. 5, Washington, D.C., 1965; S. Fanshel and J. W. Bush "A Health-Status Index and Its Application to Health Services Outcomes" Operators Research, Vol. 18, 1970, pp. 1021-1060.

24. D. F. Sullivan, "A Single Index of Mortality and Morbidity," HSMHA Health Reports, Vol. 86, 1971, pp. 347-354.

25. A. Donabedian, Aspects of Medical Care Administration, Harvard University Press, 1973, pp. 179-185.

26. American Hospital Association's, Guide to the Health Care Field, and Hospital Statistics, Chicago, Illinois.

27. A. Donabedian, op cit., pp. 183-184.

SECTION III

HEALTH RESOURCES DATA

Chapter 6

The Need for Health Resources Data

Among the primary responsibilities of HSAs and State Agencies under P.L. 93-641 is the promotion of the development of needed health resources within their health service areas.[1] In determining which health resources are needed, the health facilities, manpower and services that are available within health service areas will need to be known by HSAs. If existing resources are found to be inadequate to meet the needs of the residents in the health service areas, programs will need to be developed that more efficiently organize or expand the existing resources.

Basic to the development of needed health resources, therefore, is the availability of data concerning "the number, type and location of the area's health resources, including health services, manpower, and facilities."[2] The specific health resource information needs of HSA's and the requisite data sources and procedures for compiling health resource inventories are contained in the companion document to this Handbook, A Guide to the Development of Health Resource Inventories. The primary purpose of this section of the Handbook, therefore, is not the identification and description of the existing data sources concerning health resources, but rather, the presentation of selected techniques for the analysis of the existing data and their uses in planning, resources development and project review. The specific types of health resources discussed in this section include hospitals, physicians and institutional health manpower.

A. HOSPITAL DATA

A variety of data related to hospitals will be needed by HSAs and State Agencies. In addition to assembling and analyzing data concerning the number, type and location of their area's hospitals, HSAs are required under P.L. 93-641 to assemble and analyze data concerning the status of hospitals as part of their area's health service delivery system, and the use of these hospitals by their area's residents.[3] These data are to be used by HSAs to periodically review and assess the appropriateness of the hospital services offered in their health service areas. If such reviews and assessments include the identification of hospital service duplications and gaps, then data on the number and type of specialized beds and services contained in each area's hospital will be essential.

Data on the availability and use of hospitals will also be needed by HSAs and State Agencies in their reviews of proposed construction projects under Section 1122 of the Social Security Act and, where applicable, state certificate of need programs. Among the criteria to be used in such reviews is "the need that the population served or to be served by such services has for such services."[4] Data on the current and projected utilization of hospital beds and services among the area's residents, and the capabity of existing hospitals to produce the projected bed and service needs of the area's residents can be used in applying this criterion.

Chapter 7 in this section includes a detailed discussion of the analyses and uses of hospital data in planning, resources development and project review. Specifically, these uses are 1) assessing the availability of hospitals, beds and services; 2) measuring hospital utilization, 3) assessing the capacity of a hospital to produce service; and 4) projecting hospital bed need.

B. PHYSICIAN DATA

Contained within P.L. 93-641 are several "national health priorities" which Congress found deserve consideration in the health planning and resources development programs of HSAs and State Agencies. First among these priorities is the provision of primary medical care to medically underserved populations, especially those which are located in rural or economically depressed areas.[5] This is to be achieved, in part, through the development of certain organized systems for the provision of primary medical care, such as medical group practices and health maintenance organizations.[6] No other factor is as critical in the provision of primary medical care and in the development of the organized systems for its provision than the availability of health manpower, particularly primary care physicians.

Because the provision of primary medical care is a national health priority, HSAs have been charged with the responsibility of promoting programs which will provide the necessary health manpower to deliver such care within their health service areas.[7] Two such programs are the health maintenance organizations (HMOs) which are being developed under the Health Maintenance Organization Act of 1973 (P.L. 93-222) and the health manpower recruitment and placement programs of the National Health Service Corps.

Both P.L. 93-222 and P.L. 93-641 specifically require HSAs to be involved in the development of HMOs within their health service areas. Under P.L. 93-222, an HSA in the area to be served by a prospective HMO must be given the opportunity to review the application and to submit its recommendations respecting approval of the application to the Secretary of HEW.[8] Under P.L. 93-641, HSAs are required as part of their review authority to consider the "special needs and circumstances of health maintenance organizations for which /federal/ assistance may be provided...."[9]

In the process of reviewing HMO applic[illegible] HSAs must determine if the applicant has accurately described the medically underserved areas" (MUAs) from which membership in the prospective HMO is expected. In making these determinations, HSAs are to examine the socio-demographic structure of these areas, the availability of health resources, and indicators of health status. It is specifically required that the ratio of primary care physicians to population be utilized as a measure of the availability of health resources.[10]

The objective of the National Health Services Corps (NHSC) is to recruit and place physicians, dentists, and other health professionals in areas that have been designated as "critical health manpower shortage areas" (CHMSAs). For a community to become eligible for NHSC assistance in the placement of needed health manpower, it much first meet the criteria of a CHMSA and be so designated.* The specific criterion to be used in the designation of critical medical shortage areas (CMSAs), i.e., those areas that would be eligible for the placement of primary care physicians, is the ratio of population to primary care physicians.[11]

State and areawide health planning agencies have been requested to maintain an up-to-date inventory of selected health professionals practicing in their service areas in order to assist in the application of CHMSA criteria and in the resultant identification of areas which qualify for NHSC assistance.[12] HSAs will, therefore, require an up-to-date listing of the number of primary care physicians currently practicing in their service areas in order to identify CMSAs which would be eligible for the placement of primary care physicians. Data on the number, type, and location of physicians are also essential for an HSA's effective participation in the development of HMOs. In Chapter 8 in this section, techniques for the assemblage, analysis and use of extant physician data in the promotion of such programs are described. A technique for estimating

*The responsibility for designating areas as CHMSAs and for developing the criteria for such designation belongs to the Bureau of Health Manpower, Health Resources Administration, Building 31, Room 3B06, 9000 Rockville Pike, Bethesda, Maryland 20014. "A review of the criteria and procedures for designation of shortage areas...is now underway within the Health Resources Administration as part of special efforts to enhance the compatibility of the designation criteria and procedures for the NHSC, the Loan Repayment Program, and for other health manpower-related programs as well. This review may result in significant changes in criteria and/or the procedures currently being used in these programs. When these efforts have been completed, new lists of areas proposed for designation under the various programs will be developed through the application of the revised criteria to the latest available national data (as modified based on consultation with the appropriate agencies). Such lists of proposed shortage areas will then be circulated for review."

future requirements for primary care physicians is also presented in Chapter 8.

C. INSTITUTIONAL HEALTH MANPOWER DATA

The primary purpose of the study presented in Chapter 9 of this section is to provide HSAs with a methodology for obtaining data concerning the present and future supply of specific types of health manpower in their service areas, and the present and future demand in hospitals and nursing homes for each of ese types of manpower. In addition to the HSA's mandate to collect data on the number, type and location of the area's health manpower, two specific tasks delegated to these agencies require data of the type collected in this study.

P.L. 93-641 indicates that the planning, promotion and implementation of health manpower development programs should be considered one of the major responsibilities of HSAs.[13] Development of these programs requires that the agency identify and document current shortages of health manpower and examine past, current and future production of particular types of manpower.

HSAs must consider, as a part of the process of reviewing proposed health services, the availability of health manpower necessary for the provision of such services.[14] This requires information on both the present supply of manpower, and the number of persons currently being trained for health occupations.

Chapter 7

The Sources and Uses of Hospital Data

A. SELECTION AND DEFINITION OF DATA ITEMS

A number of data items pertaining to hospitals are available to planners from publications or other sources of extant data. Among the most useful are the following:

- hospital name
- hospital address (including county and zip code)
- hospital telephone number (including area code)
- name of hospital administrator
- beds: number of beds, cribs and pediatric bassinets regularly maintained (set up and staffed for use) for inpatients during a 12-month period*; does not include bassinets for newborn infants.[1]
- admissions: number of patients accepted for inpatient service during a 12-month period; does not include newborn.[2]
- census: average number of inpatients receiving care each day during a 12-month period; does not include newborn.[3]
- occupancy: ratio of average daily census to the average number of beds maintained during a 12-month reporting period. (The number of these "statistical beds" may differ from the bed count at the close of the reporting period.)[4]
- personnel: includes persons on payroll on a specified date; includes full-time equivalents of part-time personnel but excludes medical and dental interns and residents and other trainees. Full-time equivalents are calculated on the basis that two part-time persons equal one full-time person.[5]

---

*The American Hospital Association's Guide to the Health Care Field uses a slightly different definition. Beds are defined as the "number of beds, cribs and pediatric bassinets regularly maintained at the close of the reporting period" (p. 13).

- expense: expense for a 12-month period; both total expense and payroll components are shown. Payroll expenses include all salaries and wages except those paid to medical and dental interns and residents, and other trainees.[6]

- stay:

  - short-term: average length of stay for all patients is less than 30 days (or where separate long and short stay units exist: over 50 percent of all patients are admitted to units where average length of stay is less than 30 days)

  - long-term: average length of stay for all patients is 30 days or more (or where separate long and short stay units exist: over 50 percent of all patients are admitted to units where average length of stay is 30 days or more)

- facilities: facilities actually available within the institution. (The American Hospital Association [AHA] recognizes the following 46 categories of facilities:)

  - postoperative recovery room
  - intensive cardiac care unit
  - intensive care unit
  - open-heart surgery facilities
  - pharmacy with FT registered pharmacist
  - pharmacy with PT registered pharmacist
  - x-ray therapy
  - cobalt therapy
  - radium therapy
  - diagnostic radioisotope facility
  - therapeutic radioisotope facility
  - histopathology laboratory
  - organ bank
  - blood bank
  - electroencephalography
  - inhalation therapy department
  - premature nursery
  - self-care unit
  - extended care or long-term nursing care unit
  - inpatient renal dialysis
  - outpatient renal dialysis
  - burn care unit
  - physical therapy department
  - occupational therapy department
  - rehabilitation inpatient unit
  - rehabilitation outpatient unit
  - psychiatric inpatient unit
  - psychiatric outpatient unit

- psychiatric partial hospitalization program
- psychiatric emergency services
- psychiatric foster and/or home care
- psychiatric consultation and education services
- clinical psychologist services
- organized outpatient department
- emergency department
- social work department
- family planning service
- genetic counseling service
- abortion service (inpatient)
- abortion service (outpatient)
- home care department
- dental services
- podiatrist services
- speech therapist services
- hospital auxiliary
- volunteer services department

·<u>service</u>: the type of service that the hospital provides to the majority of its admissions. (The AHA recognizes the following 23 service categories:)

- general medical and surgical
- hospital unit of an institution (prison hospital, college infirmary, etc.)
- hospital unit within a mental retardation school
- psychiatric
- tuberculosis and other respiratory diseases
- narcotic addiction
- maternity
- eye, ear, nose, and throat
- rehabilitation
- orthopedic
- chronic disease
- other specialty
- children's general
- children's hospital unit of an institution
- children's psychiatric
- children's tuberculosis and other respiratory diseases
- children's eye, ear, nose, and throat
- children's rehabilitation
- children's orthopedic
- children's chronic disease
- children's other specialty
- institution for mental retardation
- alcoholism

·<u>approvals</u>: accreditation or approval by national approving and reporting bodies. (The following approvals are received annually by the AHA:)

- accreditation by Joint Commission on Accreditation of Hospitals
- cancer program approved by American College of Surgeons
- residency approved by American Medical Association
- internship approved by American Medical Association
- medical school affiliation, reported by American Medical Association
- hospital-controlled professional nursing school, reported by National League for Nursing
- member of Council of Teaching Hospitals of the Association of American Medical Colleges
- hospital contracting or participating in Blue Cross Plan, reported by Blue Cross Association
- certified for participation in the Health Insurance for the Aged (Medicare) Program by the Department of Health, Education, and Welfare
- accreditation by American Osteopathic Association
- internship approved by American Osteopathic Association
- residency approved by American Osteopathic Association

- control: the type of organization responsible for the management and day to day operation of the hospital. (The AHA recognizes the following control categories:)

  - government, nonfederal
    - state
    - county
    - city
    - city-county
    - hospital district or authority
  - nongovernment not-for-profit
    - church operated
    - other
  - investor owned (for-profit)
    - individual
    - partnership
    - corporation
  - government, federal
    - Air Force
    - Army
    - Navy
    - Public Health Service other than Indian Service
    - Veterans Administration
    - other federal
    - Public Health Service Indian Service
    - Department of Justice
  - osteopathic
    - church operated
    - other not-for-profit
    - other
    - individual for-profit

partnership for-profit
corporation for-profit

B. DESCRIPTION OF DATA SOURCES

There are a number of data sources from which data concerning hospitals may be available. These sources include the records of public and private agencies as well as published sources of data.

All states license health facilities in some fashion and all states participate in the federally-underwritten Hill-Burton hospital construction program. Therefore, certain data describing hospitals, including hospital name and address, have been collected and stored by state level agencies. Many states maintain a considerable file of data on hospitals. In some cases these data have been published as part of state health facilities plans, directories, etc., by state health departments, statewide CHP agencies, or under other auspices. Presumably HSA staff will have access to all of these directories, files, etc. Some data on individual hospitals may also be available directly from state hospital associations and non-profit or governmental organizations at local or regional levels. An early effort should be made on the part of HSA staff to familiarize themselves with existing state and local resources for published and unpublished data on hospitals.

The following is a list of those data items which the Hospital and Medical Facilities Construction Program requires to be collected annually on inpatient facilities for state eligibility to receive funds disbursed under this program (Hill-Burton):

- name of state
- fiscal year
- name of service area
- period covered by inventory: month, day, year
- name of facility
- name of county

- name of city or town
- control:
    - nonprofit, including church affiliated
    - government
        - state
        - county
        - city
        - city-county
        - district
- licensed bed capacity: the number of beds which each facility is licensed to operate: if a facility is not covered by a licensure program, the count of beds as established by the administration of the facility.
- total bed capacity: beds under construction, both nonconforming and conforming.
- admissions or discharges: newborns are not included; statistics for psychiatric and long-term care units in general hospitals are to be reported separately under mental and long-term care categories.
- patient days

The National Center for Health Statistics is gradually implementing a national program to provide reliable and continuing data for many types of decision-making on the part of health agencies. This program is the Cooperative Health Statistics System (CHSS). The system currently consists of seven data components, including a health facilities inventory and survey component. The inventory portion of the health facilities component has been developed. It is presently being implemented in five states and is in the planning and development phase in another ten states. States with CHSS contracts to implement the health facilities inventory eventually are required to have 100 percent coverage of all health facilities for a specified minimum set of information. The minimum data set for health facilities is as follows:

- reporting period used
- months in operation during reporting period
- facility name

- address
- telephone number
- type of ownership
- type of service provided to majority of admissions
- whether admission is restricted primarily to children
- licensed bed capacity
- number of beds certified for Medicare and Medicaid (skilled and intermediate)
- number of beds currently set up and staffed for use
- current patient census
- inpatient days of care
- number of admissions
- number of discharges
- number of employees (full and part-time)
- number of hours worked by part-time employees in last 7 days
- total expense (payroll and non-payroll)
- basic monthly charge for residents or patients
- bassinets set up and staffed for use
- total number of live births
- facilities and services offered
- number of residents who received nursing care in past 7 days
- number of visits to outpatient services

The availability of data from each of the sources described above varies from state to state. State hospital associations may be able to provide a great deal of data in some states, but in other states these agencies may be unable or unwilling to supply similar data to planners. Annual collection of Hill-Burton data became a function of State Agencies as a result of P.L. 93-641. The extent to which these agencies will be able to carry out this function in coming years, remains uncertain. Finally, the CHSS health facilities inventory component is presently available in only a few states.

For these reasons, the analysis and use section, which follows, concentrates on the analysis of data from sources known to be available to all health planners. It should be reemphasized however, that an attempt should be made to become familiar with all sources of state and local data on hospitals. Where a number of sources are available (such as CHSS, state hospital associations and Hill-Burton records) data should be selected from those sources which are able to provide data which are both recent and valid.

Whatever the availability of data at the state and local level, three national sources of data are known to be available to all health planners. These national data sources include data files of the National Center for Health Statistics and two publications of the American Hospital Association (AHA).

The National Center for Health Statistics' Division of Health Manpower and Facilities Statistics sponsors a nationwide program of data collection on all inpatient health facilities in the U.S. This program is known as the Master Facility Inventory (MFI). The MFI consists primarily of the Master Facility List (MFL) which is a computerized file containing names and addresses of all known inpatient facilities; and the Master Facility Inventory Survey, now conducted annually for hospitals and biannually for nursing homes and other health facilities, which results in a data tape or printout which is usually available in the fall of the year following conduct of the Survey. Although the inventory of nursing homes and other health facilities is conducted entirely by NCHS, for the past several years AHA and NCHS have collaborated in conducting the annual survey of hospitals. Using virtually identical questionnaires, the two organizations jointly survey all known hospitals in the United States. The overall procedures are as follows:

NCHS sends its list of new hospitals or "births" to the AHA. This list results from a program known as the Agency Reporting System (ARS) which has been set up by NCHS to determine on an annual basis the names and addresses of all newly-established inpatient facilities. The ARS consists of national voluntary organizations and federal and state agencies, including health, welfare, and voluntary religious organizations; publishers of commercial directories; state agencies which administer, regulate, license, certify, approve, list, or are otherwise concerned with medical and resident care facilities; and federal agencies which administer inpatient facilities.

AHA matches this list of births, as well as its own master file of member hospitals, against the annual publication of the Federation of American Hospitals Investor Owned Directory.

AHA adds the names of verified new hospitals to its files and prints questionnaires and address labels for all known hospitals.

AHA sends questionnaires and labels for all identified non-registered hospitals, excluding osteopathic hospitals to NCHS for mailing.

AHA sends questionnaires to the American Osteopathic Association and this organization surveys those of its hospitals which are non-registered and returns the completed questionnaires to AHA for analysis.

AHA surveys all registered hospitals, including registered osteopathic hospitals.

AHA processes and analyzes data on all registered and non-registered hospitals and updates its master file with these data.

Annual Survey data tapes are generated at AHA and sent to NCHS and these tapes are used to update the NCHS Master Facility Inventory.

Based on this joint data collection activity, NCHS prepares a tape and a printout for hospitals.* The tape contains, in addition to

*NCHS also publishes these data in Hospitals: A County and Metropolitan Area Data Book. This volume provides data on number of hospitals, number of beds and admissions, personnel, ownership, etc. These data are aggregated to the county level.

hospital name and address, data on other items included in the Survey. These include ownership, type of facility, number of beds, days of care, discharges, admissions, type of service, outpatient visits, and employees. The 1974 hospital tape (73-74 data) is now available for $200 and the printout for $150. The tape and printout can be ordered from:

Scientific and Technical Information Branch
National Center for Health Statistics
5600 Fishers Lane
Rockville, Maryland 20852

The AHA produces two annual publications based on this data collection activity; the American Hospital Association Guide to the Health Care Field (the Guide) and Hospital Statistics. These documents contain all of the data items defined in Part A. above, as well as a large number of other data items about hospitals. The Guide presents these data for individual hospitals while Hospital Statistics presents data aggregated to the SMSA, state, census division and national levels.*

The data analysis and use section which follows assumes for the purpose of exposition, that planners in Central HSA have available only the three national data sources discussed above. Obviously, planners in "real" HSAs may be able to perform more sophisticated analyses, or place more confidence in their analyses, depending on the type and quality of extant data available from state and local sources.

---

*The Guide and Hospital Statistics may be ordered in or around September of each year from the AHA, Publications Department, 840 North Lake Shore Drive, Chicago, Illinois 60611. Cost of the Guide is $20.00 and cost of Hospital Statistics is $7.50. The AHA must receive a purchase order or check before it can ship.

C. DATA ANALYSIS AND USE

A number of uses of extant hospital data are possible in health planning. Planners must develop detailed descriptions of the availability of hospital beds and services in a community. Comparison of these descriptive data with similar data for other communities or larger geographic units may be helpful in identifying shortages (or duplications) of hospital beds or hospital services. Similarly, the development of data describing the distribution of beds or services within an HSA or state may help to identify medically underserved areas. Data on the cost of hospital services and the manpower required to staff hospitals may also be compared with data from other communities, or larger geographic entities in order to identify problems.

A variety of measures dealing with the utilization of services may also be calculated from extant data. These measures are useful both as a description of current patterns of the use of hospital services within health service areas and as an indication of potential problem areas.

Since planners must also determine whether presently existing facilities are being optimally used, a section is provided which deals with ways in which extant data may be employed to determine a hospital's capacity to produce service.

Finally, the determination of the number of hospital beds needed in future years is another task which is central to health planning. Several strategies for projecting bed need which employ extant data are, therefore, discussed in the last section of this chapter.

1. Descriptive Uses of the Data

Descriptive data are useful for developing detailed information about hospitals in a community and as a method for identifying possible shortages and maldistributions of beds, facilities or services. In developing statistics and indices which describe the hospital services in a given community, the planner should attempt to locate or calculate similar statistics and indices for other communities. Comparisons among these various measures are a useful, albeit crude, method of identifying problems in the availability of services in a given community. Therefore, most tables in this chapter which describe the hospitals in Central HSA also contain similar data for larger geographic entities such as states, census divisions and the nation.

The issues of availability and distribution are dealt with at several points in the following discussion. Availability refers to the presence of a given service or bed within an area. This concept is often expressed in relative terms, e.g., the number of hospital beds per 1,000 people in the area. Distribution refers to the location of hospitals and their services within the area relative to one another and to the residents of the area.

While availability and distribution are conceptually distinct, they are sometimes difficult to separate in planning. For example, the lack of hospital beds in a county may be considered a problem of availability when the county is the area under consideration. However, when the entire state is the area under consideration, the planner may conclude that sufficient services are available and that lack of services in a particular county represents a problem of distribution.

a. Availability of Hospital Beds

Planners wishing to determine whether hospital beds are available in a given area will find useful information in the AHA's Guide and Hospital Statistics. An indication of the availability of hospital beds in a state can be obtained from Hospital Statistics. Table 1 contains data for a hypothetical state indicating the number of hospitals and beds by size of hospital as they would be presented in Hospital Statistics. The number of beds reported in Tables 1 and 2 represents the number of beds regularly maintained during the reporting period (sometimes known as "statistical beds"). The Guide reports beds regularly maintained at the end of the reporting period.* Because hospitals may have added or eliminated beds during the year these numbers may differ. A method for calculating statistical beds is presented in the section which deals with measures of hospital utilization. These publications do not contain data on the number of licensed beds available in hospitals. These data may be available from the local and state agencies discussed above or from data compiled during the HSA's inventory of health facilities.

Because the Guide's definition of beds represents an estimate of beds available at the end of a reporting period it is used throughout the present discussion of availability (i.e., throughout section 1.a.). Throughout the remainder of this chapter, however, the term "beds" refers to statistical beds.

---

*The accuracy of the data (particularly the number of beds) presented in the AHA Guide and Hospital Statistics has been the subject of some debate among many health planners. Those who use the AHA data sources, therefore, should whenever possible check their accuracy against other local and/or state data sources.

Table 1

Number of Hospitals and Beds by Size of Hospital, (state), 1974

CLASSIFICATION		HOSPITALS	BEDS
(state name)............		142	36,234
6-24 beds		3	59
25-49		14	553
50-99		38	2,757
100-199		35	4,672
200-299		12	3,015
300-399		11	3,712
400-499		10	4,336
500 or more		19	17,130

Without some information about the state's population, and about the type of hospitals and beds with which the planner is dealing, the data in Table 1 are of little use. While planners must be concerned with all types of institutions providing health care in their state or health service area, the data sources discussed here are especially suited to the description and analysis of community hospitals, i.e., nonfederal, short-term general and other special hospitals. This definition of community hospitals excludes hospital units of institutions such as prisons or universities, long-term hospitals, and federal hospitals.[7]

Hospital Statistics provides data on community hospitals at the state level. The simplest measure of the availability of community hospitals is the absolute number of community hospitals and beds, which is presented in Table 2.

Table 2

Number of Community Hospitals and Beds
by Size of Hospital, (state), 1974

CLASSIFICATION		HOSPITALS	BEDS
Community hospitals......		119	23,351
6-24 beds		2	47
25-49		10	429
50-99		37	2,705
100-199		31	4,203
200-299		10	2,590
300-399		11	3,768
400-499		10	4,381
500 or more		8	5,228

A comparison of the data for community hospitals (Table 2) with the data for all types of hospitals in the state (Table 1) demonstrates the importance of specifying the type of hospital when comparing beds available in geographic areas. About 64% of the total hospital beds in the state are in community hospitals, i.e., 23,351/36,234 = .644.

While the planner's interest often centers on community hospitals, planning decisions sometimes require data on other types of hospitals as well. Fortunately, the Guide provides data on all types of hospitals. Data provided in the Guide on ownership, length of stay, and type of service enable the planners to select specific types of hospitals comparable to the categories used in Hospital Statistics or other sources of aggregate data.

Most of the discussion in this section relates to community hospitals. In order to compile data on each of the community hospitals in a health service area, the hospitals listed in the Guide which meet the present definition of community hospitals must be determined. To compile a list of community hospitals from the Guide:

list all hospitals which are located in the counties which make up the health service area served by the HSA

remove from the list all long-term hospitals (e.g., remove all hospitals coded "L" in the "Stay" column of the Guide)

remove from the list all hospitals which are not general hospitals or other special hospitals according to AHA definitions (e.g., remove all hospitals coded 11, 12, 22, 33, 42, 62 or 82 in the "Service" column of the Guide)

remove all federal hospitals from the remaining list (e.g., remove all hospitals coded 41, 42, 43, 44, 45, 46, 47 or 48 in the "Control" column of the Guide)

These data on community hospitals may now be aggregated in order to produce tables for Central HSA which are directly comparable to the tables containing state, census division and national data for community hospitals in Hospital Statistics.

Central HSA's service area contains a total of 49 hospitals. However, only 38 of these hospitals fit the present definition of community hospitals. The other 11 hospitals include three federal military hospitals, a federal psychiatric hospital, four nonfederal psychiatric hospitals, two university health centers and a large nonfederal T.B. and respiratory disease hospital. Thus, 38/49 or 77.6 percent of Central HSA's hospitals are community hospitals. These community hospitals contained 9,759 beds, at the close of the reporting period.

Tables 3 and 4 display data on the number and percent of hospitals and beds by type of hospital. Only four categories of hospitals were selected for these tables. These categories are mutually exclusive and exhaust all possible types of hospitals, thus they add to 100 percent of the hospitals (subject to rounding error). If categories were designed in such a way that a given hospital

Table 3

Number and Percent of Hospitals by Selected Types;
Central HSA, (state), (census division), United States, 1974

Type of hospital	Number of hospitals		Percent of hospitals			
	Central HSA	(state)	Central HSA	(state)	(census division)	U.S.
	(1)	(2)	(3)	(4)	(5)	(6)
community	38	119	77.5	83.8	82.3	81.8
nonfederal psychiatric	4	14	8.2	9.9	10.2	7.6
federal	4	6	8.2	4.2	3.0	5.4
all other	3	3	6.1	2.1	4.5	5.2
Total Hospitals	49	149	100.0	100.0	100.0	100.0

Table 4

Number and Percent of Beds in Selected Types of Hospitals; Central HSA, (state), (census division), United States, 1974

Type of hospital	Number of beds		Percent of hospital beds			
	Central HSA	(state)	Central HSA	(state)	(census division)	U.S.
	(1)	(2)	(3)	(4)	(5)	(6)
community	9,759	23,351	66.2	64.4	66.0	61.2
nonfederal psychiatric	2,757	9,874	18.7	27.3	23.5	25.3
federal	1,184	1,974	8.0	5.4	6.7	8.9
all other	1,035	1,035	7.0	2.8	3.8	4.6
Total Beds	14,735	36,234	99.9*	99.9*	100.0	100.0

*Numbers total less than 100% due to rounding error.

could be omitted or counted more than once it would not be appropriate to sum the numbers or percentages. For example, if Table 3 were changed to read "community, psychiatric, federal, all other," a federal psychiatric hospital like the one in Central HSA would be counted twice. Thus, the sums of the four categories would be greater than the total number of hospitals.

Percentages in columns (3), (4), (5) and (6) of Table 4 are equivalent to the fraction of the total hospital beds represented by each type. Thus, the fraction of all U.S. hospital beds (1,512,684 beds) which are, according to the 1975 edition of Hospital Statistics, contained in community hospitals (925,996 beds), is 925,996/1,512,684 (or 61.2 percent).

Percentages reported in column (3) of these tables are calculated from the numbers reported in column (1). Data in colums (4), (5) and (6) are calculated from data in Hospital Statistics.

These comparisons of Central HSA data with data for the state, census division and nation suggest that no large differences exist in the percent of hospitals or the percent of beds in each category. These tables, however, indicate nothing about the availability of beds relative to the population. They indicate only that the distributions of hospitals by type and beds by type in Central HSA are similar to analogous distributions in larger geographic areas.

Norming

In order to derive useful information, the interpretation of descriptive data requires some kind of comparison group. For example, knowing that a community has 100 hospital beds means little without also knowing the number of beds in similar communities or the size of the community's population.

Norming is a process by which single items of descriptive data may be made more interpretable, or several items of data may be made comparable.[8] Percentages, like those used in Tables 3 and 4 represent one simple type of norming. Using this technique, the number of mental hospitals in Central HSA can be directly compared with the number of mental hospitals in the state, i.e., each may be expressed as a percent of all hospitals. A type of norming frequently used in health planning involves calculation of a ratio or rate. A ratio results from the comparison of two values "by expressing one as a multiple of the other."[9]

The Bed to Population Ratio

One of the most often used measures of the availability of hospital beds is the bed to population ratio. This ratio relates the number of beds in a community to that community's population. More precisely, beds are expressed in terms of the number of beds available per 1,000 people in the community. The bed to population ratio is calculated as follows:

$$\frac{\text{number of beds}}{\text{number of people}} \times 1{,}000 = \text{bed to population ratio}$$

The bed to population ratio requires the number of community hospital beds and an estimate of the current population. The number of community hospital beds may be taken directly from Hospital Statistics for the United States, for census divisions and for individual states. The number of community hospital beds for Central HSA has previously been calculated (Table 4). Similar calculations for SMSA's, cities, counties, or other geographic units can also be compiled from individual institution data contained in extant data sources.

Calculation of the bed to population ratio for Central HSA and the state produces the following results:

statewide (community hospitals): $\frac{23,351}{5,489,700} \times 1,000$

$= 4.3$ beds per 1,000 persons

Central HSA (community hospitals): $\frac{9,759}{2,133,100} \times 1,000$

$= 4.6$ beds per 1,000 persons

These ratios indicate that the availability of community hospital beds relative to population is slightly higher in Central HSA than in the state as a whole. The planner may also wish to compare Central HSA's bed to population ratio to that of a given community or county within Central HSA. For example SMSA-C, which is a part of Central HSA, presently has 735 beds in its two community hospitals. These beds serve a county with a population estimated to be 112,200. The bed to population ratio for SMSA-C's community hospitals is:

$\frac{735}{112,200} \times 1,000 = 6.6$ beds per 1,000 persons

The bed to population ratio, while easy to calculate, has several inherent weaknesses which make it useful only as a gross measure of the availability of hospital beds. These ratios are inadequate for more specific uses, for example, decisions concerning the construction of new hospital beds.

There are several reasons why these ratios are inadequate to make this type of decision. A primary reason is that they offer no information on the need for hospital beds in the community. It is clear that certain subsets of the population use more days of hospital care than others. Thus, a community with an unusually large population of elderly persons may need more hospital beds than a community of similar size with a younger population. Other environmental and social factors may also affect the need for services. Decisions based on a comparison of bed to population ratios implicitly assume that need is constant in the communities being compared. This assumption should not be made in the absence of data on morbidity and other indices of a community's health.

A second assumption which is made when bed to population ratios are compared is that the population residing within the geographic boundaries which define the community is also the population of the service area of that community's hospitals. This assumption may not be tenable, particularly when the community analyzed is a county or city. These communities are likely to draw many patients from outside of their geographic boundaries.

In the present example, the planner might discover that five of the six rural counties which surround SMSA-C contain no hospitals. Thus, it is possible that many of the residents of these counties rely on SMSA-C hospitals. Assuming that the service area of SMSA-C hospitals is limited to the geographic boundaries of the SMSA could be a serious error. However, an estimate of the actual population served by the hospitals in a community can be made if data are available from a patient origin study (see Chapter 12 in Section IV for an example of the use of patient origin data in estimating hospital service area population).

Some planners may be willing to assume that certain communities are sufficiently similar in need for services to make comparisons of bed to population ratios appropriate. They may also assume that the number of residents leaving the community to obtain hospital care elsewhere is equivalent to the number of "outsiders" obtaining care inside the community. To the extent that these assumptions are tenable, bed to population ratios may be used to compare the relative availability of beds in several communities. For example, the original enactment of Hill-Burton legislation specified that the federal government would not finance hospital bed construction in states having a bed to population ratio greater than 4.5-5.5 general hospital beds per 1,000 population (depending on population density).[10] Planners have continued to use the bed to population ratio as an index of the possibility that communities have too many or too few beds. Thus, planners willing to make the assumptions specified above will find the bed to population ratio useful in identifying areas which are medically underserved in terms of the availability of hospital beds. These ratios are not, however, sufficient to determine the number of beds which should be added in such a community. This determination requires a considerably more detailed analysis of a community's needs.

b. Availability of Hospital Facilities, Services and Special Beds

In order to perform their mandated functions, HSAs and State Agencies will require much more information than simply the total number of available hospital beds. For example, they will need to know the extent to which various types of special beds, services and other facilities located within hospitals are available. Examples of these facilities, services and special beds include emergency rooms, intensive care units, blood banks, psychiatric inpatient units and family planning services. Knowledge of the present availability of these services is necessary for both project review and plan development functions of HSAs and State Agencies.

A number of extant data sources may be consulted in examining these questions of availability. For example, the AHA's Guide and Hospital Statistics provide detailed information on facilities, services and special beds available in hospitals. In general, categories used in these two publications are identical; however, Hospital Statistics reports the number of extended care unit beds, family planning service visits, home care department visits, intensive care unit beds (cardiac), intensive care unit beds (mixed), psychiatric services, rehabilitation services and self care unit beds. The _Guide_ reports whether or not a hospital provides these facilities, services or special beds but does not provide data on the number of beds or units of service consumed.* With these exceptions, the categories used in the AHA _Guide's_ list of 46 "Facilities" and the _Hospital Statistics_ list of 54 "Facilities, Services and Special Beds" are identical.

The comparability of these two sources facilitates the work of the planner. The number and percent of hospitals supplying a particular service in Central HSA, or a given area of the HSA, may be easily determined from the _Guide_. These data may then be compared with the data reported in _Hospital Statistics_ for the state, census division and nation.

In making these comparisons the planner should exercise caution both in calculation and interpretation. "Percent" data reported in _Hospital Statistics_ are based on the proportion of _all_ reporting hospitals which indicate that they provide the facility, service or type of bed in question. Several sources of potential error exist. In reporting data, hospitals may not adhere to the definitions of facilities, services, beds and visits which are provided with the AHA's annual survey. Planners may incorrectly calculate the percent of hospitals in their HSA which provide these facilities, services, etc. Finally, planners may misinterpret data which show a "shortage" of a specific type of service in their health service area.

The first problem mentioned above is obviously beyond the control of the planner. It is mentioned only as a reminder that careful calculation of percentages does not insure accuracy if the initial reporting is in error. The latter two problems can, however, be prevented by careful calculation and interpretation of the data.

---

*The AHA _Guide_ does report the number of a few types of special beds such as long-term care beds in short stay hospitals.

Therefore, each problem will be briefly discussed prior to presenting techniques for displaying the data.

Calculation of the percent of hospitals offering a given service or facility seems very straightforward. The planner should note, however, that the data reported in Hospital Statistics represent the percent of all hospitals in the United States, state, etc., which offer a particular facility or service. These data are not reported for community hospitals separately. Thus, the planner should be careful to use as the denominator in calculating these percentages, all of the hospitals in the geographic area to which the calculations apply. A planner in Central HSA may discover that 34 hospitals in the HSA report that they have a "postoperative recovery room." This represents 34/49 or 69 percent of all of the hospitals in Central HSA. This number is directly comparable with the percentage data reported in Hospital Statistics. The planner may also wish to calculate the proportion of community hospitals with postoperative recovery rooms. In this case, 33 community hospitals in Central HSA which report that they have a "postoperative recovery room" are counted from the Guide. These represent 33/38 or 87 percent of the community hospitals in Central HSA. There are no comparable percentages reported in Hospital Statistics. Thus, the planner should be careful to compare only percentages based on all hospitals in Central HSA to the data reported in Hospital Statistics when dealing with facilities, services and special beds.

Interpretation of the percentages may also be problematic. Facilities and services such as postoperative recovery rooms are common in community hospitals but relatively rare in psychiatric, TB, and other long-term hospitals. A community with an unusually large number of the latter type of hospital can, therefore, be expected to be lower than the nation and state in the percent of hospitals with postoperative recovery rooms.

An analogous problem occurs with facilities or services which are rarely offered, or which represent an alternative to another form of care. For example, only 3.8 percent of U.S. hospitals offer "genetic counseling services." It is not surprising, therefore, that no hospital in Central HSA offers this service. The data in Hospital Statistics reveal that only two hospitals in the state report offering this service (1.4 percent of the state's hospitals). One of these hospitals is located quite close to the boundary of Central HSA, and can, therefore, be presumed to provide this service to Central HSA residents. Thus, the lack of a hospital providing this service in the HSA does not appear to be a problem.

At the state level the small number of hospitals providing this service may be considered a potential problem. Data from Hospital Statistics indicate that 3.8 percent of the nation's hospitals and 3.7 percent of the hospitals in the census division in which our hypothetical state belongs report offering this service. Further information would be needed however, before this situation could be labeled a problem. For example, the planner would need to know the quantity of services being provided to the state's population by the two reported counseling services, whether such services were being offered by hospitals which had not reported the service to the AHA, and whether similar services were offered in the state by non-hospital sources. In some states this information may be available from the extant data sources described above. However, in some cases only the data displayed in the AHA publications will be available without primary data collection.

Table 5 represents one method of displaying data on the availability of facilities, services and special beds. This table indicates the degree to which these services and facilities are reported to be available from hospitals in Central HSA, the state, the census division and the nation. In interpreting these data the planner should remember that:

Table 5

Reported Availability of Services, Facilities and Special Beds in Hospitals, Central HSA, (state), (census division), United States, 1974

AHA Guide number	Service, facility or bed*	Number of Central HSA hospitals	Percent of all hospitals			
			Central HSA	(state)	(census division)	United States
(39)	Abortion service (inpatient)	5	10.2	10.2	11.1	17.0
(40)	Abortion service (outpatient)	5	10.2	6.0	4.6	7.5
(14)	Blood bank	36	73.4	71.2	65.4	58.8
(22)	Burn care unit	1	2.0	2.2	3.1	2.4
(8)	Cobalt therapy	10	20.4	17.0	16.6	12.2
(42)	Dental services	18	36.7	41.3	40.4	39.8
(10)	Diagnostic radioisotope facility	22	44.9	45.0	51.0	44.0
(15)	Electroencephalography	25	51.0	44.0	50.8	43.4
(35)	Emergency department	29	59.2	78.6	79.4	76.9
(19)	Extended care unit	5	10.2	12.6	12.7	12.8
(37)	Family planning service	2	4.1	5.0	6.3	8.2
(38)	Genetic counseling service	0	0.0	1.4	3.6	3.8
. . .	. . .		. . .	. . .	. . .	. . .
(20)	Renal dialysis (inpatient)	2	4.1	5.8	12.1	11.8
(21)	Renal dialysis (outpatient)	2	4.1	5.0	11.1	10.1
. . .	. . .		. . .	. . .	. . .	. . .
(36)	Social work department	27	55.1	56.0	64.5	55.6
(44)	Speech therapist services	25	51.0	39.0	31.2	23.3
(11)	Therapeutic radioisotope facility	13	26.5	25.1	26.7	21.4
(46)	Volunteer services department	30	61.2	53.6	54.3	44.0
(7)	X-ray therapy	19	38.7	39.6	38.0	29.2

* Some Facilities, Services and Special Beds were deleted from this table in order to conserve space.

ser ces actually available from hospitals may not be reported

hospitals may occasionally report services as available which are actually available only under special circumstances

services may be available from alternative non-hospital sources

These qualifications indicate that project review decisions should not be based solely on these data. They are, however, useful as indicators of potential problems. For example, an examination of Table 5 indicates two potential problems in Central HSA. Slightly less than 60 percent of Central HSA hospitals report emergency room services available. This figure is considerably less than the comparable percentages for the state, census division or the United States. A similar shortage appears to exist in hospitals providing renal dialysis. In both the inpatient and outpatient categories Central HSA hospitals report a considerably smaller percentage of hospitals offering renal dialysis than is the case with the United States or the census division.

A partial explanation for these differences may be that these two services are normally located in community hospitals. As Table 3 indicated, Central HSA has a smaller proportion of community hospitals than the nation, census division or state. Examination of the AHA Guide indicates that 27 of these 29 emergency departments are located in community hospitals. Thus, 27/38 or 71 percent of Central HSA's community hospitals have emergency departments. This percentage is still lower, however, than the percentage of all hospitals that have emergency departments in the state, census division or nation.

Another possible explanation for the lack of emergency departments concerns the size of hospitals in Central HSA. Data in Hospital Statistics indicate that the percent of hospitals having emergency departments is greatest for hospitals with 100-399 beds and decreases for larger or smaller hospitals. The planner may, therefore, wish to compare the size of hospitals in Central HSA and the state. Table 2 indicates the distribution of community hospitals by size category in the state. A similar distribution may be created for Central HSA. Comparison of the percent of hospitals in each size category in Table 6 indicates that hospital size does not seem to be a plausible explanation for the difference in available emergency departments.

While it appears that this difference is "real," many more questions remain to be answered before it is labeled a problem. It may be that Central HSA is "better" than the nation, i.e., has less duplication of emergency department services. Thus, the smaller proportion of hospitals with emergency departments may represent a cost-saving elimination of duplicate services, or a dangerous shortage of emergency services. Data on the distribution of facilities and services are needed to illuminate this problem.

Data on the distribution of facilities and services are also needed to examine the kidney dialysis situation in Central HSA. Table 5 indicates that the percent of hospitals offering dialysis is lower than that of the state, census division or nation. Extant data indicated only two hospitals which offered dialysis; however, both of these hospitals provide both inpatient and outpatient dialysis. To examine this situation,data on the location of these hospitals, and their ability to produce services are required. The following questions require further study:

What is the location of each hospital presently providing services? (in relation to population, in relation to other hospitals providing similar services in the HSA and in neighboring HSAs?)

What is the service capacity of present dialysis facilities, i.e., how many patients can be served by the dialysis equipment presently available?

Table 6

Number and Percent of Community Hospitals by Size, Central HSA, (state), (census division), United States, 1974

Size of hospital	Number of hospitals		Percent of hospitals			
	Central HSA	(state)	Central HSA	(state)	(census division)	U.S.
	(1)	(2)	(3)	(4)	(5)	(6)
6-24 beds	0	2	0.0	1.7	2.0	5.2
25-49	3	10	7.9	8.4	10.7	20.0
50-99	7	37	18.4	31.1	24.5	25.6
100-199	12	31	31.6	26.1	24.6	22.9
200-299	3	10	7.9	8.4	15.6	11.2
300-399	5	11	13.2	9.2	8.6	6.4
400-499	3	10	7.9	8.4	6.8	3.8
500+	5	8	13.2	6.7	7.2	4.8
Total hospitals	38	119	100.1*	100.0	100.0	99.9*

*Numbers do not equal 100% due to rounding error.

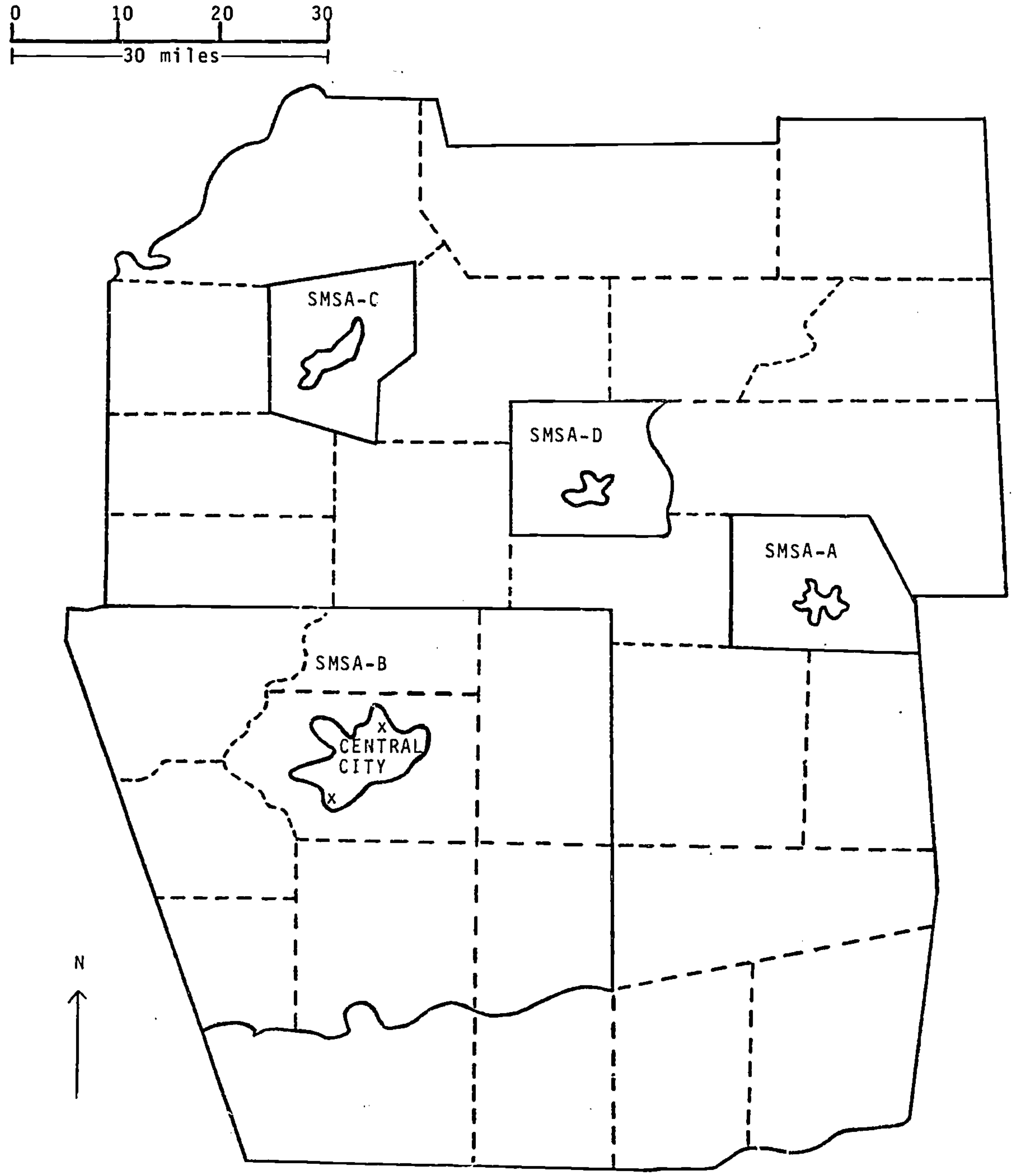
Is dialysis offered in alternative, non-hospital settings? (and if so, what is the service capacity of these alternative facilities?)

While extant data may not be sufficient to answer all of these questions, they are able to provide information on the first question. Address information in the Guide shows that both hospitals are located in Central City, which is the HSA's largest city (see Figure 1). The single largest concentration of population is, therefore, close to dialysis. However, the eight counties which make up SMSA-B represent only a little more than half of Central HSA's population. Thus, preliminary investigation based on extant data suggests that a shortage of dialysis services may exist in part of Central HSA's service area. A more complete investigation, however, is needed. Thus, the planner may wish to conduct a special study of the utilization and service capacity of kidney dialysis services in Central HSA. This study should provide answers to the three questions outlined above.

Data which relate to service capacity are particulary important. Unfortunately, data published in the Guide are not sufficient to determine the service capacity of a given type of facility, service or special bed. The Guide data indicate only whether a given type of facility, service or special bed exists in a particular hospital. The ratio of facilities, services or special beds to the population can therefore be calculated only in states where other extant data sources, which provide the needed service capacity data, are available. Hospital Statistics provides data on the number of special beds available in the state, and it is thus possible to calculate the ratio of these special beds to the state's population, and to compare these state ratios to ratios calculated for other states or for the census divisions and nation. These ratios of various types of special beds to the state's population may be useful as indicators of availability at the state level.

Figure 1

Distribution of Hospitals Providing Renal Dialysis, 1974 (x = Hospital)



Obviously, data on the availability of hospital services are crucial to the planning process. These data, however, must be combined with data on the distribution of services and beds in order to determine if a problem exists. The following section deals with the distribution issue in greater detail.

c. Distribution of Hospital Beds, Services and Facilities

The issue of the distribution of beds, etc., in the HSA is important because an analysis of the availability of services at the areawide level does not provide complete information on the ability of consumers to obtain services. Services which may seem to be available in adequate numbers may be distributed in such a way that large sections of the HSA are without locally available services.

The most straightforward method of describing the distribution of beds, services or facilities in the HSA is by way of statistical maps. By indicating on a map the precise location of each hospital, service or facility the areas in which distance from services is a potential problem can be easily identified.

Figures 2 and 3 represent maps of the distribution of hospitals in Central HSA which report offering two types of services: x-ray therapy and therapeutic radioisotope facilities. These maps indicate that there is at least one potential problem related to the distribution of services in Central HSA. Hospitals providing x-ray therapy exist in each SMSA (Figure 2), and the majority of rural counties either have, or are within 30 miles of, a hospital providing these services. This situation is not true, however, for hospitals providing therapeutic radiosotope facilities (Figure 3). No hospital in SMSA-D reports that it provides this service. Furthermore, residents of counties in the northeastern section of the HSA appear to be located far from the nearest hospital providing this facility.

Figure 2

Distribution of Hospitals Providing X-ray Therapy, Central HSA, 1974 (x = Hospitals)

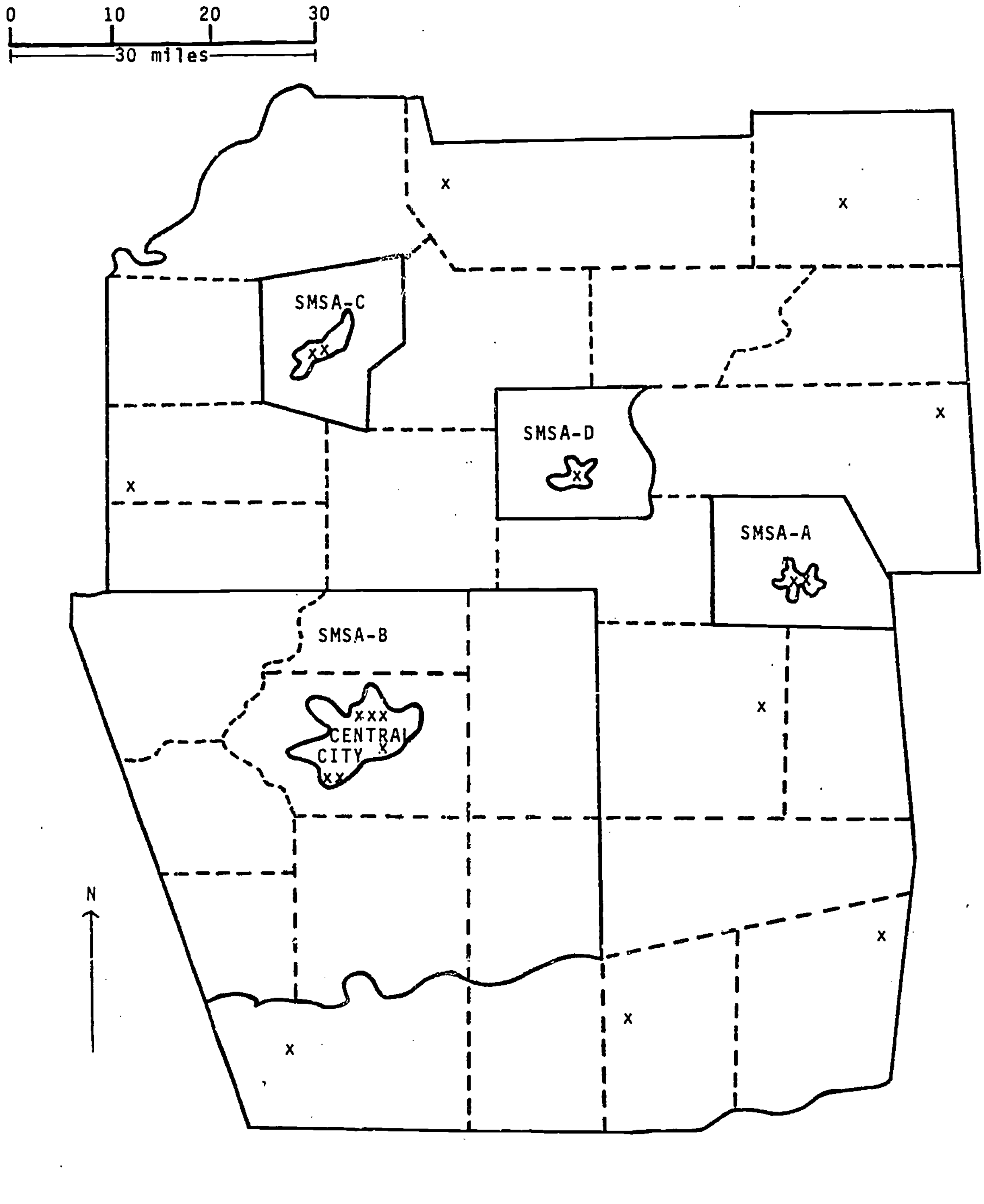
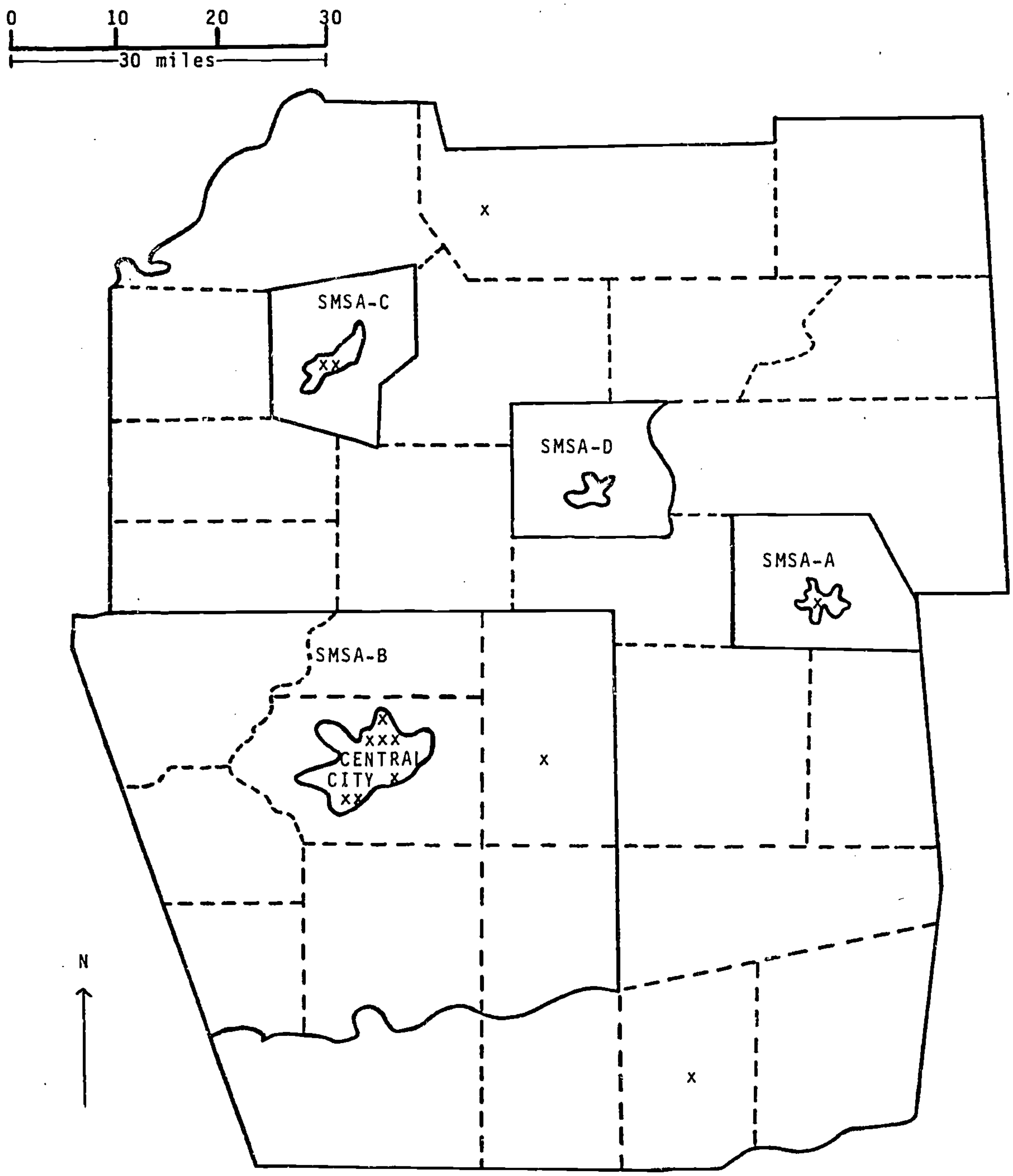


Figure 3

Distribution of Hospitals Offering Therapeutic Radioisotope Facilities, Central HSA, 1974 (x = Hospitals)



Data on the distribution of facilities may be displayed in a variety of ways. Overlays may be employed to indicate the joint distribution of two or more services. Mapping techniques may be employed to indicate areas which are beyond a specified radius from the nearest service.

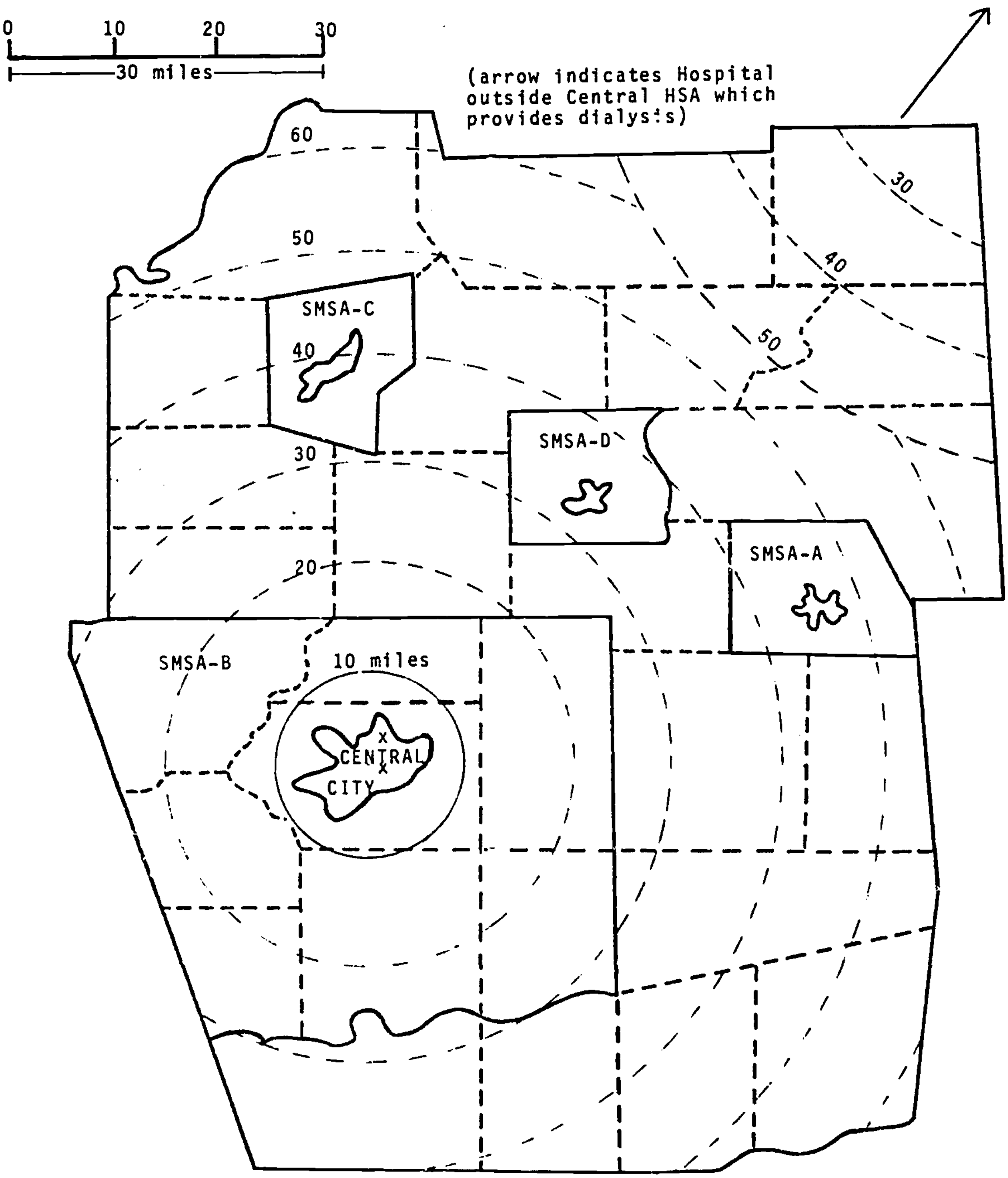
The combination of data on the distribution of services with population data is often very useful in identifying problems of distribution. Determining whether or not areas without hospital services constitute a problem is largely a function of the number of people living in those areas.

Thus, the lack of therapeutic radioisotope facilities in SMSA-D can be considered more of a problem than the lack of similar facilities in a rural county with a population of less than 5,000 persons. This situation may be expressed mathematically in several ways. The simplest expression is the number of people (in thousands) that reside outside of a specified arbitrary time or mileage radius from services, e.g., the number of persons (in thousands) residing more than 30 minutes travel time from the nearest emergency room, or the number of persons (in thousands) residing more than 50 miles from the nearest renal dialysis facility.

Figure 4 indicates that the majority of Central HSA's population lies within 50 miles of inpatient renal dialysis facilities. However, previous studies conducted by Central HSA have indicated that renal dialysis facilities are not available in sufficient numbers to meet the need of residents of the health service area. The HSA has, therefore, determined that an additional inpatient renal dialysis unit is needed in its health service area. One important consideration in selecting the location for this new facility may involve reducing travel time. Thus, it would be inappropriate to locate this facility in Central City. The planner may find it difficult to determine which of the other counties in the HSA would

Figure 4

Distance from Inpatient Renal Dialysis Facilities, Central HSA, 1974 (x = Inpatient renal dialysis facility)



provide an optimal location in terms of reducing travel for the largest number of people.

It is possible to quantify this decision by calculating an index which applies weights to the populations residing at various distances from a given facility, service or special bed. In the present example these weights represent the midpoints of the distance intervals shown in Figure 4. Thus, persons residing less than ten miles from a specific facility are assigned a weight of five, while persons living more than ten miles, but less than 20 miles from that facility are assigned a weight of 15, etc. An example of the calculation of this index is provided on the following pages for residents of the county which comprises SMSA-D. It indicates the impact of distance to inpatient renal dialysis facilities for residents of this county.

The Distance Index

In order to determine the proportion of the county's population which falls into specific distance intervals from the inpatient renal dialysis facilities, census data at the township level are used (see Figure 5). Two townships in SMSA-D fall mostly within the 20-30 mile radius, one township is mostly within the 40-50 mile area. The remainder of this county is in the 30-40 mile area. Thus, the planner calculates population as follows:

1. 20-30 miles

   township A = 2.3 thousand
   township B = 3.3 thousand
   5.6 thousand

2. 40-50 miles

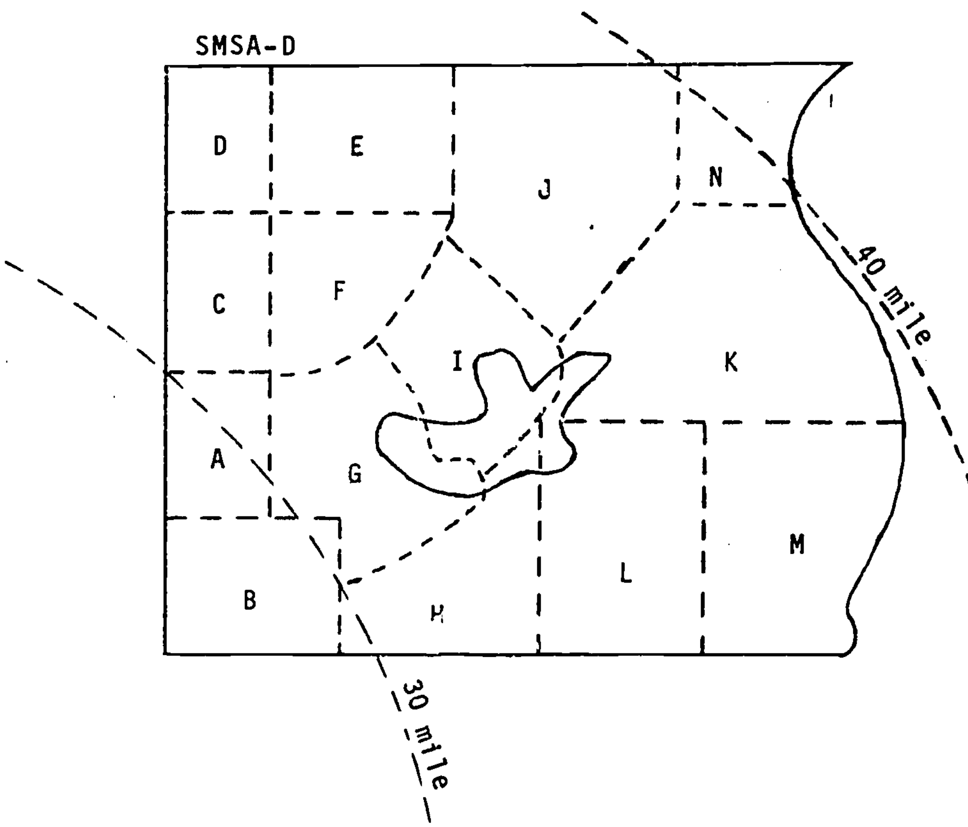
   township N = 15.4 thousand

3. 30-40 mile radius = total SMSA-D population minus townships A, B, and N

   129.2 = (2.3 + 3.3 + 15.4) = 108.2 thousand

Figure 5

Distance of Townships from Inpatient Renal Dialysis Facilities, SMSA-D



within 20-30 mile radius - townships A and B
within 30-40 mile radius - townships C thru M
within 40-50 mile radius - township N

Each of these population figures is then multiplied by the appropriate distance weight in order to calculate the distance index for SMSA-D as follows:

population	x	weight	=	index
5.6	x	25	=	140
108.2	x	35	=	3,787
15.4	x	45	=	693
		Total	=	4,620

The index for SMSA-D is then entered on Worksheet A under "county index." By calculating a similar index for each county, and adding these indices together on the worksheet, an estimate of the total impact of the present distribution of services will be indicated. The HSA can then determine the extent to which various prospective sites for the new renal dialysis facility will reduce total travel time. This can be achieved by comparing the indices that would be derived from locating the new facility at point A, point B, etc. The hypothetical location which results in the lowest index is the one which can be expected to reduce total travel time by the largest amount.

This index is dependent on the weight assigned to each distance. Since the weighting scheme is arbitrary, the location of the new facility chosen by this method may differ if weights are changed.[11]

Worksheet A

Calculation of Distance Index

Population (000's) at Specified Intervals from Renal Dialysis Services

Miles to nearest dialysis facility

Central HSA counties	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99	100-109	County Index
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11*	0	0	5.6	108.2	15.4	0	0	0	0	0	0	4,620
12												
↓												
30												
Weight	5	15	25	35	45	55	65	75	85	95	105	

Total:

*SMSA-D

2. Measures of Hospital Utilization

Most sources of data on hospitals provide a number of data items useful in describing hospital utilization. For example, the AHA's Guide provides three items of utilization data: admissions, census and occupancy. From these three data items along with others such as the number of statistical beds, numerous utilization measures may be calculated. These measures will be useful in problem identification, plan development, the projection of bed need and project review decisions.

The annual number of admissions is reported for larger geographic areas (U.S., census divisions, states, SMSAs) in Hospital Statistics. In our hypothetical state, the annual number of admissions is 859,371. It is useful to determine how many of these admissions occurred in a given community or a given type of hospital by aggregating data on individual hospitals.

Having determined the number of admissions which occurred in each community of interest, attention may be turned to comparing admissions among hospitals or communities. These comparisons may be useful in identifying potential problems related to the number of persons using inpatient hospital services or the way in which hospital facilities are used. (Thus, extant data are often helpful in identifying areas in which follow-up studies using primary data are necessary.)

For purposes of illustration, it will be assumed that Central HSA staff are interested in determining how many admissions to community hospitals occur in the state, Central HSA and in SMSA-C each year. Hospital Statistics reports 829,171 admissions to the state's community hospitals in a 12 month period. Data compiled by the HSA from extant sources show that the health service area's 38 community hospitals admit 345,342 persons annually, and that the two community hospitals in SMSA-C admit 26,235 persons annually.

Comparison of these admission data among Central HSA's communities and hospitals requires that a norming procedure be carried out. Two ratios may be calculated from these admission data. The first is the admission rate. This rate is useful whenever the size of the population of the service area of a hospital or group of hospitals is known or can be assumed. The second ratio, known as the admission per bed rate (or case flow) does not require data concerning the size of the service area's population, but, instead, expresses admissions in terms of the number of statistical beds in a hospital or community. A further discussion of each of these measures, along with others that are useful in measuring hospital utilization follows.

a. The Admission Rate

The formula for the rate of annual admissions per 1,000 persons is as follows:

$$\text{admission rate} = \frac{\text{annual number of admissions}}{\text{number of people}} \times 1{,}000$$

Using this formula, it is possible to calculate the admission rate for the state, HSA and SMSA-C. The denominators used in the calculation of these rates are the same population figures used in calculating bed to population ratios. The resulting rates are:

state: $\frac{829{,}171}{5{,}489{,}700} \times 1{,}000 = 151.0$ admissions per 1,000 persons per year

Central HSA: $\frac{345{,}342}{2{,}133{,}100} \times 1{,}000 = 161.9$ admissions per 1,000 persons per year

SMSA-C: $\frac{26{,}235}{112{,}200} \times 1{,}000 = 233.8$ admissions per 1,000 persons per year

SMSA-C appears to depart significantly from the state and the HSA. This difference might be interpreted as an indication that SMSA-C has a population which requires an unusually high number of hospitalizations per capita. However, as will be shown, this explanation seems incorrect. As noted earlier, the hospitals in SMSA-C have a service area which extends beyond the borders of the SMSA. If patient origin data are available, it is possible to more accurately estimate the population of this service area.*

In the present example, it is assumed that the population of SMSA-C hospitals' service area is 151,595. The "service area adjusted" rate of admissions per 1,000 population (using the service area population as a denominator) is:

$$\text{SMSA-C:}\quad \frac{26{,}235}{151{,}595} \times 1{,}000 = 173.1 \text{ admissions per 1,000 persons per year}$$

The service area adjusted rate for SMSA-C is only slightly higher than the admission rates for the state or HSA. The similarity of these rates suggests that the population in SMSA-C does not require an unusually high number of hospitalizations per capita relative to Central HSA or the state.

Another means of assessing the use of hospitals in SMSA-C is the development of synthetic estimates based upon census data and national estimates of hospital use. The number of admissions which would be expected from each population subgroup if national admission rates applied can be calculated, and the total number of expected admissions for a community can be estimated. The expected

*In general, the proportion of each county's population attributed to a given hospital's service area is equivalent to the proportion of the total number of patients originating from that county that utilized the hospital in question. Thus, if 60 percent of all hospitalized residents of County B use hospital X, then 60 percent of County B's population is attributed to hospital X's service area.

number of admissions can then be compared with the actual number of admissions.

An example of this type of indirect comparison, which takes into consideration the age and sex structure of SMSA-C, is presented in Table 7.

As shown in Table 7, the actual number of admissions in SMSA-C exceeds the number which would be expected for a community with SMSA-C's age-sex structure which admitted patients at the national rate (26,235 admissions as compared to 22,633 respectively). Dividing the actual admissions by the expected admissions provides one index of the use of hospital services in each community. The ratio 1.16 for SMSA-C indicates that actual admissions exceed expectations by 16 percent.

b. The Case Flow

Another use of admissions data is the calculation of the admission per bed rate or "case flow."* This rate relates the number of admissions to the number of "statistical beds" in a service area. Therefore, the first step in calculating this rate is the determination of the number of statistical beds. This measure of beds may be determined for each hospital by the following formula:

$$\text{statistical beds} = \frac{\text{average daily census}}{\text{occupancy rate}}$$

This is the measure of beds used throughout this chapter (with the exception of section 1.d.). Data on beds may now be combined with data on admissions to calculate the case flow according to the formula:

---

*Donabedian presents an analogous measure called the discharge ratio which is operationalized as "discharges per bed per year," (see reference no. 11, p. 330).

Table 7

Actual and Expected Admissions to Community Hospitals, SMSA-C, 1974

Population subgroup		Current population (000's) (1)	x	Estimated admission rate* (2)	=	Estimated admissions (3)
Males:	15	19.664		81.6		1,605
	15-44	40.123		91.0		3,651
	45-64	11.780		173.8		2,047
	65+	4.504		360.8		1,625
Females:	15	[illegible]		65.3		1,230
	15-44	37.326		216.6		8,084
	45-64	12.512		179.9		2,251
	65+	6.848		312.3		2,139

Ratio of actual admissions to expected admissions:

$$\frac{\text{Actual}}{\text{Expected}} = \frac{26,235}{22,633} = 1.16$$

*Estimated admission rates are age/sex-specific national admission rates taken from: NCHS, "Utilization of Short-Stay Hospitals" Vital and Health Statistics Series 13 Number 19, U. S. Dept. of H.E.W., Public Health Service, Health Resources Administration, Washington, U. S. Govt. Printing Office 1975.

$$\text{case flow} = \frac{\text{annual number of admissions}}{\text{number of beds}}$$

In SMSA-C, as reported in the Guide, there were 26,235 admissions to its two community hospitals, which have a total of 735 beds. The case flow for SMSA-C's community hospitals is, therefore:

$$\frac{26,235}{735} = 35.7 \text{ annual admissions per bed}$$

This case flow may be compared with similar rates calculated for other SMSAs within Central HSA, or for Central HSA, the entire state, or the nation.

Long-stay hospitals can be expected to admit fewer patients per bed in a specified time period than short-stay hospitals. Thus, in comparing one hospital with another in terms of case flows, the hospitals should be similar in terms of "stay." Similarly, comparisons between two communities in terms of their case flows will be affected by any differences in the proportion of beds which are long-stay beds in each community. It is, therefore, more appropriate to compare case flows within specific "stay" categories of hospitals even when making comparisons among communities.

c. The Occupancy Ratio

Another useful measure of the utilization of hospital services is the occupancy ratio. This measure may be taken directly from extant data sources such as the Guide, or calculated from data on average daily census (or patient days) and beds. The occupancy ratio for Central HSA hospitals may be calculated by dividing the sum of the average daily census of each hospital in the health service area by the total number of beds in the health service area.

Summing the data for the community hospitals in Central HSA indicates an average daily census of 7,263. The number of community hospital beds in the health service area is 9,759. Thus, the occupancy ratio for all Central HSA community hospitals is:

$$\text{occupancy ratio} = \frac{\text{average daily census}}{\text{number of beds}}$$

$$= \frac{7,263}{9,759} = .74 \text{ or } 74\%$$

This indicates that on an "average" day 74 percent of the health service area's hospital beds are filled. For SMSA-C, data indicate that the average daily census of its two hospitals is 291 + 284 = 575. The number of beds in SMSA-C is known to be 735. Thus, the percent occupancy of SMSA-C hospitals is:

$$\frac{575}{735} = .78 \text{ or } 78\%$$

These indices can be compared with data from Hospital Statistics which indicate a statewide percent occupancy of 77.5 percent for community hospitals. Thus, SMSA-C hospitals closely approximate the statewide occupancy level.

Occupancy data have traditionally been used as an indication of the extent to which a hospital's capacity to provide services is being fully exploited. Low occupancy indicates that a hospital usually has a large number of unfilled beds while high occupancy indicates that the hospital beds are usually nearly full. High occupancy ratios have often been cited as evidence of the need for more beds. Occupancy ratios, however, vary depending on the size of the hospital, and may be artificially high if long lengths of stay are common in a hospital. Therefore, occupancy ratios should be carefully interpreted whenever they are used in plan development or project review decisions. Occupancy ratios are discussed more fully in later sections.

d. The Length of Stay

One of the most important and commonly used measures of hospital utilization is the average length of stay, which is calculated as follows:*

$$\text{average length of stay} = \frac{\text{census} \times 365}{\text{annual admissions}}$$

Thus, for Central HSA's community hospitals the average length of stay is:

$$\frac{7,263 \times 365}{345,342} = 7.7 \text{ days}$$

This can be compared with the average length of stay calculated for SMSA-C:

$$\frac{575 \times 365}{26,235} = 8.0 \text{ days}$$

The patients admitted to hospitals in SMSA-C have longer lengths of stay than the patients admitted to community hospitals in the entire HSA. This may be explained by a variety of factors, one of which is case mix differences. Since variations among hospitals in case mix are partially related to hospital size, the planner may wish to compare SMSA-C hospitals only with other hospitals in the HSA which are in the same size category. Data published in Hospital Statistics show that the statewide average length of stay for hospitals the size of those in SMSA-C (i.e., community hospitals in the 300-399 bed range) is 8.0 days. Thus, the SMSA-C hospitals have exactly the length of stay expected of hospitals of the same size throughout the state.

*For alternative methods of calculating average length of stay, see Shoshana Folk, "Average Length of Stay in Long-term Institutions," Health Services Research, Fall 1971, pp. 251-255.

e. The Utilization Rate

Another measure of hospital utilization is the utilization rate. This rate is calculated as follows:

$$\text{utilization rate} = \frac{\text{census x 365}}{\text{population}} \times 1{,}000$$

The utilization rate is interpreted as the number of patient days (i.e., census x 365) per 1,000 persons per year. According to extant data sources, the sum of the average daily censuses for all Central HSA hospitals is 7,263. Thus, the utilization rate for Central HSA hospitals is:

$$\frac{7{,}263 \times 365}{2{,}133{,}100} \times 1000 = 1{,}242.8 \text{ patient days per 1,000 persons per year}$$

Similar utilization rates for SMSAs and counties can also be calculated from extant data. (Hospital Statistics provides the necessary data for calculating utilization rates for states, census divisions and the nation.)

The problem of a lack of correspondence between numerator (utilization) data and denominator (population) data applies here. Thus, the interpretation of utilization rates for small geographic units like counties can be misleading. In the absence of patient origin data, which can be used to correct for inflow and outflow of patients, caution should be used in interpreting utilization rates at the county or sub-county level.

A variety of other measures of hospital use may be calculated based upon the data items discussed above. These measures may be found in Aspects of Medical Care Administration by Avedis Donabedian.[12] A list of these measures and the formulae used in their calculation appears below:

- bed days per year = beds x 365
- unused bed days per year = bed days per year - (census x 365)
- turnover interval = $\frac{\text{unused bed days}}{\text{admissions}}$
- unused bed days per bed = $\frac{\text{unused bed days}}{\text{beds}}$
- average hospital cycle = $\frac{\text{bed days per year}}{\text{admissions}}$

Because several of these measures can be used in assessing the productivity of a hospital, they are discussed in the next section of this chapter on hospital productivity.

3. The Capacity of the Hospital to Produce Service

In the preceding section data on hospital characteristics were reviewed primarily in the light of their utility for describing the availability, distribution, and utilization of hospital facilities, beds, and services. The availability of hospital resources and the present demand for these as reflected by use is, therefore, assumed to have been determined by the health planner.

In this section, the available data on hospital characteristics are reviewed from the standpoint of their utility for describing and analyzing the capacity of a hospital to produce service and current service production. The primary purpose of this review is to enable the health planner to make certain resource development and project review decisions pertaining to hospital resources.

Such decisions, particularly those related to hospital bed supply, should not rest solely upon questions of the number of beds, but also upon a determination of whether hospitals are presently making optimum use of their resources. If they are not, untapped capacity exists. This capacity is capable of being translated into additional hospital service without the addition of more beds. An underlying assumption is, of course, that untapped capacity is costly, particularly if, as a result of this capacity being overlooked, new beds are added—resulting in more unused capacity and higher costs.

This section is structured in the following way: Concepts and terms which are relevant to a discussion of hospital service capacity and service load are presented and defined. A model for assessing service produced in relation to the capacity to produce service is presented and explained. Several important factors which influence a hospital's use of its capacity to produce service are discussed. Previously-introduced measures which can be constructed from the data items available from extant data are reviewed to see how they reflect variations in a hospital's use of its capacity. Problems in the interpretation and use of these measures are discussed. Finally, it is shown how the interpretation of selected measures may affect planners' decisions regarding approval of applications for the expansion of hospital facilities.

a. Concepts Related to Hospital Service Capacity

- Service Load: Service load is the amount of service a hospital actually produces by having some of its beds occupied.* This service is produced through the deployment of hospital resources.

- Service Capacity: This concept refers to the amount of service that could be produced by the hospital were its resources to be optimally deployed. The primary interest is in the unused capacity which exists in the form of unoccupied beds and inappropriately occupied beds.

- Service Reserve: A hospital's service reserve is composed of inappropriately or unproductively occupied beds, and unoccupied beds maintained as standby reserve to meet peak demand and provide for replenishment of resources between patients. The amount of standby reserve may meet normative standards, in which case it will be termed "necessary standby reserve." If it exceeds the amount provided for in estimates of necessary standby reserve it will be termed "excess standby reserve." The service reserve which consists of excess standby reserve and reserve due to inappropriately or unproductively occupied beds represents a potential increment to service load.

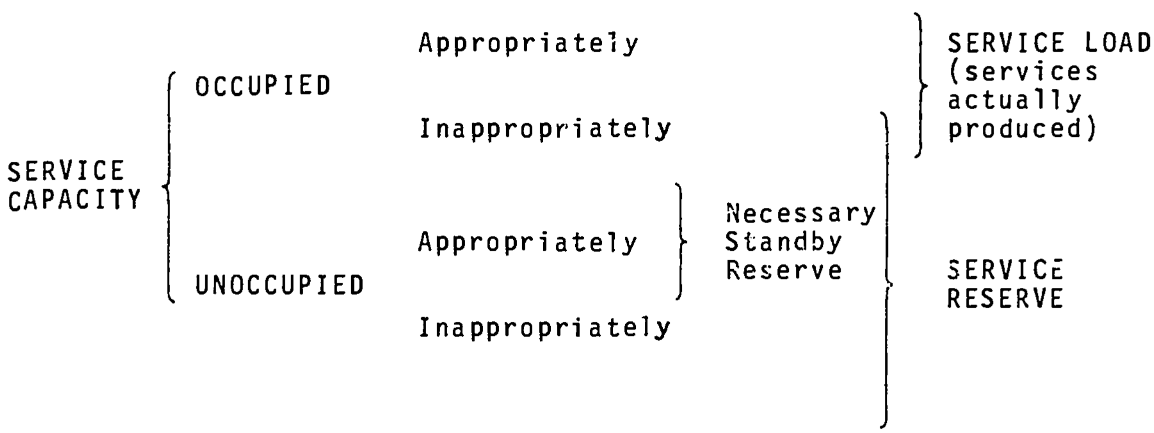
---

*"Service" is a summary term used to refer to that portion of the hospital's product which relates directly to patient care. "Services" is a term which may be used either to refer to specific facilities and services (such as an intensive care unit, cobalt therapy, etc.), or as a more concrete synonym for "service." Finally, the hospital is divided into functional and administrative units which are termed "services" (i.e., the pediatric, obstetric, or medical-surgical service). When the terms "service" or "services" have this meaning, it will be clear from the context.

- The Product: The patient day is the most commonly used measure of the hospital's product. It represents patient care actually provided. However, theoretically, there is another component of the hospital's product--its readiness to provide care or necessary standby reserve.[13]

- The Resources: Though other resources, such as manpower, are deployed in the production of hospital services, "beds" is used as a summary measure of the hospital's resources, just as "patient days" is used as a summary measure of the hospital's product.

b. A Model for the Assessment of a Hospital's Service Load Relative to Its Capacity to Produce Service*



Several features of this model require explanation. The assumption is made for purposes of simplicity that the service capacity of a hospital is uniform over time. The measures of service load available from extant data sources are also based on such an assumption (all data are reported as averages over the reporting period), and fluctuations from day to day are not apparent. In fact, it is

*This model is adapted from a more complex one presented in Donabedian (see reference no. 13).

known that both service capacity and service load, as well as demand for hospital service, fluctuate over time. These fluctuations from day to day are not apparent. In fact, it is known that both service capacity and service load, as well as demand for hospital service, fluctuate over time. These fluctuations will be discussed below under "factors affecting the hospital's maintenance of standby reserve."

Traditional measures of hospital service production such as the patient day reveal a concern only with the hospital's service load. This is that portion of overall capacity to produce that results in services actually being produced. If service reserve is not considered, the health planner must assume that all occupied beds are appropriately occupied, all unoccupied beds are appropriately unoccupied, and, therefore, that any increase in demand, such as that due to population increase, will require an increase in the supply of hospital resources, i.e., more hospital beds will be needed. The only situation in which the number of beds could legitimately be reduced would be if the population (or selected population subgroups) using the hospital or hospitals were to decline. The bed to population ratio used alone is the method of resource assessment that results from taking this perspective.

Having ascertained the current service load as represented by the number of patient days, the planner might turn to an assessment of whether all occupied beds are appropriately occupied. This points to the use of data on average length of stay as a measure of the appropriate use of resources. However, from the planner's standpoint, there are several problems in trying directly to affect changes in average length of stay. This is an issue which involves not only the organization of hospital services, but also medical judgment about quality of care (a matter which is more in the hands of the physician than the hospital *per se*). Therefore, length of stay may be less susceptible to control through planning decisions than measures which reflect other components of service reserve. Length of stay data are also useful for comparing the performance of the individual hospital or of groups of hospitals to areal norms, as described in a previous section.

There are ways other than examining length of stay to determine the appropriateness of bed occupancy. For example, patients may be utilizing a resource that provides a level of care higher than that required, such as when a patient occupying a hospital bed might more appropriately be cared for in a nursing home or elsewhere. Available data are insufficient to permit this type of an assessment of appropriateness to be made. Available data can be used in assessing the appropriateness of the deployment of unoccupied beds. The purpose of such an assessment would be to determine whether after alloting some reasonable amount of beds to standby reserve there remain inappropriately unoccupied beds which can become part of the hospital's usable service capacity. Extant data permit an analysis of the amount of standby reserve maintained by a hospital relative to the amount that should be maintained.

c. Factors Affecting the Hospital's Maintenance of Standby Reserve

Later in this chapter, the primary focus is upon the analysis of hospital service load and service capacity for the purpose of translating service reserve into usable bed capacity. It has already been noted that a portion of the hospital's service reserve might be termed "necessary standby reserve." This reserve consists of the unoccupied beds which must be maintained to accommodate peak loads and allow for patient turnover.

It will be seen that the hospital's patient care product, patient days, is related to the numbers of admissions and the average length of stay of all patients (usually in a year's time). Less well-understood is that the number of patient days which can be provided is related to how much capacity is held as standby reserve.

The current discussion focuses first upon the hospital's ability to control the admission of patients. By controlling the time of admission, the hospital is in a position to level peaks which will otherwise later cause daily census variations. Second, issues connected with the hospital's capacity to adjust to peak demand which cannot be affected by scheduling of admissions are discussed. Turnover interval, which is a component of necessary standby reserve, is the final standby reserve factor mentioned.

Controlling Time of Admission

Perhaps the most important factor influencing the relationship between a hospital's service production and its capacity to produce service is one which can be termed the non-uniformity of inflow to the hospital which results in service load fluctuations and differential demand upon capacity. Two questions of interest are: 1) To what extent is the rate of inflow susceptible to control by the hospital? 2) What practices can the hospital implement to adjust to that portion of inflow which it is not able to control? The reason for attempting to answer these questions is to suggest ways in which the hospital might deploy its resources to release unused capacity. Additional capacity may, in turn, permit the hospital to more efficiently meet fluctuations in service load.

The variable rate of inflow, or of patients presenting themselves for service to the hospital, results in variability in the number of admissions for any unit of time (hour, day, week, month). Regarding admissions whose timing is susceptible to influence by the hospital, it is relevant to review some of the findings of Anderson and Sheatsley in their study of a sample of all hospital discharges (except those of maternity patients) from 50 Massachusetts hospitals.[14]

In this study only one-third of admissions to the hospital took place under what might be called urgent conditions (hospitalization taking place on the same day that it was first recommended). Same-day admissions accounted for 15 percent of surgical cases, 51 percent of medical cases, and 42 percent of diagnostic cases.[15] In up to 85 percent of surgical cases, 49 percent of medical cases, and 58 percent of diagnostic cases, then, some flexibility in the time of admission exists. Time of admission, therefore, is potentially subject to influence by the hospital itself. Similar findings by Drosness and by London and Sigmond lend support to the contention that the hospital admitting system is amenable to administrative control.[16]

To give an appreciation of the reasons why the hospital might wish to control the timing of admissions if it is able to do so, the following research findings are of interest. In the Massachusetts study cited above, it was learned that most admissions but fewest discharges occur on Monday; that fewest admissions but most discharges, on Saturday.[17] Furthermore, the average length of stay of patients varied depending on the day of the week on which they were admitted. The average length of stay for all patients was 9.4 days; however, for patients admitted on Friday it was 11.4 days and for patients admitted on Saturday, 10.4 days.[18] Anderson and Sheatsley point out that the excess over the average length of stay resulting from Friday and Saturday admissions accounts for 3.5 percent of the total number of patient days.[19]

Were hospitals to exercise discretion over the timing of admission of patients, not only would the overall variation in admission numbers be reduced, but there is some indication that such practices would also help to reduce inappropriately long hospital stays.

Other Responses to Peak Demand

Turning from admissions whose variability might be reduced by hospital admitting practices, admissions whose variability is a given, non-controllable feature are now considered. Certain hospital practices that relate to the hospital's flexibility to handle admissions that cannot be scheduled can be examined.

Assuming that hospitals cannot control the timing of some admissions on weekends, particulary emergency or urgent admissions, it can attempt to arrange staffing patterns so that weekend admissions do not result in longer lengths of stay. An example of this would be the employment of at least some staff on the weekends who could perform necessary diagnostic or preparatory work as soon as the patient is admitted. One hospital reported in the literature on the kinds of adjustments necessary for such an adaptation and the benefits realized in terms of increasing and stabilizing occupancy.[20]

There are several other ways in which a hospital's resources can be manipulated to affect its ability to deal with fluctuations in admissions. Some hospitals have successfully used "swing beds." These swing beds can be used by different services depending upon which are experiencing the greatest demand. For example, noninfectious medical patients may alternate the use of the same beds as maternity patients. According to Donabedian, some hospitals make "provision for a number of 'flexible' beds that can be used to provide for overflow...."[21]

Another practice which increases hospital flexibility to meet increased demand is the use of more, higher level staff who can perform all the tasks of those at lower levels.[22] For example, during periods of peak demand, or when the hospital is not fully staffed, a registered nurse provides more flexibility than a practical nurse. In general, rigid, functional, hierarchical differentiation of staff reduces the hospital's ability to adjust to variation, as does departmentalization: "There may be a waiting list for beds in one department while beds go begging in another."[23]

It is of interest to note that even on a community-wide level, the organization of resources can either assist or impede the adjustment to peak demand. Donabedian points out that one can consider the existence of several small hospitals of one type in a community to be an example of departmentalization.[24] Bergwall, et al., explain how consolidation of units reduces the number of beds necessary to meet peak demand.[25]

Replenishing the Capacity to Produce Service

A certain amount of time is required between the discharge of one patient and the admission of another during which time capacity remains idle or is being replenished or reordered so that it is capable of providing service again. This period is referred to as the "turnover interval." As will be seen, certain assumptions can be made about the amount of time necessary for replenishment, and this, in turn, will have a bearing upon the assessment of the amount of necessary standby reserve maintained by a hospital.

The general reason, of course, for analyzing factors such as the above, which could contribute to an increase in the ability of the hospital to cope with periods of peak demand, is that one of the most common ways of meeting peaks has been to maintain a larger number of beds than would be necessary if other adjustments were made. This alternative has been an expensive one, since during the times when these beds are not required, the cost of maintaining them is still approximately two-thirds what it would be if they were being used.[26]

It should be acknowledged that not all of the practices mentioned in this section would be appropriate for any particular hospital to implement. In general, larger hospitals and hospitals in multi-hospital areas, have the most flexibility for adjusting the deployment of resources to meet peak demand. Furthermore, the hospital may need to yield to certain constraints, particularly with regard to staffing, such as the availability of staff in particular health manpower categories.

d. Measures of Service Load and Service Capacity

It has been pointed out that a number of measures of hospital utilization may be calculated from data items (beds, admissions, census, and occupancy) contained in the Guide. In this section, the measures which can be constructed from these items are reviewed from the standpoint of their utility for describing and analyzing a hospital's service load and service capacity. As the measures are introduced, problems associated with their interpretation and use are discussed.

Patient Days

Patient days per year is the commonly-used measure of a hospital's patient care output. There have been a number of criticisms regarding the use of this measure, particularly for comparing two or more hospitals.

First, the patient day measure represents only actual care provided but does not reflect unused capacity which exists as standby reserve. However, other measures can be used along with patient days for identifying unused capacity or standby reserve.

Second, it is claimed that a patient day "is not a useful standard unit because its specifications, in terms of the quantities and mixes of services rendered, are not uniform across hospitals or over time."[27] Standardizing for the case mix of hospitals can be done in several ways, none of which is completely satisfactory.[28] In comparisons which involve two or more hospitals of the same size and array of services, patient days may be assumed to be comparable. For broader comparisons, hospitals can be grouped by size. If this is not done, caution should certainly be exercised when comparing hospitals of varying size or service complexity.

Finally, as a standard unit a patient day does not adquately reflect heterogeneity of its consumers (age, sex, and diagnostic characteristics of patients, for example). Standardizing the patient day measure for patient characteristics is not possible using only extant data. A hospital discharge study, if available, could be used for this purpose. If a hospital's service area is known, presumptive evidence regarding users' characteristics will exist. Where these are known or assumed to be similar for hospitals, there is little difficulty in comparing the output of one hospital to that of another. In comparing groups of hospitals from widely differing geographic areas, when patient day is not standardized for patient characteristics, caution should be exercised in interpreting the results.

Bed Days

Bed days per year is the measure of a hospital's maximum service production capacity. As such, it is used along with patient days to calculate the amount of presently unused hospital capacity. Subtracting bed days from patient days yields one measure of unused capacity, i.e., unused bed days.

There are two problems associated with the interpretation and use of the bed day measure for certain purposes. Both stem from the fact that the definition of beds varies among data sources. For example, the number of beds reported in the Guide is the number available at the end of the reporting period. Difficulties can arise if this number varies substantially from the average number of beds available during the reporting period (i.e., "statistical beds"). For example, a hospital may have opened a new wing during the reporting period. The bed availability figure reported in the Guide, then, while useful for determining present capacity, is not correct for determining capacity during the study year. Fortunately, the percent occupancy reported in the Guide is based upon the average number of beds maintained during the reporting period. If occupancy is reported to be 85 percent, and the average daily census is reported to be 756, the average number of beds (889) is found by dividing the average daily census by .85. In this section, "average beds" (or "statistical beds") is used, rather than the number of beds reported in the Guide.

The second, more serious problem which exists in states where only Guide data are available, is not with the bed day measure *per se*, but with the provision of only the total bed figure in the Guide rather than figures for beds by each type of service (medical-surgical, etc.) or occupancy for various services. Particularly when estimating desirable standby reserve, it should be stressed that the optimum reserve for a unit (service, department) will vary depending upon the size of the unit, and the size of the hospital or its flexibility, as well as the nature of the demand (proportion elective admissions or those which can be scheduled) made upon it. Since data on available beds are not broken down by type of bed in the Guide, figures for standby reserve calculated from those data must be assumed to be conservative.

Average Daily Census

The average daily census represents the number of beds occupied every day averaged over a year's time. It is used to calculate average length of stay and occupancy ratio, and to calculate the appropriate amount of standby reserve which should be maintained by the hospital. The number of beds maintained on the average each day during the year minus the average daily census gives another measure of unused capacity.

Average Length of Stay

Average length of stay is an important measure, since, along with numbers of admissions, it directly influences the number of patient days a hospital produces in any given period of time. In other words, patient days can increase if either length of stay increases or if the number of admissions increase. Increasing length of stay would increase patient days, but would not necessarily reflect a desirable conversion of hospital resources.

> Long stays...raise occupancy and reduce the time during which the bed is vacant between patients at high levels of overhead. If cost is used as the measure of input and the patient-day as the measure of output, it is possible to create a spurious picture of high productivity or efficiency by simply lengthening stay.[29]

This suggests the need for other measures of efficiency than cost per patient day (such as cost per admission), and, more important for present purposes, suggests the need to examine some measure of case flow, such as admissions per bed per year.

Case Flow (Admissions per bed per year)

"Case-flow rate is the basic measure of the intensity with which hospital capacity is used."[30] It is useful as a summary measure of service production or service load in comparing one hospital's change in production over time, in comparing production among same-size hospitals, or in determining case flow for a community or other health planning area. In each case, if the number of beds remains constant, a higher rate is associated with a greater number of admissions. Also, variations in the case flow measure, when other factors remain constant, reflect variations in the use of service capacity.

According to Feldstein, substantial savings in cost per case can be achieved by increasing case flow.[31] For present purposes, since extant cost data have been questioned as to their adequacy,[32] the presumption will simply be made that decreased costs per case are associated with increased case flow. It is reasonable to assume that a response to increased demand which consists of adding more beds is a more expensive alternative than that of increasing case flow.

Occupancy Ratio

As has been pointed out, occupancy ratios may vary depending on whether "statistical beds" or beds available at the end of the reporting period are used in calculations. Occupancy ratios are often used to compare service load with service capacity. The occupancy ratio is inadequate for this purpose since two hospitals with differing admission or case flow rates may have identical occupancy ratios. The discussion of length of stay indicates why this is so. Average daily census reflects a relationship between admissions and average length of stay. If either increases, average daily census, and thus occupancy ratio, will also increase.

In a previous section there was discussion of the need for reducing fluctuations in the daily census; this would also stabilize occupancy ratios over time and permit hospitals to staff for more uniform demand on service. The ultimate aim, however, is to add to the potential for increasing occupancy without significantly increasing resource input. If this is done by increasing admissions rather than length of stay, clearly the more desirable objective from the productivity standpoint, increased occupancy _would_ reflect a more productive deployment of resources and a release of additional service capacity.

Turnover Interval

Turnover interval is a measure of the number of unused bed days per admission. A hospital's turnover interval ("idle bed time") reflects the average amount of time between patients and also the amount of standby reserve maintained by the hospital to meet peak demand. Adjustment of the turnover interval (usually downward) to meet some normative standard of the maximum time that a bed should remain idle while being replenished (for example, one day), has the effect of reducing standby reserve and bringing about an increase in usable service capacity. There are other ways to determine appropriate standby reserve besides adjusting the turnover interval.

Standby Reserve

In addition to maintaining a sufficient number of beds to meet current and future demand for service, the hospital must also calculate the number of beds which are necessary to maintain an appropriate amount of standby reserve. The sum of the two numbers is, then, the total number of beds which should be maintained by the hospital in order to meet overall demand for beds. Various methods have been used to determine the appropriate amount of standby reserve. These include: adjusting the turnover interval as indicated above; adjustment of the average daily census by applying a planned occupancy factor; adjustment based on a presumed Poisson distribution of daily censuses,* and a "rule of thumb" approach very like the Poisson approach, developed by the Commission on Hospital Care.[33]

*The suggestion that the distribution is Poisson was apparently first made by M. S. Blumberg, "Oahu's Requirements for Hospital Facilities," Menlo Park, Calif.: Stanford Research Institute, 1956. This article is referenced by Drosness, ref. no. 16; Donabedian, ref. no. 11, p. 498, references another article by Blumberg: " 'DPF Concept' Helps Predict Bed Needs," The Modern Hospital 97: 75-81, p. 170 (December 1961).

It appears from a study conducted by Drosness, et al., of daily census distributions in all 12 acute care hospitals in Santa Clara County, California,[34] that a normal, rather than a Poisson distribution more adequately describes the pattern of hospitals' daily censuses. For the medical/surgical units, the Poisson and normal distributions were "virtually the same at the 99th percentile--the tail end of the curve that gives the approximate reading for [exceeding capacity only] one day in a hundred."[35] That is, if it is acceptable for a hospital to exceed its capacity once in every 100 days, either the normal or Poisson-based estimates would be effective for determining adequate bed capacity for these units.

It would be most accurate for a hospital to study daily census variation in each of its units and base reserve estimates on this. However, extant data, where daily census is averaged, allow only the use of the Poisson-based estimate.* Using a Poisson estimate requires that the square root of the average daily census be multiplied by 2.33. This yields an estimate of standby reserve which is sufficient to meet demand 99 percent of the time.

One final qualification is appropriate. The number of patient care subunits within the hospital will affect the appropriateness of the standby estimates. (The more subunits there are within a hospital the higher will be the estimate of the number of needed beds.)** Furthermore, the more flexibility within the hospital there is (e.g., swing beds, use of empty beds in one service to accommodate suitable patients from another, etc.), and the more flexibility within the community as a whole (e.g., facilities can exchange patients with other similar facilities in the area), the fewer the number of standby beds needed.

---

*The reason for this is that the standard deviation of the daily census must be calculated, if daily census follows a normal distribution; however, for a Poisson distribution the standard deviation equals the square root of the mean.

**For a further discussion of this issue see David F. Bergwall, et al., Introduction to Health Planning, Washington, D.C.: Information Resources Press, 1974, pp.170-171.

e. Analysis and Use of Service Load and Service Capacity Data in Project Review and Resource Development Decisions

In this section example data, as well as measures discussed in the previous section which can be calculated from these data, will be presented: 1) for two community hospitals in County A of Central HSA (the only ones located in this county), and 2) for seven large community hospitals in SMSA-B of Central HSA.

In the first example, differences between the two hospitals with respect to service load and service capacity which are evident from an inspection of the measures will be discussed. Next, two types of project review decisions in which use of service load and capacity data are appropriate given the mandated responsibilities of HSAs under P.L. 93-641 are considered.

In the second example, a resource development decision is introduced in which service load and capacity data describing SMSA averages on several measures are used to establish norms. The seven community hospitals are then measured against these standards.

The first four columns of Table 8 on the following page display data similar to those contained in the Guide for the two hospitals in County A. The remainder of the measures presented in Table 8 are calculated from these data as follows:

1. column E (bed days) = column A (beds) x 365

   Hospital #1: 135,050 = 370 x 365
   Hospital #2: 135,050 = 370 x 365

2. column F (average length of stay) = $\frac{\text{column C (ADC) x 365}}{\text{column B (admissions)}}$

   Hospital #1: 7.7 = $\frac{297 \times 365}{14{,}067}$

Table 8

Service Load and Service Capacity Data for the Two Community Hospitals in County A, Central HSA

(based upon AHA *Guide* data for 1974)

	A	B	C	D	E	F	G	H	I
	beds	admiss.	ADC	% Occ.	bed days	ALOS (days)	pat. days	turnover (days	case flow
Hospital #1	370	14,067	297	80.3	135,050	7.7	108,316	1.9	38.0
Hospital #2	370	12,058	281	76.0	135,050	8.5	102,493	2.7	32.6

Hospital #2: 8.5 $= \frac{281 \times 365}{12,058}$

3. column G (patient days) = column B (admissions) x column F (length of stay)

Hospital #1: 108,316 = 14,067 x 7.7

Hospital #2: 102,493 = 12,058 x 8.5

4. column H (turnover interval) = $\frac{\text{column E (bed days) - column G (patient days)}}{\text{column B (admissions)}}$

Hospital #1: 1.9 $= \frac{135,050 - 108,316}{14,067}$

Hospital #2: 2.7 $= \frac{135,050 - 102,493}{12,058}$

5. column I (case flow) = $\frac{\text{column B (admissions)}}{\text{column A (beds)}}$

Hospital #1: 38.0 $= \frac{14,067}{370}$

Hospital #2: 32.6 $= \frac{12,058}{370}$

Beds and Bed Days

Since no change in beds occurred during the year for either Hospital #1 or Hospital #2, the percent occupancy in column D accurately reflects the average occupancy, and the bed day measure in column E accurately reflects the maximum service capacity in each hospital, both during and at the end of the reporting period. Nothing is known about the percentage of beds allocated in either hospital to the patient care subunits within them.

Table 9 makes the relationship between the two hospitals on each measure clear in an arithmetic way. This table expresses the percent by which Hospital #1 is greater or less than Hospital #2 on each measure. This percentage difference is calculated by dividing

Table 9

Relationship of Hospital #1 to Hospital #2 on Selected Service Load and Service Capacity Measures (expressed as percent differences)

ADMISSIONS

$$\frac{\text{Hospital \#1: } 14{,}067}{\text{Hospital \#2: } 12{,}058} \times 100 = 116.7\% \qquad \begin{array}{r} 116.7\% \\ -\ 100.0\% \\ \hline 16.7\% \end{array} \text{ difference}$$

AVERAGE DAILY CENSUS

$$\frac{\text{Hospital \#1: } 297}{\text{Hospital \#2: } 281} \times 100 = 105.7\% \qquad \begin{array}{r} 105.7\% \\ -\ 100.0\% \\ \hline 5.7\% \end{array} \text{ difference}$$

OCCUPANCY RATIO (percent)

$$\frac{\text{Hospital \#1: } 80.3}{\text{Hospital \#2: } 76.0} \times 100 = 105.7\% \qquad \begin{array}{r} 105.7\% \\ -\ 100.0\% \\ \hline 5.7\% \end{array} \text{ difference}$$

PATIENT DAYS

$$\frac{\text{Hospital \#1: } 108{,}316}{\text{Hospital \#2: } 102{,}493} \times 100 = 105.7\% \qquad \begin{array}{r} 105.7\% \\ -\ 100.0\% \\ \hline 5.7\% \end{array} \text{ difference}$$

AVERAGE LENGTH OF STAY

$$\frac{\text{Hospital \#1: } 7.7}{\text{Hospital \#2: } 8.5} \times 100 = 90.6\% \qquad \begin{array}{r} 90.6\% \\ -\ 100.0\% \\ \hline -\ 9.4\% \end{array} \text{ difference}$$

TURNOVER INTERVAL

Hospital #1:	1.9		70.4%
		x 100 = 70.4%	- 100.0%
Hospital #2:	2.7		
			- 29.6% difference

$$\frac{1.9}{2.7} \times 100 = 70.4\% \qquad 70.4\% - 100.0\% = -29.6\% \text{ difference}$$

CASE FLOW

$$\frac{38.0}{32.6} \times 100 = 116.6\% \qquad 116.6\% - 100.0\% = 16.6\% \text{ difference}$$

(Hospital #1: 38.0; Hospital #2: 32.6)

the measure for #1 by the measure for #2, multiplying the result by 100, and subtracting 100 from the resulting product.

Admissions, Case Flow

Hospital #1 had 16.7 percent more admissions in 1974 than Hospital #2. Since they have the same number of beds, the case flow (admissions per bed per year) is also greater for #1 than for #2. Both service load and use of capacity are greater for Hospital #1.

Occupancy Ratio, Average Daily Census, Patient Days

The hospitals differ from each other in service load and use of service capacity on all three of these measures to the same degree, as can be seen from Table 9. Occupancy ratio is based on average daily census. The reason that Hospital #2's average daily census is as high as it is, is that patients stay longer, rather than that more patients are admitted. Since the relative difference in patient days in the two hospitals is so small due to the longer stays in Hospital #2, the patient day measure does not adequately reflect differences in the two hospitals' use of capacity.

Using occupancy ratios, daily census, or patient days, then, to measure variation in hospitals' use of capacity is not sufficient. These hospitals differ more in their respective deployment of resources than the differences in these measures reveal.

Length of Stay

The difference in length of stay between the two hospitals is nearly 10 percent. This difference is one of the factors that makes the patient day measure of the hospitals' respective outputs differ. The remainder of the difference in patient days is due to the greater number of admissions for Hospital #1. Even if length of stay <u>and</u> numbers of admissions are considered, there is still another way in which use of capacity between the two hospitals varies.

Turnover Interval

Hospital #1 has a turnover interval which is 29.6 percent less than that for Hospital #2. The turnover interval measure reflects the greatest difference in use of capacity. Included in the turnover interval is the standby reserve of each hospital. Hospital #1 presently maintains an average of 73 unused beds (370 beds minus the ADC of 297). Using 2.33 $\sqrt{ADC}$, 40 beds would be a sufficient reserve. Hospital #2 presently maintains an average of 89 unused beds (370 beds minus the ADC of 281). Its standby reserve should be 39. However, these figures may be too conservative since the beds in one unit are not necessarily interchangeable with the beds in another.

If Hospital #1 continues to have the same number of admissions and average length of stay but maintains a standby reserve of 40 rather than 73 beds, the bed complement can be reduced by 33 beds to 337 beds. This would affect the turnover interval as follows:

$$\frac{\text{bed days - patient days}}{\text{admissions}} = \frac{(337 \times 365) - 108{,}316}{14{,}067} = \text{1.0 days turnover}$$

If Hospital #2 continues to have the same number of admissions and average length of stay but maintains a standby reserve of 39 rather than 89 beds, the bed complement can be reduced by 50 beds to 320. This would affect the turnover interval as follows:

$$\frac{(320 \times 365) - 102{,}493}{12{,}058} = \text{1.2 days turnover}$$

A second possibility, however, is that the hospitals maintain 370 beds but use the released standby capacity to serve increased admissions, as follows:

Hospital #1:

1. present unused bed days = 73 x 365 = 26,645
2. unused bed days with standby of 40 beds = 40 x 365 = 14,600
3. difference between 1 and 2 (bed days released for additional admissions) = 12,045
4. increment to admissions = $\frac{12,045}{7.7\text{(ALOS)}}$ = 1,564
5. 4 plus present admissions = 1,564 + 14,067 = 15,631

Revised turnover = $\frac{135,050 - (15,631 \times 7.7)}{15,631}$ = .9 days turnover

Revised case flow = $\frac{15,631}{370}$ = 42.2 admissions per bed (per year)

Hospital #2:

1. present unused bed days = 89 x 365 = 32,485
2. unused bed days with standby of 39 beds = 39 x 365 = 14,235
3. difference between 1 and 2 (bed days released for additional admissions) = 18,250
4. increment to admissions = $\frac{18,250}{8.5\text{ (ALOS)}}$ = 2,147
5. 4 plus present admissions = 2,147 + 12,058 = 14,205

Revised turnover = $\frac{135,050 - (14,205 \times 8.5)}{14,205}$ = 1.0 days turnover

Revised case flow = $\frac{14,205}{370}$ = 38.4 admissions per bed (per year)

Project Review Decisions

Two examples of project review decisions in which service load and service capacity data could be used will be presented. In the first example, Hospital #2 requests that it be permitted to add beds to one of its units. P.L. 93-641 states that the primary responsibility of HSAs is to promote the development of services and facilities that not only meet identified needs, but also "reduce documented inefficiencies."[36] In examining the present operation of Hospital #2, it appears that, given a turnover interval of 2.7 days, the hospital is not presently making optimal use of its resources to provide service.

While data on beds, admissions, and turnover for the unit for which beds are requested would be useful, the HSA planner could conclude that by improving overall use of its capacity, the hospital can meet present, and even increased, demand for service. This could be done by converting beds from another service, by making use of swing beds as discussed earlier, or by adopting a policy of scheduling elective admissions more carefully. Were any or all of these approaches to be taken, the turnover interval, which reflects the hospital's average unused capacity, would be shortened, thus releasing capacity to accommodate periods of peak demand, as well as overall increases in demand. Of course, if the hospital can show a substantial increase in admissions since the reporting period for the most recent extant data, or otherwise document demand which cannot be met given present capacity, these factors will be considered in making the decision.

In a second example, both Hospital #1 and Hospital #2 request to add a service neither presently offers, a burn care unit. Assuming that the need for such a service in County A has been documented, the HSA planner should decide which of the two hospitals ought to establish the new unit, since "unnecessary duplication of services" is to be avoided, according to P.L. 93-641.[37]

In general, and assuming no substantial change in the most recent extant data, Hospital #1 is using its capacity to provide service at a higher rate than Hospital #2 (i.e., a case flow of 38 admissions per bed and 32.6 admissions per bed, respectively). Therefore, Hospital #1 might be expected to provide the new service as well in a manner which would be more efficient and less costly than would Hospital #2.

It is of interest to note that were Hospital #2 to justify its longer average length of stay (for example, by presenting data on case mix), the discrepancies in use of capacity between the two hospitals might need to be re-evaluated. For example, with only a slight increase in admissions (say to 13,000), Hospital #2 would produce more patient days than Hospital #1 (110,500 for #2 compared to 108,316 for #1). However, the case flow rate for Hospital #1 (38.0) would remain higher than the case flow rate for #2 (35.1). Therefore, in some circumstances, both measures (case flow and patient days) must be examined to give an accurate picture of a hospital's relative use of capacity.

Resource Development Decisions

Resource development standards for a health service area or other health planning area can be based upon extant service load and service capacity data. These standards may then be used to set resource development goals, as well as implemented through project review decisions.

The particular standards set may vary from one area to another. The standards chosen for community hospitals in SMSA-B, Central HSA include:

1) hospitals' admission per bed ratios (case flow) should equal or exceed the average for the SMSA, and

2) turnover intervals should equal or be below the average for the SMSA, and

3) excess standby reserve maintained should be no greater than the average excess standby reserve for all hospitals in the SMSA.

When hospitals improve their performance as reflected by selected service load and capacity measures, areal averages will, of course, change. For hospitals with already high performance on these measures, standards may be based on state, rather than SMSA or HSA, norms. In Maryland, for example, optimal admission per bed ratios are set according to present performance, with higher standards applied to hospitals already above the area or county average.[38]

The following procedure can be followed in order to calculate SMSA-B averages for standards 1), 2), and 3) above (refer to Table 10 for appropriate data):

Part I: Determine Average Case Flow for SMSA-B Hospitals

Step 1: From the AHA Guide, abstract, for each community hospital in SMSA-B, the number of admissions, the average daily census, and the percent occupancy.

Step 2: Find the average number of beds maintained by each hospital during the reporting period by dividing the average daily census by the percent occupancy/100.

Step 3: Sum the admissions (Step 1) to calculate the total admissions for SMSA-B.

Step 4: Sum the average number of beds maintained by each hospital (Step 2) to calculate the average number of beds maintained in SMSA-B.

Step 5: Determine the admission per bed ratio (case flow) for SMSA-B by dividing the number of admissions (Step 3) by the average number of beds maintained (Step 4).

Part II: Determine Average Turnover Interval for SMSA-B Hospitals

Step 1: Calculate the patient days for each hospital in SMSA-B by multiplying the average daily census (Step 1 Part I) by 365.*

Step 2: Calculate bed days for each hospital by multiplying the average number of beds (Step 2, Part I) by 365.

Step 3: Calculate the turnover interval for each hospital by the following formula:

$$\frac{\text{bed days - patient days}}{\text{number of admissions}}$$

Step 4: To determine the average turnover nterval for SMSA-B hospitals, sum the bed days for e ch hospital (Step 2, Part II) and subtract from these the summed patient days for each hospital (Step 1, Part II). Divide the result by the total number of admissions for SMSA-B hospitals (Step 3, Part I).

Part III: Determine Average Excess Standby Reserve for SMSA-B Hospitals

Step 1: Calculate optimal standby reserve for each hospital by multiplying 2.33 by ADC (Step 1, Part I).

Example, using data for Hospital #1, Table 10:

$$\text{optimal standby reserve} = 2.33 \times \sqrt{ADC}$$
$$= 2.33 \times \sqrt{925}$$
$$= 71 \text{ beds}$$

Step 2: Calculate average standby reserve maintained by each hospital by subtracting the average daily census (Step 1, Part I) from the average number of beds maintained (Step 2, Part I).

Example, using data for Hospital #1:

*This is an alternative method of calculating patient days to the one presented earlier (admissions times average length of stay).

average standby reserve = average number of beds - ADC

= 1064 - 925

= 139 beds

Step 3: To calculate the percent excess standby reserve for each hospital (i.e., the percent of average standby reserve currently maintained by each hospital which is in excess of the optimal standby reserve required by each hospital), divide the result of Step 1 by the result of Step 2, multiply by 100 and subtract the result from 100.

Example, using data for Hospital #1:

$$\text{percent excess standby} = 100 - \left(\frac{\text{optimal standby}}{\text{average standby}} \times 100\right)$$

$$= 100 - \left(\frac{71}{139} \times 100\right)$$

$$= 49\%$$

Step 4: To determine the average optimal standby reserve for all SMSA-B hospitals, sum the optimal standby reserve for each hospital (Step 1).

Step 5: To determine the average standby reserve maintained by all SMSA-B hospitals, subtract the total average daily census for the SMSA from the total average number of beds maintained by hospitals in SMSA-B.

Step 6: To calculate the average percent excess standby maintained by SMSA-B hospitals, divide the result of Step 4 by the result of Step 5 (average total optimal standby ÷ average total average standby), multiply by 100 and subtract the result from 100.

The results of the steps described in Parts I, II, and III are presented in Table [illegible] on the following page for all hospitals in SMSA-B which are larger than 300 beds (N = 7). These hospitals, because of their relatively large size, can be expected to have the most flexibility to implement policies which would result in a more effective use of their capacity to provide service. It can be seen in the table that five hospitals have case flow ratios which exceed

Table 10

Selected Service Load and Service [illegible]acity Measures for Large Community Hospitals in S[illegible]-B, Central HSA

(based upon A[illegible]de data for 1974)

	beds	admiss.	ADC	% Occ.	bed days	pat. days	ALOS	turn.	case flow	excess standby
Hospital #1	1064	37,740	925	86.9	388,360	337,625	8.9	1.3	35.5*	49%
Hospital #2	657	15,343	422	64.2	239,805	154,030	10.0	5.6	23.4	80%
Hospital #3	656	23,203	537	81.9	239,440	196,005	8.4	1.9	35.4*	55%
Hospital #4	579	18,123	476	82.2	211,335	173,740	9.6	2.1	31.3	50%
Hospital #5	436	15,594	354	81.2	159,140	129,210	8.3	1.9	35.8*	46%
Hospital #6	403	14,042	361	89.4	147,095	131,765	9.4	1.1	34.8*	5%
Hospital #7	327	12,996	283	86.5	119,355	103,295	7.9	1.2	39.7*	11%
TOTALS	4122	137,041	3358		1,504,530	1,225,670		2.0	33.2	54%

the SMSA average for large hospitals, i.e., 33.2 (these are starred).

Of these five, all also have turnover intervals below the SMSA average for large hospitals, i.e., 2.0. However, only four of these hospitals (#1, 5, 6, and 7) maintain an excess standby reserve which is less than the average of 54%. Therefore, the fifth hospital (#3) does not meet all three criteria. Such criteria can be used in the review of the hospitals' applications for expansion conducted by the HSA.

The application of standards s h as those presented in Table 10 encourages ospitals to respond to increased demand for service by tapping unused capacity, rather than by expanding their bed complement. This alternative is likely to serve at least one of the goals of HSAs as set forth in P.L. 93-641: "restraining increases in the cost of providing...health services."[39]

4. Bed Need Projections

The estimation of the need for hospital beds and services in future years has been one of the foremost areas of concern for health planners. While it is desirable to base this crucial determination on detailed utilization data which have been collected directly from each hospital, this is not always feasible. Planners often have to establish a projection of the beds needed in some future year based solely on the limited data available from extant sources. This section presents several strategies which use these extant data to project the future need for hospital beds.

The formula most commonly used over the last two decades for projecting the need for hospital beds is the Hill-Burton formula. This basic model along with several modifications of it, is discussed first. Two alternatives to the Hill-Burton formula are then discussed. Examples of the application of each formula in estimating the need for community hospital beds in a rural county in Central HSA (County 30) are employed in the discussion.

a. The Hill-Burton Formula

Calculation of the basic Hill-Burton formula includes four steps:

Step 1: estimate the county's current utilization rate

$$\text{current utilization rate} = \frac{\text{patient days}}{\text{population}} \times 1{,}000$$

For County 30, the current utilization rate may be calculated from census and hospital utilization data. County 30's one hospital, "Hospital A," has an average daily census of 208. This average daily census may be multiplied by 365 to find the number of patient days needed in Step 1. With the number of patient days and the population, Step 1 for County 30 may now be applied:

$$\text{current utilization rate} = \frac{208 \times 365}{80{,}700} \times 1{,}000 = 940.77 \text{ patient days per 1,000 persons}$$

Step 2: estimate the county's future utilization in terms of the total number of patient days

future utilization (patient days) = current utilization rate x (projected population/1,000)

It is assumed that the population projection for 1980 for county 30 is 86,311.

$$\text{future utilization} = 940.77 \times \frac{86{,}311}{1{,}000} = 81{,}199 \text{ patient days}$$

Step 3: estimate the county's projected average daily census

$$\text{projected average daily census} = \frac{\text{future utilization}}{365}$$

$$\text{projected average daily census} = \frac{81{,}199}{365} = 222.46$$

Step 4: apply a planned occupancy factor to estimate bed need

$$\text{bed need} = \frac{\text{projected average daily census}}{\text{planned occupancy}}$$

In this case, it is assumed that community hospitals will have an occupancy ratio of .85. Thus the bed need for this county is:

$$\text{1980 bed need} = \frac{222.46}{.85} = 261.7 = 262 \text{ beds}$$

The 236 licensed beds in Hospital A would therefore be considered inadequate to meet 1980 needs on the basis of the Hill-Burton formula.

Several aspects of the Hill-Burton formula deserve consideration. These include the assumption that the current utilization rate will (or should) continue in the future, the assignment of a standard (.85) occupancy ratio to every community hospital regardless of size, and failure to take into account the origin of patients. The latter problem is similar to the problem noted in the calculation of the admission rate and the bed to population ratio, i.e., the size of the county population used in computing Steps 1 and 2 above may not be the same as the size of the population residing in the hospital's medical ser rea.

Since many HSAs will not ha al origin data available, this section will not deal with the estimation of hospital service area population. An example of the use of patient origin data in the Hill-Burton formula is, however, included in Chapter 12 of Section IV.

b. Utilization-Adjusted Hill-Burton Formula

The basic Hill-Burton formula assumes that the current utilization rate will remain constant over time. It is possible, however, to adjust the current utilization rate to reflect changes over time before performing the Hill-Burton calculations. One method for predicting future utilization is the trend-line extrapolation approach. This approach was found to be a reasonably accurate method of predicting future utilization by Feldstein and German who described d empirically evaluated several methods of projecting bed need.[40]

In the utilization-adjusted method, future utilization rates are estimated on the basis of past trends in the number of patient days per 1,000 population. The data required include population estimates and the average daily census for each year in the last decade. The average daily census for past years may be available from a number of sources including records of the state Hill-Burton agency or past issues of the Guide and its predecessor the "Guide Issue" of Hospitals, Journal of the American Hospital Association. These publications can be found in most large libraries.

Worksheet B on the following page demonstrates how the trend in utilization can be calculated. Patient days in column (2) are calculated from the average daily censuses for past years, while population data in column (3) are taken from population estimates for past years. By expressing the population in thousands, the utilization rate displayed in column (4) is calculated by dividing column (2) by column (3). Column (5) is calculated by subtracting the previous year's utilization rate from that of the present year. Thus, the change in utilization from 1964 to 1965 is equal to 831.26 - 839.04 = -7.78. The percentage change shown in column (6) is calculated by expressing each year's change as a percentage of the previous year's utilization rate. For example, the change of -7.78 represents -7.78/839.04 = -.0093 = -.93% of the previous year's utilization rate.

The 1.35% average annual increase derived from Worksheet B may be used to adjust the utilization rate used in the 1980 bed need projection calculated above. Applying this annual increase to the observed 1974 data indicates the following utilization rates for the next six years:

year	utilization rate from previous year	x 1.0135 = (101.35%)	projected utilization rate
1975	940.77	x 1.0135 =	953.47
1976	953.47	x 1.0135 =	966.34
1977	966.34	x 1.0135 =	979.39
1978	979.39	x 1.0135 =	992.61
1979	992.61	x 1.0135 =	1,006.01
1980	1,006.01	x 1.0135 =	1,019.59

The 1980 utilization rate of 1,019.59 would, therefore, be substituted into Step 2 of the Hill-Burton formula. Calculation of Steps 2-5 would proceed as follows:

Worksheet B

Utilization of Hospital A, County 30, 1964-74

Year	Patient days	Population (000's)	Utilization rate (2) ÷ (3)	Change in utilization	Percentage change
(1)	(2)	(3)	(4)	(5)	(6)
1964	54,719	65.216	839.04		
1965	55,247	66.462	831.26	-7.78	-.93%
1966	56,582	67.732	835.38	4.12	.50%
1967	56,795	69.025	822.82	-12.56	-1.50%
1968	56,116	70.344	797.74	-25.08	-3.05%
1969	56,780	71.687	792.05	-5.69	-.71%
1970	56,001	73.056	766.55	-25.50	-3.22%
1971	54,255	74.452	728.72	-37.83	-4.94%
1972	55,755	75.876	734.82	6.10	.84%
1973	67,384	77.800	866.12	131.30	17.87%
1974	75,920	80.700	940.77	74.65	8.62%

Mean Percentage Change = +1.35%

Step 2: estimate the county's future utilization in terms of the number of patient days

future utilization (patient days) = projected (1980) utilization rate x (projected population/1,000)

$$\text{1980 utilization} = 1{,}019.59 \times \frac{86{,}311}{1{,}000} = 88{,}002 \text{ patient days}$$

Step 3: estimate the county's projected average daily census

$$\text{projected average daily census} = \frac{\text{future utilization}}{365}$$

$$\text{projected 1980 average daily census} = \frac{88{,}002}{365} = 241.1$$

Step 4: apply a planned occupancy factor to estimate bed need

$$\text{bed need} = \frac{\text{projected average daily census}}{\text{planned occupancy}}$$

$$\text{1980 bed need} = \frac{241.1}{.85} = 283.65 = 284 \text{ beds}$$

When utilization trends are incorporated into the basic Hill-Burton formula, the number of beds needed in 1980 in county 30 increases from 262 to 284. This utilization-adjusted method will normally result in slight increments or decrements to the bed need projection derived from the basic Hill-Burton formula. Occasionally, however, this method will result in bed need projections which are very inaccurate. This problem occurs whenever the last decade's trend in utilization has been the result of forces which cannot reasonably be expected to continue. An example should help to clarify this issue:

> County X is located in a poor rural area and has traditionally had very low utilization. In 1964 the utilization rate was 553.2. Due to the construction of a new hospital and the migration of several physicians into the county the utilization rate has increased over the last ten years at an annual rate of 7.5 percent. The 1974 utilization rate was approximately 1,140.2. This rate of utilization is still somewhat below Central HSA's rate of 1,242.8. However, if the 7.5 percent annual increase is projected to

1985, a utilization rate of 2,350 would result. Projections of 1985 bed need for county X which use a projected utilization rate of 2,350 are likely to result in a gross over-estimate. Thus, it is suggested that upper and lower limits be placed on projected utilization rates. These limits may be based on observed utilization rates in similar areas of the state which are believed to have an appropriate number of beds available or on other norms established by the HSA.

c. Utilization/Occupancy-Adjusted Hill-Burton Formula

The basic Hill-Burton formula utilizes an occupancy ratio of .85 for Hospital A. However, data in Hospital Statistics indicate that occupancy ratios in community hospitals throughout the nation increase with the size of the hospital. The lowest occupancy ratio (.495) occurs in community hospitals with less than 25 beds. Data from Hospital Statistics for the state in which county 30 is located show the same positive association between occupancy and the size of the hospital.

From these data it is found that hospitals in the 200-299 bed category have the following occupancy ratios:

nation:	.775
census division:	.782
state:	.785

Relative to these ratios, the planned occupancy ratio of .85 may be unrealistically high. Incorporating a planned occupancy ratio of .80 into the utilization-adjusted Hill-Burton formula for county 30 results in the following change in Step 4:

Step 4: apply the planned occupancy factor to estimate bed need

$$\text{bed need} = \frac{\text{projected average daily census}}{\text{planned occupancy}}$$

$$\text{1980 bed need} = \frac{241.1}{.80} = 301.38 = 302 \text{ beds}$$

Thus, when an occupancy-adjusted ratio of .80 is incorporated into the utilization-adjusted Hill-Burton formula, the number of beds needed in county 30 in 1980 increases from 284 to 302 beds.

Whenever bed need projections include a number of small hospitals, some adjustment of occupancy ratios may be necessary since it may be unreasonable to require a thirty bed hospital to maintain a .85 occupancy ratio before adding new beds. For example, an examination of Table 6 indicates that 22 of Central HSA's 38 community hospitals presently have less than 200 beds. Therefore, a bed need projection for all of Central HSA might include an adjusted occupancy ratio of less than .85. An adjusted occupancy ratio can be calculated as follows:

Central HSA Community Hospitals

	hospitals	hospital size (beds)	number of patient days, 1974	proportion of all patient days	state occupancy ratio
	(1)	(2)	(3)	(4)	(5)
	0	6-24	0	0	.457
	3	25-49	32,872	.0124	.616
	7	50-99	152,697	.0576	.667
	12	100-199	522,766	.1972	.715
	3	200-299	205,982	.0777	.785
	5	300-399	483,541	.1824	.816
	3	400-409	392,877	.1482	.812
	5	500+	860,260	.3245	.821
Total:	38	--	2,650,995	1.0000	--

The proportion of Central HSA patient days occurring in each hospital size category appears in column (4). Approximately 73 percent of patient days presently occur in hospitals with 200 or more beds. These hospitals can reasonably be expected to exhibit occupancy ratios in the .80-.85 range. Hospitals in the 100-199 bed category may be unable to maintain an occupancy ratio this high, and hospitals of less than 100 beds clearly cannot be expected to have an occupancy ratio of .85 if state occupancy ratios are to constitute the norm. The data on 1974 proportioned patient days may be used in estimating 1980 patient days by hospital size. Total 1980 patient days are projected to be 2,916,095 in Central HSA. It is assumed that the proportion of these patient days occurring in each hospital size category remains constant from 1974 to 1980. Hospitals in the 25-49 bed category can, therefore, be expected to account for .0124 of these patient days. The total number of patient days for each size category is, therefore:

hospital size (beds)	proportion of 1974 patient days	number of expected 1980 patient days
(1)	(2)	(2) x 2,916,095
6-24	0	0
25-49	.0124	36,160
50-99	.0576	167,967
100-199	.1972	575,054
200-299	.0777	226,581
300-399	.1824	531,895
400-499	.1482	432,165
500+	.3245	946,273

It is now possible to relate the number of patient days expected in each hospital size category with a planned occupancy ratio which is appropriate to this size category. These planned occupancy ratios may be based on present occupancy ratios in the HSA, or some other norm which is intended to represent an appropriate occupancy ratio for hospitals of a particular size. These norms may be based on national or regional ratios or may result from a review of literature regarding hospital size.

In the present example, hospitals in the 6-24 and 25-49 bed categories are assigned occupancy ratios of .65. Occupancy ratios of

.75 are selected for hospitals with at least 50 beds but less than 200 beds. Hospitals with 200 to 299 beds are expected to maintain an occupancy ratio of .80 while hospitals with 300 or more beds are expected to maintain a .85 occupancy ratio. These occupancy ratios all represent increases over present statewide ratios, but are considered feasible for hospitals of each size category. The use of these occupancy ratios on Worksheet C results in a 1980 bed need projection of 9,801 beds for Central HSA. This can be compared to the estimate resulting from the use of a .85 occupancy ratio for all Central HSA hospitals. If no adjustments to occupancy ratio are made in Central HSA, the bed need projection is:

Step 3: $\text{projected average daily census} = \frac{2{,}916{,}095}{365} = 7{,}989.3$

Step 4: $\text{bed need} = \frac{\text{average daily census}}{\text{occupancy}} = \frac{7{,}989.3}{.85} = 9{,}399 \text{ beds}$

The use of adjusted occupancy ratios results in an increase of 402 beds in the projected 1980 bed need for Central HSA. This is a sizable difference, and is indicative of the importance of considering adjusted planned occupancy in projecting bed need.

d. The Normile Model

One alternative to the Hill-Burton formula which has been used in projecting bed need is the Normile model. The Normile model is based on the assumption that the incidence of the underlying biological need for hospital care is a random event in the general population and, therefore, follows a normal, Gaussian, distribution in probability of occurrence. The model additionally assumes that the day of week and time of year when patients are admitted are random.[41]

These assumptions are obviously not met equally well by all types of hospital services. For example, obstetrical admissions may

Worksheet C

Bed Need Calculation Adjusting for Occupancy Ratio, Central HSA, I

Number of hospitals	Hospital size (beds)	Number of expected 1980 patient days	Projected average daily census (3)/365	Planned occupancy ratio
(1)	(2)	(3)	(4)	(5)
0	6-24	0	0	.65
3	25-49	36,160	99.1	.65
7	50-99	167,967	460.2	.75
12	100-199	575,054	1,575.5	.75
3	200-299	226,581	620.8	.80
5	300-399	531,895	1,457.2	.85
3	400-499	432,165	1,184.0	.85
5	500 +	946,273	2,592.5	.85

Total 38

Total

Worksheet C

d Calculation Adjusting for Occupancy Ratio, Central HSA, 1980

spital size (beds)	Number of expected 1980 patient days	Projected average daily census (3)/365	Planned occupancy ratio	Bed need (4)/(5)
(2)	(3)	(4)	(5)	(6)
6-24	0	0	.65	0
25-49	36,160	99.1	.65	153
50-99	167,967	460.2	.75	614
100-199	575,054	1,575.5	.75	2,101
200-299	226,581	620.8	.80	776
300-399	531,895	1,457.2	.85	1,714
400-499	432,165	1,184.0	.85	1,393
500 +	946,273	2,592.5	.85	3,050

Total: 9,801 beds

indeed be based on an underlying biological need for services, and obstetrical admissions may therefore occur at random times of the day, week and year. However, it is clear that this assumption is not met in medical or surgical wards where a large proportion of admissions are scheduled days or weeks in advance. Planners may therefore wish to restrict use of this model to specific types of beds for theoretical reasons. Nevertheless, tentative studies indicate that this model produces reasonably accurate bed need projections for services in which admissions do not represent random events.[42]

As a result of random variation, the probability of any given patient occupying a hospital bed on a particular day may be estimated by dividing the hospital's average length of stay by 365. This probability is symbolized by "P." By subtracting P from 1, the probability of any given patient not being in the hospital may also be calculated. This probability is symbolized by "Q" ($Q = 1-P$).

The standard deviation around the average daily census may then be calculated according to the formula:

$$\sqrt{NPQ}$$

(where N = number of admissions in a year)

When N is large this binomial distribution will approximate a normal distribution. Thus, tables of the standard normal distribution may be used to express multiples of the standard deviation in terms of confidence intervals.

Using this information one can predict at a given level of confidence that a patient can be placed in a bed on a certain number of days of the year without resorting to administrative or selective discharges. Stated differently, at the 99 percent confidence level, no more than one day out of one hundred will the patient not be placed in a bed.[43]

An application of this formula to county 30 will help to clarify this procedure. Calculation of the standard deviation of the average daily census requires data on future utilization in terms of patient days and admissions. Future patient days and projected average daily census have already been calculated for county 30 in Steps 1, 2, and 3 of the Hill-Burton formula. The future utilization was calculated to be 81,199 patient days. The projected average daily census was calculated at 222.46.

Step 4: estimate the county's current admission rate

$$\text{current admission rate} = \frac{\text{admissions}}{\text{population}} \times 1{,}000$$

Extant data sources indicate that current annual admissions to county 30's one hospital equal 9,703. Thus, the current admission rate is:

$$\frac{9{,}703}{80{,}700} \times 1{,}000 = 120.24 \text{ admissions per } 1{,}000 \text{ persons}$$

Step 5: estimate the county's future utilization in terms of future admissions (N)

$$\text{future admissions} = \text{current admission rate} \times (\text{projected population}/1{,}000)$$

$$\text{future admissions} = 120.24 \times \frac{86{,}311}{1{,}000} = 10{,}378 = N$$

Step 6: estimate future average length of stay

$$\text{average length of stay} = \frac{\text{patient days}}{\text{admissions}}$$

(In the present example, current average length of stay will equal future average length of stay because the same population projection was used in calculating both future patient days and future admissions.)

$$\text{future average length of stay} = \frac{81,199}{10,378} = 7.82$$

Step 7: calculate the probability that a given patient will be in the hospital on any given day (P)

$$P = \frac{ALS}{365} = \frac{7.82}{365} = .0214$$

Step 8: calculate the probability that a given patient will <u>not</u> be in the hospital on any given day (Q)

$$Q = 1-P = 1-.0214 = .9786$$

Step 9: calculate the standard deviation (σ) around the county's projected average daily census

$$\text{standard deviation} = \sqrt{NPQ}$$

$$\sqrt{10,378 \times .0214 \times .9786} = \sqrt{217.34} = 14.74 = \sigma$$

Step 10: estimate the beds needed in county 30 in 1980 (at the .99 confidence level)

$$\text{bed need} = \text{projected average daily census} + 2.33\ \sigma$$

$$1980 \text{ bed need} = 222.46 + 2.33\ (14.74) = 256.8 = 257 \text{ beds}$$

A higher or lower confidence level may be chosen. For example, an estimate at the .90 level of confidence adds 1.28σ to the projected average daily census, which would result in a smaller estimate of bed need [222.46 + 1.28 (14.74) = 241.33]. A table of "areas under the normal curve," which is contained in most elementary statistics texts, indicates the level of confidence associated with each σ.

Approximation of σ in the Normile Model

Because of the difficulty of calculating NPQ, an approximation of σ is often used in the Normile model. This approximation of σ is simply the square root of the projected average daily census ($\sqrt{ADC}$).[44] In the present example, $\sqrt{ADC} = \sqrt{222.46} = 14.92$. With this approximation, it is possible to project bed need as follows:

$$\text{bed need} = ADC + 2.33\sqrt{ADC}$$

(where ADC = average daily census)

This formula assumes that $2.33\sqrt{ADC}$ accounts for approximately 99 percent of the variation around the average daily census.

Application of this formula to county 30 produces the following estimate of bed need:

1. projected 1980 average daily census = 222.46 (from Step 3 of the Hill-Burton formula)
2. 1980 bed need = $222.46 + 2.33\sqrt{222.46}$ = $222.46 + 34.75 = 257.21 = 257$ beds

This simpler computational method produces the same estimate of bed need for county 30 as the Normile model, i.e., 257 beds.

Like the Hill-Burton formula, the Normile model assumes that the rate of utilization remains constant. Thus, the Normile model can also be adjusted for observed changes in utilization rates in the same way as was done in the utilization-adjusted Hill-Burton formula.

The simplest method of adjusting for utilization trends involves substituting the projected 1980 average daily census from Step 3 of the utilization-adjusted Hill-Burton formula. The bed need projection which results from a utilization-adjusted Normile model is:

$$241.1 + 2.33\sqrt{241.1} = 277.3 = 277 \text{ beds}$$

e. The Commission on Hospital Care Method

A similar method of estimating hospital bed need is the method developed by the Commission on Hospital Care. This method is intended to provide a hospital with sufficient beds to meet day to day and seasonal variations in demand for beds. This method assumes that under normal conditions hospitals will have sufficient beds to meet demand when:[45]

$$\text{bed need} = \text{ADC} + 3\sqrt{\text{ADC}}$$

Projection of ADC requires assumptions about trends in utilization. A simple application with no data on utilization trends, and a utilization-adjusted model are presented below:

simple model:

$$1980 \text{ bed need} = 222.46 + 3\sqrt{222.46} = 267.20 = 267 \text{ beds}$$

utilization-adjusted model:

$$1980 \text{ bed need} = 241.1 + 3\sqrt{241.1} = 287.68 = 288 \text{ beds}$$

f. Summary of Bed Need Projections

Each of the bed need projection methods presented above provides a slightly different estimate of the beds needed in county 30. All of the methods take into consideration data on projected 1980 population, and current use rate. They differ in assumptions concerning changes in use rates and appropriate occupancy ratios for various sizes of hospitals. These differing assumptions result in the following distribution of bed need projections for county 30:

Method	Bed Need
Hill-Burton Formula	262
Utilization-Adjusted Hill-Burton Formula	284
Utilization/Occupancy-Adjusted Hill-Burton Formula	302
Normile Model (.99 confidence level)	257
Utilization-Adjusted Normile Model	277
Commission on Hospital Care Method	267
Utilization-Adjusted Commission on Hospital Care Method	288

All of these methods indicate that the present 236 licensed beds in Hospital A of county 30 will not be sufficient to meet the projected demand for hospital beds in 1980. A minimum projection of 257 beds is indicated by the Normile model, and a maximum projection of 302 beds is indicated by the utilization/occupancy-adjusted Hill-Burton formula.

Based upon the above projections (and assuming that Hospital A has no excess standby reserve), the addition of beds to Hospital A seems justifiable. In the present case, the Normile model indicates that an addition of 21 beds are needed to meet the projected demand of the county's population (257 - 236 = 21). However, the utilization/occupancy-adjusted Hill-Burton method suggests that 66 additional beds are needed (302 - 236 = 66).

The decision regarding the actual number of beds Hospital A will be permitted to add is obviously a function of many factors, including economies of scale in adding new beds. Application of the above bed need projections provides, however, a reasonable range within which the bed addition should fall, i.e., from 21 to 66 additional beds.

Other methods for projecting bed need besides those discussed here have been developed. Certain of these methods require data that are not generally available, such as daily hospital census, occupancy ratios by type of bed (e.g., obstetrical, medical-surgical, pediatric etc.), size of waiting lists and the health status of the population. A discussion of these methods can be found in the following sources:

1) Avedis Donabedian, Aspects of Medical Care Administration: Specifying Requirements for Health Care. Cambridge: Harvard University Press, 1973, Chapter V.

2) Mara Minerva Mellum, Assessing the Need for Hospital Beds: A Review of Current Criteria. Minneapolis, Minnesota: Inter Study, December 1975.

3) School of Public Health and Tropical Medicine, Tulane University, Reference Manual for Project Review Standards and Criteria, January 7, 1974, Appendix V.

4) David F. Bergwall, Philip N. Reeves and Nina B. Woodside, Introduction to Health Planning, Washington, D.C.: Information Resources Press, 1973, Chapter 9.

Chapter 8

The Sources and Uses of Physician Data

A. SELECTION AND DEFINITION OF DATA ITEMS

For purposes of this chapter, <u>physicians</u> are defined as all licensed federal and non-federal doctors of medicine (M.D.) and osteopathy (D.O.), including interns and residents, who are engaged in the practice of medicine or surgery in a health service area. Physicians who reside in a health service area but are inactive are not included. Also excluded are those active physicians who reside in a health service area but who practice in another service area. Chiropractors are not included in this definition of physicians.

In order for HSAs to develop and maintain an up-to-date listing of their area's physicians, it will be necessary to obtain the names and addresses of each physician as defined below:

- <u>name</u>: last name, first name, and middle initial
- <u>address</u>: the location of the physician's practice including the street number, street name, city, state and zip code

The experiences of local health planning agencies have shown that in many health service areas the only way to determine the exact number of physicians practicing in these areas is to have a listing of their names and addresses. Address-based lists of physicians will also be needed by HSAs wishing to participate in the National Health Service Corps (NHSC) program. For those areas seeking identification as critical medical shortage areas (CMSAs) which are "not defined in terms of entire counties, a road map of the proposed service area and surrounding areas must be submitted, showing the boundaries of the proposed service area and the approximate location of all primary care physicians' offices...."[1]

HSAs will need more than the names and addresses of physicians in their service areas, however, in order to participate in the HMO and NHSC programs. Each of these programs have specific criteria regarding the types of physicians that are to be included in the identification of medically underserved areas (MUAs) and CMSAs respectively. These criteria will be discussed later in this chapter. The specific data items for each physician required by these criteria that are available from extant sources are as follows:

- <u>federal employment status</u>: whether or not the physician is on active duty in one of the military services (U.S. Air Force, Army or Navy) or employed by the U.S. Public Health Service or Veterans Administration. For this data item, as well as those which follow, each HSA will need to determine the number and type of data item sub-categories best suited for its planning and resources development purposes. The selection of these sub-categories will be dependent in part upon the unique characteristics of each HSA's service area. The sub-categories suggested here are those which are suitable for purposes of participation in the HMO and NHSC programs. They are:

    non-federal

    federal
        U.S. Public Health Service
        other

- <u>primary specialty</u>: the particular field of medicine in which the physician spends the largest percentage of his/her professional time. The 68 physician (M.D.) specialties currently recognized by the American Medical Association and the 58 physician specialties (D.O.) currently recognized by the American Osteopathic Association are presented in Appendix A. Only six physician specialties are required for purposes of MUA and CMSA designation, however. These are:

    general practice

    family practice

    internal medicine

    pediatrics

    obstetrics and gynecology

    general surgery

    other

• <u>professional activity</u>: the type of activity in which the physician spends the largest percentage of his/her time. The sub-categories of professional activity (sometimes referred to as "type of practice") needed for purposes of MUA and CMSA designation are:

- active
  - direct patient care
  - training (interns and residents)
  - other (teaching, research, and administration)
- inactive (retired, semi-retired, permanently disabled, temporarily not in practice, other)

• <u>age</u>: the physician's year of birth

B. DESCRIPTION OF DATA SOURCES

The principal sources of extant data concerning physicians are the American Medical Association (AMA) and the American Osteopathic Association (AOA). About every three to four years the AMA publishes the American Medical Directory, which contains a listing of all physicians known to the AMA who possess a degree of Doctor of Medicine.* The 1973 issue of the American Medical Directory is comprised of four volumes or parts. Part I contains an alphabetical listing of all physicians. Parts II, III and IV contain a geographical listing of all federal physicians arranged alphabetically within each branch of service and non-federal physicians arranged alphabetically within each state and city. Contained in this Directory are the following data on each physician:

- name: last name, first name and middle initial
- AMA membership: the use of bold face type for physician's name indicates AMA membership
- address: the professional mailing address (i.e., the address at which the physician receives medically related materials) including the street number, street name and postal zip code**
- year of birth

---

*Some state officials and health researchers have recently raised the question of the completeness of the AMA's Directory, particularly in regard to foreign medical graduates. For a discussion of these issues, see American Medical Association, American Medical News, Chicago, September 8, 1975; and Joel C. Kleinman, et. al., "Physician Manpower Data: The Case of the Missing Foreign Medical Graduate," Medical Care, Vol. 12, no. 11, November 1974, pp. 906-917.

**In many instances, a physician's "professional mailing address" listed in the AMA Directory will not be the same as the location of the physician's practice.

- *medical education number*: a multi-digit code indicating the state (or country in the case of foreign medical graduates) and medical school in which the physician received his/her M.D. degree, and the year in which the degree was received

- *year of license*: the year of license in state of present professional mailing address

- *primary specialty*: the field in which the physician indicates his/her major practice (i.e., the particular field of medicine in which the largest percentage of professional time is spent)

- *secondary specialty*: another field in which the physician indicates he/she has a limited practice

- *type of practice*: the type of activity in which the largest percentage of the physician's professional time is spent; the sub-categories include, training (intern or resident), direct patient care, administration, medical teaching (with or without some patient care), medical research (with or without some patient care), other activities (with or without some patient care), inactive (retired, semi-retired, permanently disabled, temporarily not in practice, not active for other reasons), and no classification

Copies of the *American Medical Directory* can usually be found in medical school and university libraries. It may also be purchased from the AMA. The cost of the four-volume 1973 issue of the *Directory* is $125.00.

An updated list of the physicians and information contained in the 1973 issue of the *Directory* can also be obtained from the AMA. Information concerning updated listings of the physicians located in a particular state or group of counties, including costs, can be obtained by writing to the Center for Health Services Research and Development, AMA, 535 North Dearborn Street, Chicago, Illinois 60610.*

---

*Any requests for physician information must meet with the approval of the AMA. Requests for physician listings must be accompanied by statements concerning the specific uses that will be made of the lists. Requests can generally be met within three months, but can take as long as six months.

A listing of all osteopathic physicians known to the AOA is published annually in its Yearbook and Directory of Osteopathic Physicians. The 1975 issue contains the name, address, date of birth, specialty, type of practice, medical school and year of graduation of each federal and non-federal physician. Copies of the Yearbook and Directory of Osteopathic Physicians can sometimes be found in medical school and university libraries. Copies of the 1975 issue can be purchased from the AOA, 212 East Ohio Street, Chicago, Illinois 60611. The cost of the 1975 issue is $30.00.

Other sources of extant data on physicians include the U.S. Physician Reference Listing, the Directory of Medical Specialists, state licensure files, and membership listings maintained by state and local medical societies. For a more detailed discussion of these sources as well as the publications of the AMA and AOA, see A Guide to the Development of Health Resource Inventories.[2]

C. DATA ANALYSIS AND USE

In discussing the uses of physician data in this chapter, it will be assumed that the HSA has developed an inventory of the physicians in its service area and has obtained for each physician the data items suggested above. A methodology for developing such an inventory is contained in A Guide to the Development of Health Resource Inventories.[3]

1. Assessing the Availability of Physicians and Identifying Shortage Areas

The most basic statistical tool for assessing the availability (i.e., supply) of physicians in a given geographic area is the physician to population ratio. Although it is not without its limitations, the physician to population ratio can also be a useful, albeit crude, indicator of an area's need for additional physicians.* This ratio, in fact, is presently being used in the identification of MUAs and CMSAs. The calculation of this ratio and its use in MUA and CMSA identification is described below.

a. Identification of Medically Underserved Areas (MUAs)

Under the regulations pertaining to the identification of MUAs, the area under consideration can be either 1) a county, minor civil division, or census county division in non-metropolitan areas (i.e., areas that are outside of SMSAs); 2) a census tract in metropolitan areas (i.e., SMSAs); or 3) a group of census tracts, minor civil divisions, or census county divisions which constitute a "natural neighborhood."[4]

Data on four indicators of underservice are to be used by HSAs in the calculation of an Index of Medical Underservice (IMU).** These indicators include 1) the ratio of primary care physicians to population in the county which contains the area under consideration; 2) the infant mortality rate in the county containing the area; 3)

*For a discussion of some of the limitations in using physician to population ratios, see William L. Kissick, "Forecasting Health Manpower Needs," Hospitals, Journal of the American Hospital Association, Vol. 41, September 16, 1967, p. 49.

**For a discussion of some of the problems associated with the calculation and use of the Index of Medical Underservice, see the articles related to this Index in Health Services Research, Volumn 10, Number 2, Summer 1975.

the percentage of population aged 65 or over in the area; and 4) the percentage of population with family incomes below the poverty level in the area. Regardless of the particular area under consideration (i.e., a county, minor civil division, census tract, etc.), the ratio or primary care physicians to population is to be calculated for the county which contains the area under consideration. Under the current regulations pertaining to MUA designation, therefore, it is not necessary to calculate sub-county physician to population ratios.*

For purposes of computing the IMU, primary care physicians are defined as follows:

> active doctors of medicine (M.D.) and doctors of osteopathy (D.O.) who spend at least 50 percent of their time engaged in direct patient care in the fields of general or family practice, internal medicine, pediatrics, or obstetrics and gynecology. In metropolitan areas the computations should include all non-Federal physicians meeting the above definition. In nonmetropolitan areas the computation should include Public Health Service physicians in addition to non-Federal physicians.[5]

Interns and residents are not to be included in the above definition unless it can be determined that they are spending 50 percent or more of their professional time in the delivery of direct patient care in an organized outpatient department of a hospital.[6]

The above definition of primary care physicians differs in an important manner from the way in which physician data are collected by the previously discussed extant physician data sources (i.e., the AMA's Directory and the AOA's Yearbook). This important difference is the specific requirement that physicians must spend "at least 50 percent of their time engaged in direct patient care." The AMA and

*"The use of sub-county physician ratios is being studied by the Health Resources Administration and the Health Services Administration. A uniform definition of medical service areas (primary-care service areas) is being developed for programs delivering primary medical care. County physician-to-population ratios will be used in the IMU until a medical service area definition is established." Federal Register, September 2, 1975.

AOA, however, attribute to physicians the professional activity of "direct patient care" whenever that activity is the one in which the physician spends the largest percentage (which may or may not be 50 percent or more) of his/her time. Unless data to the contrary are available, it is not unreasonable to assume that the physicians listed in the extant data sources with the professional activity of "direct patient care" do, in fact, spend at least 50 percent of their time engaged in that activity.*

For use in the calculation of IMU, the physician to population ratio is to be computed as follows:

$$\frac{p}{r} \times 1{,}000$$

where,

p = the number of primary care physicians, as defined above, in the county containing the area being proposed for MUA designation

r = the resident population of the county minus the resident members of the Armed Forces and inmates of institutions

Presented in Table 1 is a listing (column 1) of the 30 counties which make up Central HSA's health service area. Metropolitan counties are designated with an asterisk (*). Counties without an asterisk are nonmetropolitan. Column 2 of Table 1 lists the most current population estimate for each county.** Column 3 contains the most current

*An analogous situation exists in reference to the physician's primary specialty designation. That is, the particular field attributed to physicians in the extant data sources as their primary specialty is the field in which the physician spends the largest percentage of time, which may or may not be 50 percent or more.

**These estimates can be obtained from the Census Bureau's Current Population Reports, Series P-25 and P-26, U.S. Government Printing Office, Washington, D.C. 20402. The county population estimates presented in these reports do not, however, exclude resident members of the Armed Forces or inmates of institutions. For estimates of these population subgroups, which can then be subtracted from the total county estimates presented in the Current Population Reports, 1970 census publications will need to be used.

Table 1

Population and Physician Data
by County, Central HSA

County * metro-politan area	Population Estimates	Total Primary Care Physicians: GPs, FPs, IMs, Peds, OB-GYNs		General Surgeons
		Non-Federal	Public Health Service	Non-Federal
(1)	(2)	(3)	(4)	(5)
1	21,900	7	0	0
2	88,300	34	1	8
3	84,400	34	2	7
4	11,200	2	0	0
*5	112,200	47	N/A	N/A
6	18,200	5	0	0
7	30,800	5	0	1
8	16,000	5	0	1
9	8,500	3	0	0
10	34,400	7	1	1
*11	139,400	50	N/A	N/A
12	24,100	5	0	1
13	18,600	4	1	0
14	16,000	5	0	0
*15	130,600	65	N/A	N/A
*16	32,000	10	N/A	N/A
*17	65,100	26	N/A	N/A
*18	40,300	11	N/A	N/A
19	53,200	19	0	2
20	29,500	8	0	2
*21	792,500	510	N/A	N/A
*22	59,900	15	N/A	N/A
*23	67,800	25	N/A	N/A
*24	39,100	8	N/A	N/A
25	79,200	33	1	6
*26	47,300	12	N/A	N/A
27	20,900	5	0	1
28	27,400	6	0	1
29	17,600	3	0	0
30	6,700	0	0	0
Total	2,133,100	969	6	31

N/A - not applicable

number of non-federal primary care physicians, as defined above, in each county. Column 4 contains the most current number of Public Health Service physicians in each nonmetropolitan county. Column 5 contains the most current number of general surgeons in each nonmetropolitan county. The figures in column 5 will be needed in identifying CMSAs to be discussed later in this chapter. The figures in columns 3, 4 and 5 can be obtained from the extant data sources discussed earlier.

The data in Table 1 can be used in calculating the physician to population ratio as follows:

- for nonmetropolitan areas:

$$\frac{\text{columns } 3 + 4}{\text{column } 2} \times 1{,}000$$

e.g., County 2: $\frac{34 + 1}{88{,}300} \times 1{,}000 = .396$

- for metropolitan areas (*):

$$\frac{\text{column } 3}{\text{column } 2} \times 1{,}000$$

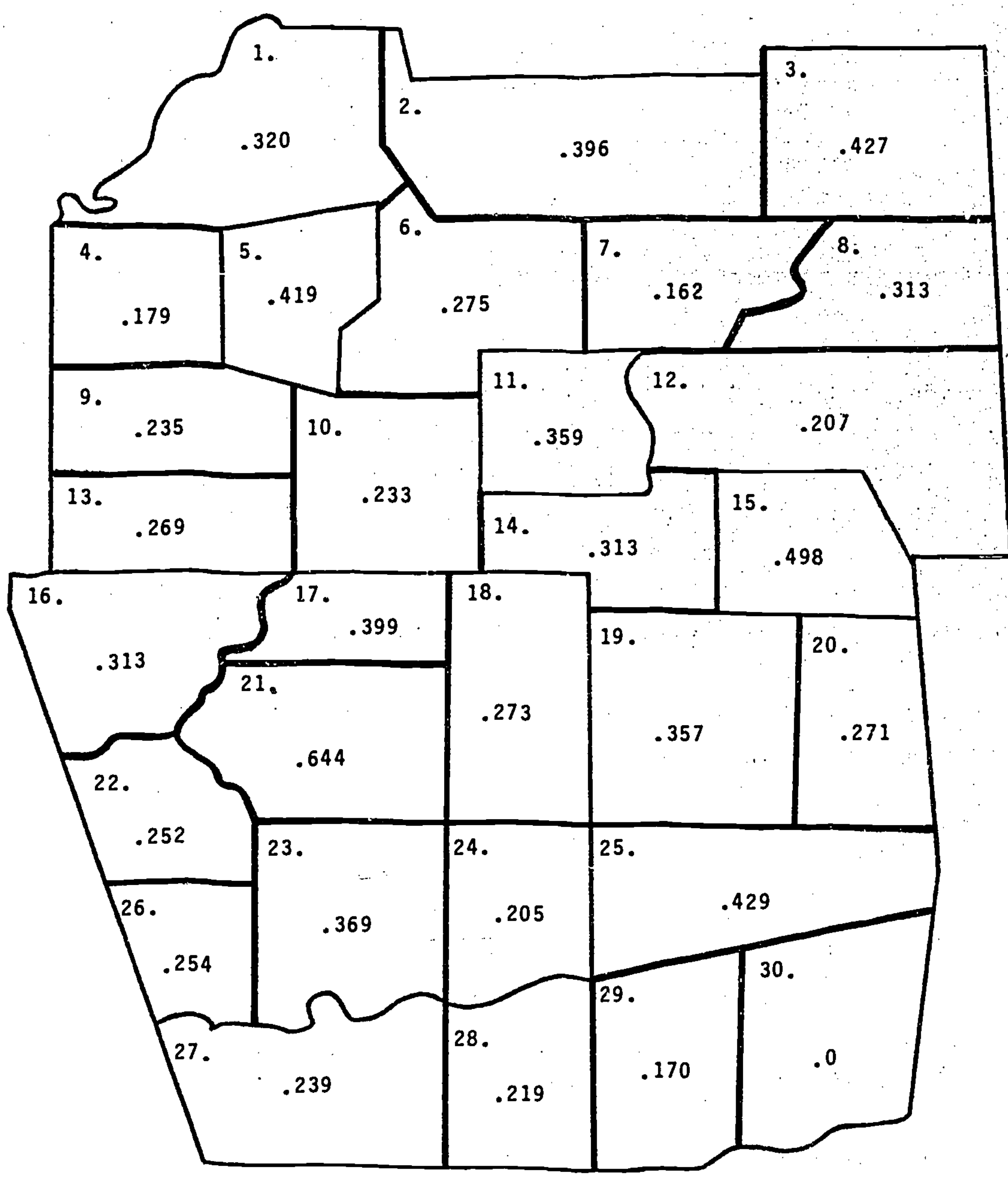
e.g., County 5: $\frac{47}{112{,}200} \times 1{,}000 = .419$

Presented in Figure 1 are the physician to population ratios calculated according to the above formula for each of the 30 counties in Central HSA's health service area.

The ratios presented in Figure 1 can be used to compare the relative availability of primary care physicians in the 30 counties. The reader should refer to the regulations governing MUA designation concerning the incorporation of these ratios into the IMU, i.e., the particular weights that are to be applied to specific ratios and their inclusion along with the weighted values of the other three indicators in the calculation of this composite index. The most current (1975) list of MUAs also appears in the publication containing these regulations.[7] Continuous modifications to this list will be

Figure 1

Number of Primary Care Physicians Per 1,000
Residents by County, Central HSA



made. HSAs should send their recommendations for changes to these lists to:

Director, Bureau of Community Health Services
Department of Health, Education, and Welfare
Room 6A-14
5600 Fishers Lane
Rockville, Maryland 20852

b. Identification of Critical Medical Shortage Areas (CMSAs)

Under the regulations pertaining to the identification of CMSAs, the area under consideration can be either 1) a county or several contiguous counties; 2) a portion of a county or an area made up of portions of more than one county in nonmetropolitan areas (where possible, such portions are to be defined in terms of one or more minor civil divisions or census county divisions); 3) a group of census tracts in metropolitan areas. For purposes of illustration, it will be assumed that the areas under consideration are individual counties.

According to CMSA designation criteria, primary care physicians are defined as follows:

> non-Federal...physicians /M.D.s and D.O.s/...practicing general or family medicine, internal medicine, pediatrics, and obstetrics and gynecology, and those general surgeons who spend 50 percent or more of their patient care time in primary care practice. (Unless data to the contrary is supplied, general surgeons in non-metropolitan areas will be assumed to fall in this category while general surgeons in metropolitan areas will be assumed not to fall in this category.)[8]

Interns and residents are not to be included in the above definition unless it can be determined that they are spending 50 percent or more of their professional time in the delivery of direct patient care in an organized outpatient department of a hospital.[9]

Unlike the ratio that is to be used in the identification of MUAs (which is the ratio of physicians to population), the ratio that is to be used in the identification of CMSAs is, instead, the ratio of population to physicians. This ratio is to be computed as follows:

$$\frac{r}{p}$$

where,

r = the resident civilian population of the particular area under consideration

p = the number of full-time equivalent primary care physicians, as defined above, in the area under consideration*

The formula for calculating the population to physician ratio using the data in Table 1 is as follows:

- for nonmetropolitan areas:

$$\frac{\text{column 2}}{\text{columns 3 + 5}}$$

e.g., County 2: $\frac{88,300}{34 + 8} = 2,102$

- for metropolitan areas (*):

$$\frac{\text{column 2}}{\text{column 3}}$$

*As specified in the regulations, "the use of 'full-time equivalent' refers to percentage of a full work week, and is intended to make allowances for physicians who are semi-retired, who operate a reduced practice due to infirmity or other limiting conditions (which should be specified), or who are available to the population of the shortage area on a part-time basis only. Thus advanced age alone shall not constitute justification for counting an individual physician as less than full time." (Bureau of Health Manpower) The specific criterion to be used in the determination of full-time equivalency, that is, the number of hours that constitute a full work week, is not included in the regulations. However, even if it was, data concerning the number of hours each physician works per week are not available from the extant data sources.

e.g., County 5: $\frac{112,200}{47} = 2,387$

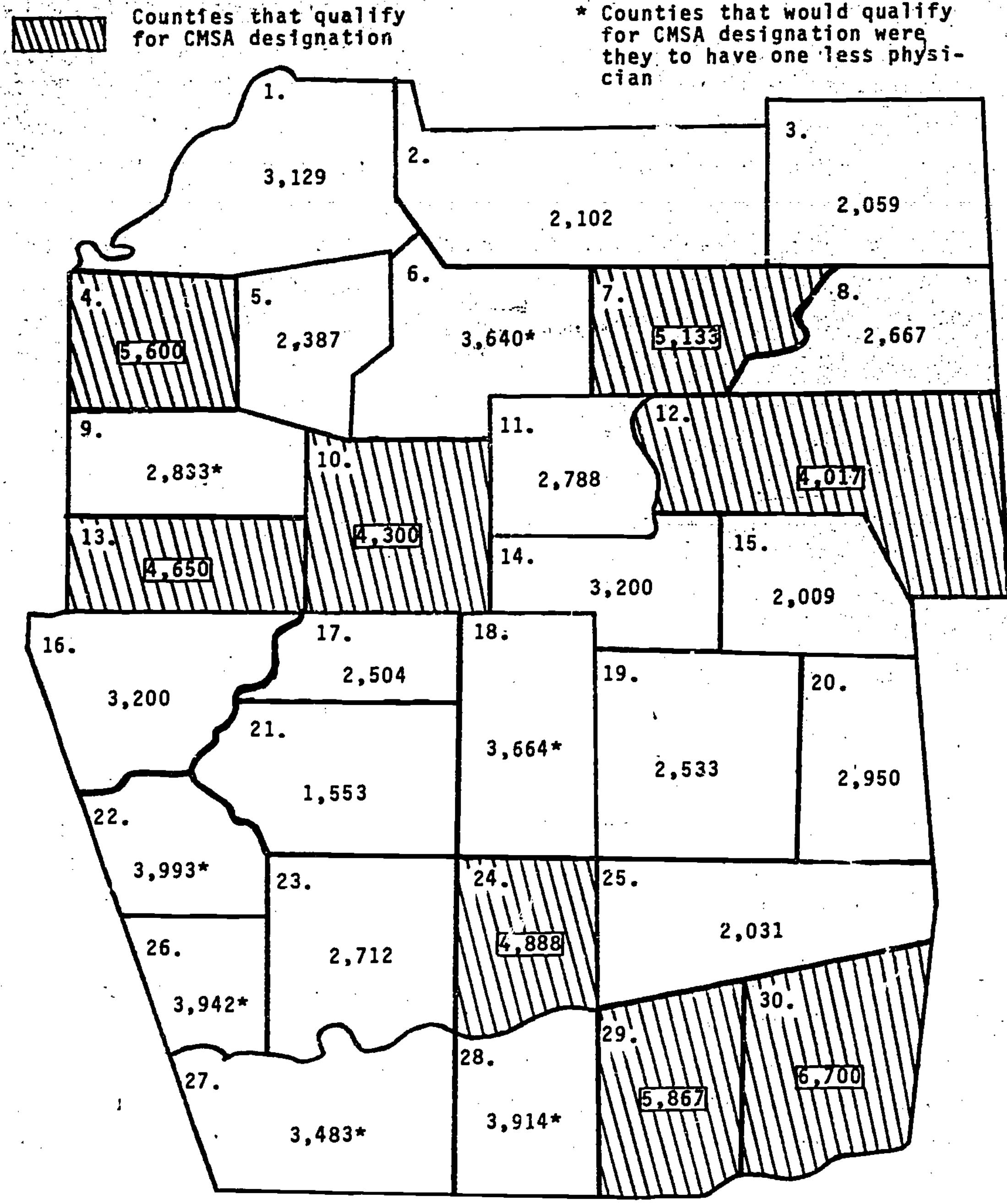
Presented in Figure 2 are the population to physician ratios calculated according to the above formula for each of the 30 counties in Central HSA's health service area. For an area to be designated a CMSA, the population to physician ratio must be greater than 4,000 to 1. The counties within Central HSA's health service area which have ratios greater than 4,000 to 1 are easily identifiable in Figure 2. Thus, it would appear, that eight counties qualify for CMSA designation: counties 4, 7, 10, 12, 13, 24, 29 and 30. Before final designation can be made, however, the regulations governing CMSA designation require that certain contiguous area considerations be made. Where the proposed CMSA is a county, the contiguous area considerations are as follows:

> the population-to-primary care physician ratios in contiguous counties (or independent cities) will be examined. A ratio of more than 2000:1 in a contiguous county will be taken to indicate that the county has no excess capacity which might alleviate the shortage situation in the county for which designation is requested. In the event that a contiguous county or independent city has a ratio less than 2000:1, the distance by road between population centers of the two counties and the topography of the region will be considered in order to determine whether the resources of the contiguous county should disqualify the proposed county from designation.[10]

As shown in Figure 2, county 24 is contiguous to a county which has a ratio of less than 2,000 to 1, namely, county 21 (1,553 to 1). As required above, the distances between the population centers of counties 21 and 24 and the topography of the region will need to be considered before it can be determined whether or not county 24 can be disqualified from CMSA designation. Other potential CMSAs which may be contiguous to counties with ratios less than 2,000 to 1 are counties 4, 12, 13, 29 and 30. Before final determinations can be made for these counties, data for contiguous counties which are outside Central HSA's health service area need to be obtained and analysed.

Figure 2

Number of Residents Per Primary Care Physician by County, Central HSA



The regulations governing CMSA designation also allow considerations which may nelp an area to qualify for CMSA designation where the criteria outlined above are nearly but not fully met in that area. Among these considerations is "advanced age of a significant number of the area's physicians."[11] The regulations do not, however, specify how many aged physicians constitute a "significant number."

As noted in an earlier section, the year of birth of each physician is provided in the extant data sources. From each physician's year of birth, the age distribution of an area's physicians can be estimated. It might be found, for example, that the three primary care physicians who practice in county 9 are aged 45, 65 and 67. It would be reasonable to suspect that the physicians aged 65 and 67 are not in full-time practice. A telephone call could be made to these two physicians for the purpose of determining the extent of their practice. Should it be determined that both have a halftime practice (i.e., have 20 or less office hours per week), the population to physician ratio for county 9 could be recalculated on the basis of two rather than three full-time equivalent primary care physicians. The resulting ratio, 4,250 to 1, would qualify county 9 for CMSA designation.

The data in Table 1 can be updated whenever new editions of the previously described extant data sources become available. These data should, of course, be updated whenever new information is available regarding changes in an area's supply of physicians. For example, changes in the supply of an area's physicians can occur from a death or retirement, which are often reported in local newspapers. Obviously, HSAs will want to closely monitor the supply of physicians in areas that would qualify for CMSA designation were they to have one less physician (e.g., those counties in Figure 2 with ratios followed by an asterisk).

Lists of the most current CMSAs are periodically published in the Federal Register. To date, the most current list is for 1975.[12]

HSAs wishing to request additions to or deletions from these lists should contact:

Chief, Manpower Analysis Branch
Bureau of Health Manpower
Attention: Shortage Area Designation Staff
NIH, Building 31, Room 3B06
9000 Rockville Pike
Bethesda, Maryland 20014

Any area which has been designated as a CMSA is eligible to apply for placement of National Health Service Corps personnel to alleviate its shortage situation. By assisting such areas in the CMSA designation and NHSC application process, HSAs will have significantly contributed toward one of the national health priorities recognized by Congress, namely, the provision of primary medical care to medically underserved populations.

2. Estimating Future Requirements for Primary Care Physicians

The problem of accurately projecting the need for physicians and other health manpower has long been a matter of concern for health planning agencies. The actual task of deriving such projections, however, has remained what is aptly described by Reinhardt as "Mission Impossible."[13] Over the years, the methodologies used in projecting physician manpower needs have evolved from simple physician to population ratios[14] to complex multivariate, multi-equation econometric models that employ a computerized regression-simulation approach.[15] Although the methods are varied, one fact remains concerning physician manpower projections—they are all largely "guesstimates."

The most convincing evidence to support this fact is contained in Hansen's critique of several physician manpower projections for 1975. Hansen's examination of these projections showed that depending upon which projection was used, it was possible to conclude that either a surplus as high as 21,700 or a deficit as high as 65,000 would occur in the number of physicians needed in the United States in 1975.[16]

The fact that physician manpower projections are hazardous, however, does not preclude their necessity in the execution of an HSA's planning and resources development functions. For example, an HSA will need physician manpower projections in deciding how many physicians and/or physician extenders need to be recruited to prevent the number of CMSAs in its service area from becoming more numerous and the shortage of physicians in existing CMSAs from becoming more acute. In the remainder of this chapter, a methodology for projecting the number of physicians required for primary medical care is described. This methodology was designed specifically for use with existing sources of physician resource and utilization data. Other methods for projecting physician manpower needs are described in the literature.[17]

The purpose of this methodology is to estimate the number of primary care physicians (as defined above according to the MUA and CMSA criteria) that will be required at some future date to provide adequate medical care to the residents of a particular geographic area. For purposes of illustration, counties 9, 10 and 13 will constitute the geographic area (see Figure 2).

Various assumptions concerning the boundaries of medical service areas, the morbidity levels of an area's population (need), their patterns of utilizing the area's physicians (demand), the levels of productivity of the area's physician's, etc., are inherent in nearly all physician manpower projection methodologies. For example, it will be assumed here that these three contiguous, rural counties constitute a single primary medical care service area. While the residents of these three counties may go outside the area for certain types of specialty care, it is assumed that they do not go outside the area to utilize primary care physicians. Other assumptions that are used in the application of this methodology are indicated at the appropriate points in the discussion of the methodology. Depending upon its unique health service area characteristics, the availability of other data, etc., each HSA will, of course, have to judge the appropriateness of each of these assumptions in its own application of this methodology.

The specific parts and steps in estimating the number of primary care physicians that will be required in this three county area in 1980 are as follows:

Part I. Estimate Future Demand

The demand for primary care physicians is measured in terms of the annual number of primary care physician visits made by the residents of the three county area. The definition of a physician visit used here is similar to the definition used by the National Center for Health Statistics (NCHS), which is as follows:

> A physician visit is defined as consultation with a physician...for examination, diagnosis, treatment, or advice. The visit is considered to be a physician visit if the service is provided directly by the physician or by a nurse or other person acting under a physician's supervision....[18]

Only those physician visits which occur in the physician's office are included in the above definition. Excluded, therefore, are physician visits which occur in hospitals, nursing homes, the patient's home, company or industry health units, or any other place outside the physician's office where a physician consultation might take place. Also excluded are physician consultations which occur via telephone. The reason for restricting the place of physician contact to the physician's office is to make the definition of "physician visit" as consistent as possible with the definition of "office visit" to be discussed later under physician productivity.

Refer to Worksheet A in the discussion of the specific steps under Part I.

Step 1: estimate the number of total physician visits per person per year

In column (1) of Worksheet A, list the most recent annual number of total physician visits per person for the various age and sex categories indicated. It is assumed that a health interview survey has not been conducted in the three county area under consideration.* Thus, the required age/sex-specific physician vi :s will need to be synthetically estimated from national rates which are available from extant national data sources. Even in areas where local health

*A methodology for conducting a health interview survey is described in Chapter 15 of Section IV.

Worksheet A

Estimating Future Demand

Sex/age	Number of total physician visits per person per year, U.S., 1974		Office to total physician visit ratio, U.S. farm population, 1971		Farm to total population physician visit ratios, U.S., 1971		Primary care to total physician visit ratio, U.S., 1971		Population Projections 1980
	(1)	X	(2)	X	(3)	X	(4)	X	(5)
Male									
under 17	4.3		.808		.767		.83		11,625
17-24	3.2		.808		.767		.83		5,175
25-44	3.5		.808		.767		.83		8,925
45-64	4.9		.808		.767		.83		7,350
65-74	6.8		.808		.767		.83		1,350
75+	6.3		.808		.767		.83		1,725
Female									
under 17	4.0		.808		.782		.83		11,175
17-24	5.8		.808		.782		.83		5,550
25-44	6.3		.808		.782		.83		9,600
45-64	6.1		.808		.782		.83		8,100
65-74	6.9		.808		.782		.83		1,050
75+	6.6		.808		.782		.83		3,375

Total physician v

Worksheet A

Estimating Future Demand

er of to- physician its per son per ir, U.S., 1974	Office to total physician visit ratio, U.S. farm population, 1971	Farm to total population physician visit ratios, U.S., 1971	Primary care to total physician visit ratio, U.S., 1971	Population Projections, 1980	Estimated annual number of primary care physician visits, 1980
(1) X	(2) X	(3) X	(4) X	(5) =	(6)
4.3	.808	.767	.83	11,625	25,713
3.2	.808	.767	.83	5,175	8,518
3.5	.808	.767	.83	8,925	16,068
4.9	.808	.767	.83	7,350	18,525
6.8	.808	.767	.83	1,350	4,722
6.3	.808	.767	.83	1,725	5,590
4.0	.808	.782	.83	11,175	23,442
5.8	.808	.782	.83	5,550	16,882
6.3	.808	.782	.83	9,600	31,718
6.1	.808	.782	.83	8,100	25,913
6.9	.808	.782	.83	1,050	3,800
6.6	.808	.782	.83	3,375	11,682
				Total physician visits:	192,573

interview surveys have been conducted, the national rates can be used as a planning standard. These rates may be obtained for the United States population from the NCHS Vital and Health Statistics publication entitled, "Current Estimates from the Health Interview Survey: United States." To date, the most current estimates are for 1974.[19]

It is assumed that the residents of the three county area utilize physicians at the same rate as the U.S. population as a whole. Furthermore, it is assumed that these age/sex-specific rates will remain constant through 1980.

Certain adjustments will need to be made to the rates listed in column (1). For one thing, the definition of physician visits used by NCHS is not consistent with the above definition. For example, telephone contacts are included in the rate of total physician visits listed in column (1). An adjustment to include only those visits which occur in the physician's office is made in Step 2. Other adjustments are made in Steps 3 and 4 to reflect the type of population and type of physician visit under consideration in this example application of the methodology.

Step 2: estimate the ratio of physician visits which occur in the physician's office to total physician visits

In column (2) of Worksheet A, list the ratio of physician visits which occur in the physician's office to total physician visits. According to the most recent data (1971) from NCHS, this ratio is .808 for the U.S. farm population, i.e., 80.8 percent of the total number of outpatient physician visits reported to have occurred among the U.S. farm population in 1971 occurred in the physician's office.[20]

Step 3: estimate the ratio of physician visits among the farm population to physician visits among the total population

In column (3) of Worksheet A, list the sex-specific ratios of physician visits among the farm population to physician visits among the total population. This adjustment is made to the rate of physician visits in column (1) because of the fact that the residents of the three counties under consideration are predominantly a farm population, and, as shown by national data, the rate of physician utilization among the farm population is less than among the population as a whole.* Similar adjustments can be made for non-farm populations residing outside SMSAs and for populations residing within SMSAs.

According to the most recent data (1971) from NCHS, the ratios of physician visits among the U.S. farm population to physician visits among the total U.S. population is .767 for males and .782 for females.[21] Not presented in column (3), but calculable from the NCHS national data, are certain age/sex-specific ratios.

Step 4: estimate the ratio of primary care physician visits to total physician visits

In column (4) of Worksheet A, list the ratio of primary care physician visits to total physician visits. According to the most recent data (1971) from NCHS, this ratio is .83, i.e., 83 percent of all physician visits in 1971 were visits made to general practitioners, pediatricians, internists, obstetrician/gynecologists, and surgeons.[22] Age/sex-specific ratios are not available. Ratios by type of residence (i.e., farm, non-farm, etc.) are also not available.

*The farm population includes persons living on places of ten acres or more from which sales of farm products amounted to $50 or more during the previous 12 months or on places of less than ten acres from which sales of farm products amounted to $250 or more the preceding 12 months.

Step 5: project the population to 1980

In column (5) of Worksheet A, list the age/sex-specific populations projected to 1980. It is assumed here that Central HSA has available such projections for the three county area.

Step 6: estimate the annual number of primary care physician visits in each age/sex-specific category

In column (6) of Worksheet A, list the annual number of age/sex-specific primary care physician visits. These figures are the product of columns (1) x (2) x (3) x (4) x (5).

Step 7: estimate the total number of primary care physician visits made by the residents of the three county area in 1980

The total number of primary care physician visits that is estimated to be made by the residents of the three county area in 1980 is obtained by summing the age/sex-specific numbers of visits in column (6). This figure, 192,573 primary care physician visits, represents the area's estimated future demand for primary care physicians.

Part II. Estimate Future Supply

The supply of primary care physicians is measured in terms of the number of general and family practitioners, internists, pediatricians, obstetrician/gynecologists, and general surgeons practicing in the three county area. Refer to Worksheet B in the discussion of the specific steps under Part II.

Worksheet B

Estimating Future Supply

Age of physicians, 1973	Specialty of physicians, 1973/1980	Age of physicians, 1980
(1)	(2)	(3)
64	GP	71 retired
62	GP	69 retired
59	GP	66 retired
57	GP	64
55	GP	62
51	IM	58
50	GP	57
48	OB-GYN	55
46	GS	53
45	GP	52
43	GP	50
42	OB-GYN	49
40	IM	47
39	Ped	46
38	GP	45
35	Ped	42
32	IM	39

Step 1: determine the number of primary care physicians by age and specialty currently practicing in the three county area

In column (1) of Worksheet B, list the ages of the primary care physicians currently practicing in the three county area. This information can be obtained from the extant data sources described earlier in this section. In column (2) list the specialty of each of the physicians listed in column (1). These data can also be obtained from the previously described extant sources. It is assumed here that the physicians listed under column (1) are all full-time physicians who spend at least 50 percent of their time engaged in direct patient care.

Step 2: estimate the number of primary care physicians by age and specialty who will be practicing in the three county area in 1980

In column (3) of Worksheet B, list the ages of the primary care physicians in 1980. These figures are obtained by simply adding seven years to each of the ages under column (1). It is assumed that physicians in this three county area retire at age 65, thus those physicians aged 66, 69 and 71 in 1980 are assumed to be retired. It is also assumed that there are no deaths among these physicians between 1973 and 1980, that none of them move their practices outside the area, and that none of them change their specialties. Furthermore, it is assumed that no new physicians move into the area between 1973 and 1980. To the extent that an HSA has available local data pertaining to the retirement ages of its physicians, their death rates, and their in and out-migration rates, adjustments to the expected distribution of physicians in 1980 can be made.*

*1969-1973 age/sex and specialty-specific death rates of physicians are available for the U.S. and geographic divisions from Louis J. Goodman, "The Longevity and Mortality of American Physicians, 1969-1973," Health and Society. Summer 1975, pp. 353-375.

Part III. Estimate Future Productivity

The producti ity of the primary care physicians practicing in the three county area is measured in terms of their total number of annual office visits. An office visit is defined as a patient visit which occurs in the physician's office.[23] This definition is consistent with the earlier definition of "physician visits". Refer to Worksheet C in the discussion of the specific steps under Part III. Note that the distribution of primary care physicians in Worksheet C is derived from columns (2) and (3) of Worksheet B.

Step 1: estimate the average number of office visits per week by physician specialty

In column (1) of Worksheet C, list the average number of office visits per week by each of the primary care physician specialties. It is assumed that Central HSA has not yet conducted an ambulatory care survey among its area's physicians.* Thus, the required number of office visits will need to be synthetically estimated from national rates which are available from extant national data sources. These rates may be obtained for all U.S. physicians from the AMA publication entitled Reference Data on Profile of Medical Practice. To date, the most current estimates are for 1973.[24] These rates are available for nonmetropolitan and metropolitan areas. The figures under column (1) are for nonmetropolitan areas.

It is assumed that the rates of productivity among the primary care physicians in the three county area are the same as those for the total population of primary care physicians in the U.S. Furthermore, it is assumed that these specialty-specific rates will remain constant through 1980.

*A methodology for conducting an ambulatory care survey is described in Chapter 11 of Section IV.

Worksheet C

Estimating Future Productivity

Estimated physician specialty/ age, 1980	Average number of office visits per week, U.S., non-metropolitan areas, 1973		Age-specific to total office visit ratios, U.S., 1969		Average number of weeks worked per year, U.S., non-metropolitan areas, 1972		Age-specific to total weeks worked ratios U.S., 1968
	(1)	X	(2)	X	(3)	X	(4)
General and Family Practice							
64	162.8		.939		47.8		.998
62	162.8		.939		47.8		.998
57	162.8		1.064		47.8		.996
52	162.8		1.133		47.8		1.000
50	162.8		1.133		47.8		1.000
45	162.8		1.055		47.8		1.013
Internal Medicine							
58	90.0		1.064		47.0		.996
47	90.0		1.055		47.0		1.013
39	90.0		1.005		47.0		1.017
Pediatrics							
46	151.3		1.055		47.5		1.013
42	151.3		1.041		47.5		1.013
Obstetrics - Gynecology							
49	109.5		1.055		45.8		1.013
45	109.5		1.064		45.8		.996
General Surgery							
53	91.5		1.133		45.9		1.000

Total office visit

Worksheet C

Estimating Future Productivity

Average number f office visits er week, U.S., on-metropolitan areas, 1973	Age-specific to total office visit ratios, U.S., 1969	Average number of weeks worked per year, U.S., non-metropolitan areas, 1972	Age-specific to total weeks worked ratios U.S., 1968	Estimated number of primary care office visits, 1980
(1) X	(2) X	(3) X	(4) =	(5)
162.8	.939	47.8	.998	7,293
162.8	.939	47.8	.998	7,293
162.8	1.064	47.8	.996	8,247
162.8	1.133	47.8	1.000	8,817
162.8	1.133	47.8	1.000	8,817
162.8	1.055	47.8	1.013	8,317
90.0	1.064	47.0	.996	4,483
90.0	1.055	47.0	1.013	4,521
90.0	1.005	47.0	1.017	4,323
151.3	1.055	47.5	1.013	7,681
151.3	1.041	47.5	1.013	7,579
109.5	1.055	45.8	1.013	5,360
109.5	1.064	45.8	.996	5,315
91.5	1.133	45.9	1.000	4,758
			Total office visits:	92,804

Step 2: estimate the age-specific to total office visit ratios

In column (2) of Worksheet C, list the ratios of the average number of office visits per week by age of physician to the average number of office visits per week among all physicians. This adjustment is made because, as shown by national data, the age of a physician is related to the average number of weekly office visits. The ratios listed under column (2) are derived from the most current national data (1969).[25]

Step 3: estimate the average number of weeks worked per year by physician specialty

In column (3) of Worksheet C, list the average number of weeks worked per year by each of the primary care physician specialties. These rates may also be obtained from the AMA publication cited above in Step 1. The most current data are for 1972.[26] The figures under column (3) are for nonmetropolitan areas.

Step 4: estimate the age-specific to total weeks worked ratio

In column (4) of Worksheet C, list the ratios of the average number of weeks worked per year by age of physician to the average number of weeks worked per year among all physicians.

This adjustment is made because, as shown by national data, the age of a physician is related to the average number of weeks worked per year. The ratios listed under column (4) are derived from the most current national data (1968).[27]

Step 5: estimate the annual number of office visits by age and specialty of physician

In column (5) of Worksheet C, list the estimated annual number of office visits provided by each physician according to age and specialty. These figures are the product of columns (1) x (2) x (3) x (4).

Step 6: estimate the total number of office visits provided by the physicians in the three county area in 1980

The total number of office visits that is estimated to be provided by the primary care physicians in the three county area in 1980 is obtained by summing the age/specialty-specific number of visits in column (5). This figure, 92,804 primary care office visits, represents the area's estimated future productivity of its primary care physicians.

Part IV. Estimate Future Requirements

Step 1: estimate surplus/deficit of physician visits in 1980

To determine whether or not the projected number of office visits provided by the area's primary care physicians is adequate to meet the area's projected demand for primary care, the following formula can be used:

PV - OV = surplus/deficit visit demand

where,

PV = projected demand for physician visits in 1980

OV = projected provision of office visits in 1980

For the three county area under consideration, PV is derived from the total of column (6) in Worksheet A, and OV is derived from the total of column (5) in Worksheet C. Applying these figures to the above formula, we have:

192,573 - 92,804 = 99,769 visits

It is estimated that in 1980 there will be a surplus visit demand (SVD) from the area's residents of 99,759 annual physician visits which cannot be provided by the area's primary care physicians. If a negative number of visits were derived instead, it would suggest that the projected supply and productivity of the area's primary care physicians would be adequate to meet the expected demand.

Step 2: estimate the average number of annual office visits provided by primary care physicians

The weighted mean derived from column (5) of Worksheet C can be used as the estimated average number of annual office visits provided by a primary care physician. The formula for calculating the weighted average of annual office visits is as follows:

$$\sum_{i=1}^{n}(p_i \times a_i) = \text{average number of annual office visits per primary care physician (AOV)}$$

where,

p_i = the proportion of primary care physicians in specialty i

a_i = the estimated average number of annual office visits provided by primary care physicians in specialty i

n = the number of primary care physician specialties

Using the data from Worksheet C in the above formula, we have:

n	p		a		
GPs	.4286	x	8,131	=	3,485
IMs	.2143	x	4,442	=	952
Peds	.1429	x	7,630	=	1,090
OB-GYNs	.1429	x	5,338	=	763
GSs	.0714	x	4,758	=	340
				Σ =	6,630

It is estimated that, on the average, a primary care physician in the three county area will provide 6,630 office visits in 1980.

Step 3: estimate the number of additional primary care physicians required in 1980 to meet the expected surplus visit demand

It is assumed that new primary care physicians who locate in the three county area will provide, on the average, the same number of annual office visits as was estimated in 1980 for the current supply of primary care physicians in this area, i.e., 6,630 visits. Based upon this assumption, the formula for estimating the number of additional primary care physicians required is as follows:

$$\frac{SVD}{AOV} = \text{number of additional primary care physicians required}$$

where,

SVD = surplus visit demand

AOV = annual office visits

Using the figures derived from Steps 1 and 2 in the above formula, we have:

$$\frac{99,769}{6,630} = 15 \text{ additional primary care physicians}$$

Thus, in this three county area of Central HSA, it is estimated that 15 additional primary care physicians will be required in 1980 to meet the demand for primary medical care.

Unless past in-migration rates of physicians have been favorable and/or projected sizes of the graduating classes from nearby medical schools are favorable, Central HSA staff may feel that the recruitment of 15 new primary care physicians for this three county area is an unrealistic expectation. Or it may be the case that the recruitment of this many physicians would not be unrealistic, but less costly and/or more efficient procedures for meeting the anticipated surplus visit demand are desired by Central HSA staff. Increasing the level of productivity of the current supply of physicians could be one such procedure. For example, Central HSA may choose to assist these counties and their present primary care physicians in the recruitment of physician extenders and/or additional auxiliary personnel. Although only limited data are available, they have shown that a physician's productivity can increase as much as 50 percent with the addition of a physician extender and about 20 percent per additional aide.[28]

Whatever alternative programs are promoted by an HSA in response to an anticipated surplus demand for primary care in its health service area, the above methodology can provide a useful, albeit rough, estimate of the number of primary care physicians or their equivalents that are required to meet this demand. And, as has been demonstrated, the advantage of this methodology is that it can be employed solely with the use of extant physician data.

Appendix A: Specialty Categories of Physicians (M.D.s and D.O.s) as of 1974

Physician (M.D.) Specialty Groups
(Source: American Medical Association

- Aerospace medicine
- Allergy
- Anesthesiology
- Broncho-Esophagology
- Cardiovascular diseases
- Dermatology
- Diabetes
- Emergency medicine
- Endocrinology
- Family practice
- Gastroenterology
- General practice
- General preventive med.
- Geriatrics
- Gynecology
- Hematology
- Hypnosis
- Infectious disease
- Internal medicine
- Laryngology
- Legal medicine
- Neoplastic diseases
- Nephrology
- Neurology
- Neurology, Child
- Neuropathology
- Nuclear medicine
- Nutrition
- Obstetrics
- Obstetrics & Gyne.
- Occupational medicine
- Ophthalmology
- Otology
- Otorhinolaryngology
- Pathology
- Pathology, Clinical
- Pathology, Forensic
- Pediatrics
- Pediatrics, Allergy
- Pediatrics, Cardiology
- Pharmacology, Clinical
- Physical med. & rehab.
- Psychiatry
- Psychiatry, Child
- Psychoanalysis
- Psychosomatic medicine
- Public health
- Pulmonary diseases
- Radiology
- Radiology, Diagnostic
- Radiology, Pediatric
- Radiology, Therapeutic
- Rheumatology
- Rhinology
- Roentgenology, Diagnostic
- Surgery, Abdominal
- Surgery, Cardiovascular
- Surgery, Colon and Rectal
- Surgery, General
- Surgery, Hand
- Surgery, Head and Neck
- Surgery, Neurological
- Surgery, Orthopedic
- Surgery, Pediatric
- Surgery, Plastic
- Surgery, Thoracic
- Surgery, Traumatic
- Surgery, Urological
- Other specialty

Physician (D.O.) Specialty Groups
(Source: American Osteopathic Association)

- Allergy
- Anesthesiology
- Cardiology
- Dermatology
- Dermatology & Syphil.
- Diseases, Cardiovascular
- Diseases, Peripheral vasc.
- Diseases, Pulmonary
- Endocrinology
- Gastroenterology
- General Practice
- Geriatrics
- Gynecology
- Hematology
- Medicine, Industrial
- Medicine, Internal
- Medicine, Psychosomatic
- Neurology
- Neurology & Psychiatry (or Neuropsychiatry)
- Nuclear medicine
- Obstetrics
- Obstetrics & Gynecology
- Ophthalmology
- Ophthal. & Otolaryn.
- Ophthal. & Otorhino.
- Otolaryngology
- Otorhinolaryngology
- Pathology
- Pathology, Anatomic
- Pathology, Anat. & Clin.
- Pathology, Anat. & Clin. and Cytopathology
- Pathology, Clinical
- Pathology, Forensic
- Pediatrics
- Phys. Med. & Rehab. (or Physiatry)
- Proctology
- Psychiatry
- Psychiatry, Pediatric
- Public health
- Radiation therapy
- Radioactive isotopes
- Radiology
- Rheumatology
- Roentgenology
- Roentgenology, Diagnostic
- Roentgenology, Radiology
- Sclerotherapy
- Surgery
- Surgery, Cardiovascular
- Surgery, Gynecological
- Surgery, Neurological
- Surgery, Obstet.-Gyne.
- Surgery, Orthopedic
- Surgery, Peripheral vasc.
- Surgery, Plastic
- Surgery, Thoracic
- Surgery, Urological
- Urology

Chapter 9

A Study of the Demand for Health Manpower in Hospitals and Nursing Homes

I

STUDY METHODOLOGY

As was noted in Chapter 6 of this section, the purpose of this study is to provide HSAs with a methodology for obtaining data concerning the present and future supply of specific types of health manpower in their service areas, and the present and future demand in hospitals and nursing homes for each of these types of manpower.

A. Definition of Concepts

In order for this study to be successfully implemented a number of concepts must be clearly defined. The definition of concepts aids in both study design and in the development of operational definitions of study variables. The concepts of prime importance in this study are as follows.

> Health Occupations: includes all persons possessing skills and knowledge unique to the health field.[1] These health occupations include physicians, dentists and registered nurses, as well as all those in the various health occupations known collectively as allied health manpower.[2] This definition excludes persons who perform the "business, clerical and maintenance services essential to the operation of health facilities and agencies," but who are employed in occupations that are not unique to the health field,[3] (e.g., janitors, computer programmers, switchboard operators, etc.).*

---

*While the concept of health occupations can be specified, translation of this concept into generally accepted lists of health occupations is problematic. The way in which each health occupation is defined affects who is counted, particularly when the definitions are not mutually exclusive, or when titles are redefined so that persons are switched from one category to another.[4] Later in this chapter a strategy is presented for developing a list of health occupations suitable for use in this study.

Demand: includes, for every health occupation, the number of budgeted positions which are currently filled and the number of vacant budgeted positions for which active recruiting of employees is currently taking place.*

Hospital: is defined as a type of medical care institution whose function is to provide diagnosis and treatment for patients. It has an organized medical staff, permanent facilities which include inpatient beds, and medical services, including physician services and continuous nursing services.[5]

Nursing Home: is defined as an establishment with at least three beds which meets the definition of a nursing care home, personal care home with nursing, personal care home without nursing or domiciliary care home. Each of these types of home can be defined as follows:[6]

> A nursing care home is one in which at least 50 percent of the residents receive one or more nursing services, and where at least one registered nurse (RN) or licensed practical nurse (LPN) is employed 35 hours or more per week. Nursing services include nasal feedings, catheterization, irrigation, oxygen therapy, full bed bath, enema, hypodermic injection, intravenous injection, temperature-pulse-respiration, blood pressure, application of dressing or bandage, and bowel and bladder retraining.
>
> A personal care home with nursing is one in which (1) some of the residents but less than 50 percent receive nursing care or (2) more than 50 percent of the residents receive nursing care but no RN's or LPN's are employed full time on the staff.
>
> A personal care home without nursing is one in which the facility routinely provides three personal services or more but no nursing service. Personal services include rub or massage service or assistance with bathing, dressing, correspondence or shopping, walking or getting about, and eating.

*This definition excludes positions which are budgeted, but for which no attempt is being made to recruit employees. It also excludes positions which are "desired" or "considered optimal" by administrators of hospitals and nursing homes, but for which no salaries are budgeted. It should be noted that demand, as herein defined, may be quite different from the health manpower needs of a community based on an objective measure of the community's health status. Demand expressed by employers is primarily determined by financial factors which affect the institution's ability to budget new staff positions, and not by objective measures of the health care needs of a community's residents.

A domiciliary care home is one in which the facility routinely provides less than three of the personal services specified in the definition above and no nursing service. This type of facility provides a sheltered environment primarily to persons who are able to care for themselves.

Health Occupation Educational Institutions: includes all institutions which offer accredited or approved programs leading to the granting of a degree, certificate or other award required for entry into employment in a health occupation.

Graduates of Health Occupation Educational Institutions: includes all persons receiving a degree, certificate or other award required for entry into employment in a health occupation, from one of the health occupation educational institutions included in this study.

B. Definition of Variables

The unique characteristics and data needs of individual HSAs will be important determinants in the selection of the variables to be included in this type of study. The variables included in this study, therefore, represent an example of the variables which might be collected by a particular HSA, not a definitive list of study variables.

Careful selection and definition of the health occupations to be included in this type of study is of primary importance. These occupations may be selected from nationally available dictionaries or glossaries of occupational titles, such as the glossary of occupational titles prepared by the American Society of Allied Health Professions,[7] or from lists and definitions of health occupations developed by agencies of the state government which deal with manpower education or employment.* These dictionaries, glossaries or lists will include a number of health occupations which are of little interest to health planners because of the setting in which they are usually employed or

*A list of health occupations is also provided in the NCHS publication entitled Health Resources Statistics: Health Manpower and Health Facilities, 1974. Most occupations included in this publication are defined within the text of the chapter dealing with each occupational grouping.

because they represent an insignificant proportion of the total health manpower work force (e.g., medical photographers, medical illustrators, etc.).

In order to keep the study manageable, it will be desirable to select a subset of health occupations from among the total work force. The process of selecting this subset of health occupations should be carefully carried out. It is recommended that representatives of hospitals, nursing homes and educational institutions be asked to participate in this process. The derived list of health occupations should then be submitted for review to a number of other persons knowledgeable about health manpower. Suggestions for additions or deletions of particular health occupations may be incorporated into the final list.

The final list of health occupations and their definitions will be used in developing and implementing study instruments. The list of health occupations and their definitions included in this study is as follows:[8]

- audiologist: specializes in diagnostic, habilitative, and rehabilitative services, and research related to hearing.

- biomedical equipment technician: tests and repairs electromedical equipment such as electrocardiographs, sterilizers, operating room lamps and tables, and diathermy machines, following schematic diagrams and using hand tools and test meters.

- biomedical instrumentation specialist: devises, sets up and operates electronic instrumentation and related electromechanical apparatus used in health care facilities and medical diagnostic and research laboratories. May check out sensing, telemetering, and recording instrumentation and circuitry and perform preventive and corrective maintainance of equipment.

- clinical dietetic technician: a skilled worker in nutritional care who works under supervision on a hospital dietetic staff.

·clinical dietitian (registered): directs and supervises hospital personnel concerned with planning, preparing, and serving food to patients and staff. May plan, organize, and conduct dietetics educational programs for nurses, medical and dental interns, medical residents, dietetics interns, and other personnel. May plan and direct preparation of modified diets prescribed by medical staff for patients with therapeutic diet needs.

·clinical social worker (registered): aids patients and their families with personal and environmental difficulties which predispose illness or interfere with obtaining maximum benefits from medical care.

·cytotechnologist: strains, mounts, and studies cells of human body to detect evidence of cancer and other pathological conditions. May be trained in cytogenetics to detect chromosome abnormalties.

·electrocardiograph technician: records electromotive variations in action of heart muscle on an electrocardiograph for diagnosis of health ailments. Studies patient records to record identifying information for EKG records. Prepares and positions patients, attaches electrodes, reviews recording for clarity and deviations. Keeps machine in general working order, and reports malfunctions to repairman.

·electroencephalograph technician: measures by means of an electroencephalograph (EEG), impulse frequencies and differences in electrical potential between various areas of the brain, to obtain data for use in diagnosis of brain disorders. Analyses patients' records for pertinent history in order to be alert for symptoms during test. Studies characteristics of tracings and calls attention to unusual patterns which may indicate brain disorders. Worker does not interpret wave patterns.

·emergency medical technician: an emerging occupation. Specializes in acute care techniques for handling of emergency patients in a hospital or in the field.

·hospital food service supervisor: supervises and coordinates activities of kitchen workers preparing and cooking food for hospital patients, staff, and visitors. Reviews menus, etc. with administrative dietitian, may inspect trays for attractiveness, etc., inspects purchased foods for standards of quality.

·licensed practical nurse: performs a wide variety of patient care activities and accomodative services for assigned patients, as directed by head nurse or team leader. Assists nurse, staff, or physician.

·<u>medical laboratory technician (medical laboratory assistant)</u>: conducts routine tests in clinical laboratories for use in treatment and diagnosis of disease, and performs related duties. Supervised by medical technologist.

·<u>medical records administrator (registered)</u>: supervises and coordinates activities of personnel engaged in analyzing, compiling, coding, indexing, and filing permanent medical records of patients; assists medical staff in research; prepares periodic and statistical reports; and provides information to authorized persons. Designs systems and methods to make data more accessible; brings unusual or interesting material to attention of medical staff.

·<u>medical records technician</u>: prepares statistical reports, codes diseases and operations according to accepted classification; maintains indexes according to established plans and procedures, and takes records to court. Handles requests for information, carries out routines of follow-up systems for patient care, may tabulate simple data for research and study.

·<u>medical technologist (registered)</u>: performs various chemical, microscopic, and bacteriologic tests to obtain data for use in diagnosis and treatment of disease. Obtains laboratory specimens, such as blood, urine, and sputum from ward or directly from patients, using established laboratory techniques. Posts all test findings for study by pathologist or other laboratory supervisor.

·<u>nuclear medicine technologist</u>: an emerging occupation. Operates radioscopic equipment, such as scintillation detectors and scanners, to produce scanograms, and measure concentrations of radioactive isotopes in specified body areas and body product to obtain information for use of physicians in diagnosing patient illnesses. Prepares radioactive isotopes for administration to patient. Performs laboratory tests and computes results for use by physician. Is responsible for disposal of radioactive waste, safe storage of radioactive material, and inventory and control of radiopharmaceuticals.

·<u>nurse (registered) (includes supervisor)</u>: renders professional nursing care to patients in support of medical care as directed by physician. Maintains patients' medical records on nursing observations and actions. Observes emotional stability of patients, expresses interest in their progress, and prepares them for continuing care after discharge. Licensed.

· nurse, anesthetist: a registered nurse who has post-graduate training of a formal nature in a recognized program which provides him (or her) with specialized skills in anesthesiology. Registered as a specialist. Licensed as R.N.

· nursing aide (orderlies): performs various patient care services and related services necessary in caring for the personal needs and comfort of patients, under supervision of registered nurse or staff.

· occupational therapist: organizes and conducts program to facilitate rehabilitation of physically or mentally handicapped, involving such activities as manual arts and crafts, practice in functional prevocational and home-making skills and activities of daily living, and participation in a variety of sensor-motor, educational recreational, and social activities designed to help patients regain physical or mental functioning or adjust to their handicaps. Registered.

· occupational therapy assistant: works with the professional occupational therapist in the conduct of a program to facilitate rehabilitation of physically or mentally handicapped patients.

· pharmacist: compounds and dispenses medications and other pharmaceutical supplies using standard physical and chemical procedures to fill written prescriptions issued by physicians, dentists, and other qualified prescribers. Licensed.

· pharmacy assistant: assists pharmacist to perform routine duties in hospital pharmacy.

· physical therapist: plans and conducts medically prescribed physical therapy program for individual patients, involving physical means, such as exercise, massage, heat, water, light and electricity. Applies diagnostic and prognostic muscle, nerve, joint, and functional ability tests. Licensed.

· physical therapy assistant: a skilled, technical worker who performs physical therapy treatments and related duties as assigned by the physical therapist. Works under supervision of physical therapist.

· physician assistant (registered): an emerging occupation. Skilled person, qualified by academic and clinical training, to provide patient services under the supervision and responsibility of a physician. Academic background may vary.

- <u>radiation therapy technologist</u>: an emerging occupation. Administers radiation and x-ray equipment, plastic hyperbaric chambers, and other laboratory equipment. Keeps informed of latest methods; compiles data for research physician. Shares responsibility for treatment records and assists in maintaining proper operation of controlling devices and equipment used in treatment.

- <u>radiologic technologist (registered)</u>: operates x-ray equipment to make radiographs of designated portions of the body and to provide x-ray therapy to patient under the supervision of an M.D. Radiologist, and performs related duties.

- <u>rehabilitation counselor (clinical)</u>: concerned with the guidance of the disabled person toward the full utilization of his employment potential. Reviews hospital charts and interprets data to field counselors. Engages in individual and group therapy with patients and families as an integral part of vocational guidance and planning. Coordinates evaluation of patients to determine degree of disability.

- <u>respiratory therapist (registered) (inhalation therapist)</u>: sets up and operates various types of oxygen and other therapeutic gas and mist inhalation equipment, such as iron lungs, tents, masks, catheters, cannulas, and incubators, to administer prescribed doses of medicinal gases and aerosolized drugs to hospital patients. Checks patients regularly, inspects, tests, and maintains equipment, orders repairs, keeps up with new developments, enforces safety rules.

- <u>respiratory therapy technician (inhalation therapy technician)</u>: sets up and operates various types of oxygen and other therapeutic gas and mist inhalation equipment to administer prescribed doses of medicinal gases and aerosolized drugs to hospitalized patients. Works under close supervision.

- <u>surgical technician (scrub technician, operating room technician)</u>: performs a variety of duties in an operating room to assist the surgical team.

The names and definitions of health occupations may vary from state to state as a result of licensure laws and the types of manpower training required. Where licensure or registration is an integral part of recognition as a member of a health occupation, this should be indicated in the definition.

In addition to the 33 health occupations defined above, the following variables are included in this study:

(For Survey of Demand in Hospitals and Nursing Homes)

- employer identifier: name and address of the employer, including zip code.
- type of employer: includes hospital, nursing care home, personal care home with nursing, personal care home without nursing or domiciliary care home.
- current employment: for each of the 33 health occupations previously defined, the number of full-time equivalent persons currently employed (excluding trainees). Full-time equivalents are to be calculated on the basis of hours worked.
- vacancies: the number of budgeted positions in each health occupation that are currently vacant and for which active recruiting of qualified persons is currently taking place.
- expected 1980 employment: the number of positions realistically expected on July 1, 1980 for each of the 33 health occupations.
- recruitment problems: the specific types of health manpower that are currently the most difficult to recruit.

(For Survey of Health Occupation Educational Programs)

- educational institution identifier: the name and address of the educational institution, including zip code.
- type of educational institution: includes vocational-technical high school, junior college, senior college, university and hospital.
- current status: whether for each manpower type a training program:
  - does not exist and is not planned
  - does not exist but is planned
  - currently exists but is being phased out
  - currently exists (no phase out planned)
- length of program: the normal time period in years, between entrance and completion of all existing and planned programs (rounding off to the nearest year).
- enrollment capacity: the approved student capacity for each existing and planned program per year.

- current enrollment: the number of students currently enrolled in each existing program.
- number of graduates previous year: the number of students who completed each program during the calendar year September 1 to August 31 preceding the study.
- expected graduates: the total number of graduates expected during the next four years from all existing and planned programs (i.e., the sum of the annual number of graduates expected for those years).
- percent of graduates locating in Central HSA: the percentage of graduates over the last two years who have located in Central HSA after completion of their training.
- jointly operated program: whether, for each health occupation, the institution's program is operated in cooperation with another institution (such as a hospital or college) which shares responsibility for the education or training of students. If yes, the name of the cooperating institution should also be indicated.
- placement problems: the specific types of health manpower that are currently the most difficult to place.

Although not included in this particular study, some HSAs may want to include a question in the Survey of Demand in Hospitals and Nursing Homes dealing with the attrition or turnover of certain types of health occupations. Such estimates would be useful in projecting future need for these occupations.

C. Study Instruments

Figures 1 and 3 present instruments designed to collect the variables defined above. Because these instruments are to be completed by respondents from the employers and educational institutions rather than by agency staff, it is important that they be kept straightforward and simple to complete. It is also important to limit the number of questions to those which are necessary for the purposes of the study. Each mailing should include, in addition to the appropriate instrument and cover letter, a list of study instructions and definitions (see Figures 2 and 4).

All information is to be entered by the respondent, with the exception of the employer identifier and type of employer (or educational institution identifier and type of educational institution) which will be completed on the instrument by the HSA staff prior to mailing the study materials. The study instruments are designed so that the data from this study can be processed, stored and analyzed without the use of a computer.

This study is designed to be implemented at the HSA level. A number of states have implemented statewide manpower studies.[9] In general, these studies utilized more complex instruments and more sophisticated techniques of data analysis than the study presented here.* However, they required a level of staff effort, expense, and computer capability which may exceed the resources of many HSAs.

*An example of one such study implemented at the state level is Allied Health Manpower in Texas 1973: A Report on Manpower Requirements, Resources and Education. This study was a joint effort of the Texas Hospital Association and the Texas Medical Foundation under a grant from the Regional Medical Program of Texas.

FIGURE 1

CENTRAL HSA SURVEY OF MANPOWER DEMAND IN HOSPITALS AND NURSING HOMES

SURVEY QUESTIONNAIRE

Do Not Use This Space
Editor________ Date________

John Q. Smith, Administrator
Anytown Community Hospital
456 Main Street
Anytown, State 12345

(Hospital)

Return to:

Manpower Study
Central HSA
123 State Street
Central City, State 01234

NOTE: *Before entering the requested information, please read the instructions attached to this questionnaire.*

Manpower Type	Current FTE Employment, Excluding Trainees July 1, 1976 (1)	Current Budgeted Positions Vacant and Open For Employment (2)	Expected FTE Employment on July 1, 1980 (3)
AUDIOLOGIST			
BIOMEDICAL EQUIPMENT TECHNICIAN			
BIOMEDICAL INSTRUMENTATION SPECIALIST			
CLINICAL DIETETIC TECHNICIAN			
CLINICAL DIETITIAN (REGISTERED)			
CLINICAL SOCIAL WORKER (REGISTERED)			
CYTOTECHNOLOGIST			
ELECTROCARDIOGRAPH TECHNICIAN			
ELECTROENCEPHALOGRAPH TECHNICIAN			
EMERGENCY MEDICAL TECHNICIAN			
HOSPITAL FOOD SERVICE SUPERVISOR			
LICENSED PRACTICAL NURSE			

Manpower Type	Current FTE Employment, Excluding Trainees July 1, 1976 (1)	Current Budgeted Positions Vacant and Open For Employment (2)	Expected FTE Employment on July 1, 1980 (3)
MEDICAL LABORATORY TECHNICIAN (MEDICAL LABORATORY ASSISTANT)			
MEDICAL RECORDS ADMINISTRATOR (REGISTERED)			
MEDICAL RECORDS TECHNICIAN			
MEDICAL TECHNOLOGIST (REGISTERED)			
NUCLEAR MEDICINE TECHNOLOGIST			
NURSE, REGISTERED (INCLUDES SUPERVISORS)			
NURSE ANESTHETIST			
NURSING AIDE (ORDERLIES)			
OCCUPATIONAL THERAPIST			
OCCUPATIONAL THERAPY ASSISTANT			
PHARMACIST			
PHARMACY ASSISTANT			
PHYSICAL THERAPIST			
PHYSICAL THERAPY ASSISTANT			
PHYSICIAN ASSISTANT (REGISTERED)			
RADIATION THERAPY TECHNOLOGIST			
RADIOLOGIC TECHNOLOGIST (REGISTERED)			
REHABILITATION COUNSELOR (CLINICAL)			
RESPIRATORY THERAPIST (REGISTERED) (INHALATION THERAPIST)			
RESPIRATORY THERAPY TECHNICIAN (INHALATION THERAPY TECHNICIAN)			
SURGICAL TECHNICIAN			

Please list the HEALTH OCCUPATIONS not covered by this survey which are essential to your operation. Exclude chiropractors, dentists, administrators, physicians and persons without specific health related knowledge and skills (such as secretaries, computer programmers, clerks, etc.).	Current FTE Employment, Excluding Trainees July 1, 1976 (1)	Current Budgeted Positions Vacant and Open For Employment (2)	Expected FTE Employment on July 1, 1980 (3)

(4) What health occupations specified above do you currently find most difficult to recruit?

________________ ________________

________________ ________________

________________ ________________

________________ ________________

________________ ________________

Signature____________________ Date____________
Representative completing this report

Name and Title ____________________ Phone____________
(please type)

FIGURE 2

Study Instructions

1. Please complete the requested information for each health occupation listed on the questionnaire whether or not you currently employ persons in these categories.

2. Before completing the information for each health occupation please familiarize yourself with the definitions of health occupations at the end of these instructions.

3. Enter in Column (1) the number of full-time equivalent persons currently employed on July 1, 1976 in each health occupation. Full-time equivalents should be calculated on the basis of hours worked with 35 hours equaling one full-time equivalent. Round off to one decimal place.

   Example: Anytown Hospital employs one full-time audiologist and two part-time audiologists. Hours worked are:

full-time person	40
part-time person 1	25
part-time person 2	30
total hours worked:	95

   $$\frac{95}{35} = 2.7$$

   Dividing total hours worked by 35 results in 2.7 full-time equivalents. The number 2.7 will be entered on the questionnaire.

4. Enter in Column (2) the number of budgeted positions in each health occupation which are currently vacant and for which you are now attempting to hire qualified persons. (Do not count positions you are not actively trying to fill.)

5. Enter in Column (3) the number of persons expected to be employed on July 1, 1980 for each health occupation, i.e., the number of positions expected to be budgeted for each health occupation at that time. (Do not count positions which are only "desirable" or "thought optimal".)

6. Complete Question (4) based upon your experiences in recruiting manpower over the past year.

IF YOU HAVE QUESTIONS ABOUT THIS STUDY TELEPHONE CENTRAL HSA AT: AREA CODE (000) 123-4567

Definitions of Health Occupations

- audiologist: specializes in diagnostic, habilitative and rehabilitative services, and research related to hearing.
-
-
-
-
- surgical technician (scrub technician, operating room technician): performs a variety of duties in an operating room to assist the surgical team.

FIGURE 3

CENTRAL HSA SURVEY OF HEALTH MANPOWER EDUCATION

SURVEY QUESTIONNAIRE

James A. Jones, Dean
Anytown Community College
Academic Drive
Anytown, State 12345

(Junior College)

'OTE: *Before entering the requested information, please read the instructions attached to this questionnaire.*

Manpower Type	Status (please refer to instructions)	Length of Program (In Years)	Enrollment Capacity	Current Enrollment
	(1)	(2)	(3)	(4)
AUDIOLOGIST				
BIOMEDICAL EQUIPMENT TECHNICIAN				
BIOMEDICAL INSTRUMENTATION SPECIALIST				
CLINICAL DIETETIC TECHNICIAN				
CLINICAL DIETITIAN (REGISTERED)				
CLINICAL SOCIAL WORKER (REGISTERED)				
CYTOTECHNOLOGIST				
ELECTROCARDIOGRAPH TECHNICIAN				
ELECTROENCEPHALOGRAPH TECHNICIAN				
EMERGENCY MEDICAL TECHNICIAN				
HOSPITAL FOOD SERVICE SUPERVISOR				
LICENSED PRACTICAL NURSE				
MEDICAL LABORATORY TECHNICIAN (MEDICAL LABORATORY ASSISTANT)				
MEDICAL RECORDS ADMINISTRATOR (REGISTERED)				
MEDICAL RECORDS TECHNICIAN				

Do Not Use This Space
Editor________ Date________

Return to:

Manpower Study
Central HSA
123 State Street
Central City, State 01234

Graduates Sept 1975-Aug 1976	Graduates Expected from Sept 1976-Aug 1980	Percent of Recent Graduates Locating in Central HSA Area	Jointly Operated Program (Check)	Name of Cooperating Institution
(5)	(6)	(7)	(8)	(9)

Figure 3, continued

Manpower Type	Status (please refer to instructions)	Length of Program (In Years)	Enrollment Capacity	Current Enrollment
	(1)	(2)	(3)	(4)
MEDICAL TECHNOLOGIST (REGISTERED)				
NUCLEAR MEDICINE TECHNOLOGIST				
NURSE, REGISTERED (INCLUDES SUPERVISORS)				
NURSE, ANESTHETIST				
NURSING AIDE (ORDERLIES)				
OCCUPATIONAL THERAPIST				
OCCUPATIONAL THERAPY ASSISTANT				
PHARMACIST				
PHARMACY ASSISTANT				
PHYSICAL THERAPIST				
PHYSICAL THERAPY ASSISTANT				
PHYSICIAN ASSISTANT (REGISTERED)				
RADIATION THERAPY TECHNOLOGIST				
RADIOLOGIC TECHNOLOGIST (REGISTERED)				
REHABILITATION COUNSELOR (CLINICAL)				
RESPIRATORY THERAPIST (REGISTERED) (INHALATION THERAPIST)				
RESPIRATORY THERAPY TECHNICIAN (INHALATION THERAPY TECHNICIAN)				
SURGICAL TECHNICIAN				

(10) Please list any of the active programs specified above which currently are experiencing problems in placing their graduates in appropriate employment

____

____

____

Graduates		Percent of Recent Graduates Locating in Central HSA Area	Jointly Operated Program (Check)	Name of Cooperating Institution
Sept 1975-Aug 1976	Expected from Sept 1976-Aug 1980			
(5)	(6)	(7)	(8)	(9)

Signature ______________________________ Date ______________

Name and Title ______________________________ Phone ______________
(please type)

FIGURE 4

Study Instructions

1. Please read all of these instructions and familiarize yourself with the definitions of manpower types provided before completing any information.

2. In Column (1) indicate the status of each type of health occupation training program in your institution using the following codes:

   1 - Does not exist and is not planned.
   2 - Does not exist but is planned.
   3 - Currently exists but is being phased out.
   4 - Currently exists (no phase out is planned).

3. In Column (2) indicate for each program coded 2, 3 or 4 in Column (1), the length of time normally required from entrance into the program until completion of the program. Round to the nearest year. (Programs requiring less than six months should be entered as 0. Programs requiring from six months to one year and five months should be entered as 1, etc...).

4. In Column (3) indicate the approved student capacity of all training programs for which you entered a code of 2, 3 or 4 in Column (1).

5. In Column (4) indicate the current enrollment of all programs coded 3 or 4 in Column (1).

6. In Column (5) indicate the number of graduates during the time period September 1, 1975 to August 31, 1976 from each training program coded 3 or 4 in Column (1).

7. In Column (6) indicate the total number of graduates expected during the four year period September 1, 1976 to August 31, 1980 from each training program coded 2, 3 or 4 in Column (1).

8. Consult mailing addresses of 1974-1976 graduates of all programs coded 3 or 4 in Column (1) to determine the percent of these graduates now living in Central HSA's service area (i.e., County 1, County 2, etc.). Enter this information in Column (7).

9. In Column (8) indicate whether any programs coded 2, 3 or 4 in Column (1) are operated jointly with another institution.

10. If any programs are checked in Column (8) place name of the institution(s) which cooperates in offering the joint program in Column (9).

11. Complete Question (10) based upon your experience in placing graduates of each training program over the past two years.

IF YOU HAVE QUESTIONS ABOUT THIS STUDY TELEPHONE CENTRAL HSA AT: AREA CODE (000) 123-4567

Definitions of Manpower Types

- audiologist: specializes in diagnostic, habilitative and rehabilitative services, and research related to hearing.
-
-
-
-
- surgical technician (scrub technician, operating room technician): performs a variety of duties in an operating room to assist the surgical team.

D. Data Collection Design

It is of paramount importance that a thorough investigation of the availability of extant data concerning health occupations be conducted before undertaking this study.* Some HSA's may find, for example, that state agencies concerned with the administration and governance of higher education programs have already collected all of the data on the education of health manpower which are needed to determine their current and future production. Similarly, data on the current supply of some types of health manpower may be available from agencies concerned with manpower accreditation and licensure, or from professional associations representing specific health occupations (e.g., medical societies, state licensure boards, state nurses' associations, etc.). The data from these sources may eliminate altogether the need to conduct this study, or they may be used to supplement the data which are collected in this study.

It is assumed that Central HSA has completed the inventories described in A Guide to the Development of Health Resource Inventories. The availability of these inventories is an important factor in the design and implementation of this study, because of the similarity of the data that are contained in these inventories and the data that are collected in this manpower study.

Definition of Universe

The study universe consists of all hospitals, nursing homes and health manpower educational institutions located in Central HSA's health service area. Health manpower educational institutions located outside Central HSA's health service area, but educating a significant proportion of specific-health occupations in the state are also included in the universe. The study involves a complete census of this universe.

*In several states, data of this type are available from the Health Manpower Statistical System (HMSS), which is a component of the Cooperative Health Statistics System (CHSS). The HMSS is described in A Guide to the Development of Health Resource Inventories (pp. 127-129).

Unit of Analysis

The unit of analysis for this study is the institution. Institutions include hospitals and nursing homes which employ persons in health occupations and educational institutions which train these persons.

Case Selection

Since data are to be gathered from all institutions in the study universe, case selection simply involves compiling a list of all entities which are a part of this universe. The inventories of hospitals and nursing homes described in A Guide to the Development of Health Resource Inventories provide the listing for these institutions. The listing of health manpower educational institutions described in the Guide is, however, not deemed sufficient for purposes of this study because it does not include all of the health occupations that are to be included in this study.

The list of educational institutions used in this study should include vocational-technical high schools, junior colleges, senior colleges, universities and hospitals. Names and addresses of vocational-technical high schools should be available from state agencies concerned with secondary education. The names and addresses of colleges and universities can be found in a number of sources. Most large libraries contain several publications which list these institutions.[10] State agencies administering higher education programs may also be able to provide up-to-date lists of colleges and universities. As noted above, a list of hospitals is available as a result of the agencies inventory activities.

Many HSAs may want to include in this study certain educational institutions which lie outside the boundaries of their health service areas. This may be particularly desirable when specific types of manpower are trained in relatively few institutions. For example, all radiologic technologists in Central HSA are trained in an institution which is not located in the agency's health service area. In order to include an appropriate proportion of this institution's graduates in estimates of the future availability of health manpower in Central HSA, it is necessary that this institution be included in the study. Institutions which train manpower for the entire state may be identified (and thereby added to the list) by consulting the following directories of health manpower educational programs:

Allied Medical Education Directory

American Medical Association
Council on Medical Education
535 North Dearborn
Chicago, Illinois 60610
(price - $7.50)

Allied Health Education Programs in Junior and Senior Colleges, 1973. Health Planners Edition

Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402

These publications list all institutions that provide training in health occupations by type in each city and state. They can, therefore, be used to identify health occupations for which only a limited number of training sites exist, or for which training takes place outside of the HSA. The colleges, universities or hospitals which are identified as training sites for these occupations may be added to the list of health manpower educational institutions that is to be used in this study.

In addition to identifying the names and addresses of institutions, the case selection process should identify the individual to whom the study materials will be mailed. For hospitals and nursing homes this individual will generally be the administrator. When the institution is a college or university, the study materials will usually be mailed to the dean. The extant data sources pertaining to institutions often provide the names of these persons. However, if the appropriate person cannot be identified from these sources the agency should telephone the institution in order to determine the name of the person to whom the materials should be sent. Similar telephone calls may be necessary to identify the person in vocational-technical high schools to whom the materials should be sent.

E. Field Procedures

Pretesting

Because this study deals with numero health occupations and employer and educational settings, it is important that it be pretested. A number of institutions representing hospitals, nursing homes and educational institutions should be asked to participate in the pretest. This pretest is useful for several purposes. It allows definitions of concepts and variables to be examined and criticized by persons with different backgrounds. The ability of the respondents from various institutions to quickly and accurately complete the study instrument can also be assessed. Finally, it allows adjustments to the lists and definitions of health occupations that are to be used in the study before a major commitment of time and effort has been expended.

The format of this pretest can be quite simple. The instrument to be pretested may be mailed to cooperating institutions with a request to complete the pretest by a specified date. HSA staff can then contact the institution and arrange a personal or telephone interview with the person assigned to complete the instrument. This interview should include a discussion of the appropriateness of definitions employed, the ease and accuracy with which each question may be completed, and the respondents' suggestions for changes in the instruments or in aspects of the study design.

Procedures to Elicit Study Participation

A number of procedures may be employed to elicit participation in the study. It is suggested that each institution in the study universe be contacted approximately two weeks prior to the mailing of the final study instruments. If possible, this contact should consist of a letter to the individual participant from an appropriate professional association or governing agency which endorses the conduct of the study and which encourages participation. A follow-up letter one week later from the HSA which outlines the purposes of the study and the type of data to be collected is also suggested.

The instrument, study instructions and definitions, and appropriate cover letters for the various types of institutions should be sent out in a single mailing. The cover letters should specify the date by which the completed questionnaires should be returned. A stamped, addressed return envelope should be included in this mailing.

Institutions which have not responded within one week after the date specified in the cover letter for return of the questionnaires should receive a second mailing. This mailing should be followed-up by a telephone call from the HSA or from the professional association or governing agency which initially contacted the institution. This call may offer the institution assistance in completing the questionnaires or suggest that a member of the HSA staff meet with a representative of the institution to complete the questionnaires.

Careful follow-ups which stress the usefulness of the data to participating institutions and the need for response from all institutions should help to increase the response rate.

Data Collection Personnel

The nature of this study makes it difficult to identify beforehand the person who will actually complete the questionnaires or to train such individuals. It is, therefore, doubly important that definitions and study instructions be carefully developed, and that the cover letter and instrument clearly stipulate that the respondent should carefully conform to the study instructions and definitions.

Data Collection Procedures and Quality Control

Data collection consists of the return of the completed questionnaires. Institutions which indicate that they desire assistance in completing the questionnaires should be allowed to schedule a visit between a member of the HSA staff and their representative. While the HSA should be reasonably fexible in scheduling this meeting, the institution should be encouraged not to delay the completion of the questionnaires.

A HSA staff member should also be assigned the task of checking all completed questionnaires returned by the respondents. This check would include a search for missing data and a check to make certain that all responses appear to be consistent with the size and type of institution. For example, if a small hospital indicates that it currently employs 30 nuclear medicine technologists and no registered nurses, the staff member may suspect that the responses to these two health occupations were entered in the wrong rows on the questionnaire. All such inconsistencies should be followed-up by telephoning the person who completed the questionnaire.

F. Data Processing and Storage

This study does not require coding or keypunching of data. The data may be stored on the original instruments. It is suggested, however, that employer data for hospitals and nursing homes be stored separately. When the data collectio ocess is completed, the data for indidual institutions may be aggreg. ted to provide data for counties or for the service area of the HSA. This process is very straightforward for the health employer questionnaire (Figure 1). Data from each column of this questionnaire is summed for each manpower type.

Similar summations should be performed for columns (3) through (6) of the educational institution questionnaire (Figure 3). Columns (1), (2), (8) and (9) of this questionnaire will be used to identify active programs, differentiate among training programs of different lengths*, and prevent duplication of data from jointly operated programs.

Editing of the data on the educational institution questionnaires should precede their aggregation. The following steps are suggested in the editing process: 1) identify jointly operated programs for which duplicate data have been reported by examining columns (8) and (9), and eliminate these duplications; and 2) identify occupations for which the length of training programs varies among institutions and determine whether these programs should be analyzed separately or combined in the data analysis.

After the editing of these data is completed, the data can be aggregated. In aggregating these data, however, the number of future graduates that can be expected to locate in Central HSA must be estimated. This requires that the numbers of graduates reported in column (6) be multiplied by a proportion equal to the percent locating in Central HSA reported in column (7). Thus, if a hospital reports that it expects to graduate 500 R.N.'s in the following four years and indicates that 80 percent of recent graduates have located in Central HSA's health service area, the number of graduates expected to locate in Central HSA is estimated as follows:

years	total expected graduates	x	proportion located in Central HSA	=	number expected to locate in Central HSA
1976-80	500		.80		400

Both the total graduates reported in columns (5) and (6) and the number locating in Central HSA may be of interest in data analysis. Thus, the data aggregated over the HSA for each occupation should include: enrollment capacity, current enrollment, number of graduates

*For example, HSA staff may wish to differentiate graduates of two or three year R.N. programs from graduates of four year (baccalaureate) programs. The data recorded in column (2) of the educational institution questionnaire can be used for this purpose.

last year, expected future graduates (next four years), and number of future graduates expected to locate in Central HSA.

Aggregated data will be the prime source of data analysis. However, the original instruments should be retained since the data may be useful for other purposes, such as the analysis of health manpower needs in specific cities or counties of the HSA.

II

DATA ANALYSIS AND USE

The data from this study can be used to address a number of issues directly related to the legally mandated function of HSAs. For example, development of general descriptions of the number, type and location of persons employed in health occupations in the health service area's hospitals and nursing homes is a necessary part of the larger task of HSAs to gather data concerning their area's health resources. Identification and documentation of shortages of personnel in certain health occupations will be useful to HSAs in making project review decisions. Similarly, data indicating current and future demand for health occupations in hospitals and nursing homes are useful to HSAs in fulfilling their responsibilities for the planning, promotion and implementation of health manpower development programs.

A. Identifying Manpower Shortages

The data from this study may be examined from a variety of perspectives. For example, the data may be examined by an HSA's facilities planner for the purpose of determining whether or not there is presently, or will be, an adequate supple of persons trained in health occupations needed to staff a proposed new facility or service. Conversely, the data may be examined by educators in order to determine whether openings presently exist, or will exist at some future time, for graduates of proposed or existing programs which train persons in specific health occupations. In the first case the emphasis is placed on the availability of manpower; in the second case the emphasis is placed on the availability of jobs. Whatever the emphasis, it is clear that comparisons need to be made between the data on the demand for persons in health occupations collected from hospitals and nursing homes, and the data on the production of persons in health occupations collected from educational institutions.

Table 1 presents data collected in this study for ten health occupations. Column (1) indicates the number of persons presently employed in these occupations in Central HSA hospitals and nursing homes. Obviously, these persons represent only one component, albeit a major one, of the total number of persons employed in each health occupation in Central HSA. Other settings such as doctors' offices, clinics, laboratories and industries also employ persons in these health occupations. The numbers presented in this column should, therefore, be interpreted only as accurate estimates of the number of persons employed in hospitals and nursing homes for each health occupation, not as estimates of the total number of persons employed in each health occupation in Central HSA's service area.

Column (2) of the Table indicates the number of budgeted vacant positions in each health occupation which the hospitals and nursing homes responding to the study questionnaire are trying to fill. These vacancies represent the demand for persons trained in health occupations which is not currently being met. Examination of the numbers in this column indicate that openings currently exist in all of the health occupations.

Column (3) indicates the expected demand for persons in health occupations in July, 1980. It represents the hospitals' and nursing homes' estimates of the number of positions in each health occupation which will be budgeted at that time. It is important to note that these estimates may be quite different from the actual number of positions which will be budgeted in July, 1980. Overestimates of growth by the persons completing the questionnaire, changes in the health service area's economy and a variety of other factors may make this estimate inaccurate. Nevertheless, it does provide an indication of the future need for persons in health occupations in the health service area.

Presented in column (4) are estimates of the number of graduates of the educational institutions included in this study who can be

TABLE 1

Employment and Production Data, Selected Health Occupations, Central H

Health Occupation	Current Demand Employ-ment (1)	Current Demand Vacancies (2)	Expected Demand July, 1980 (3)	1976-80 Grads Locating in Central HSA (4)	1980 Suppl (5) (1)+(
Clinical dietetic technician	75	2	81	8	83
Clinical dietitian	275	50	361	47	322
Electroencephalograph technician	18	2	24	6	24
Licensed practical nurse	3,300	167	3,801	512	3,812
Medical laboratory assistant	375	21	432	34	409
Medical records technician	267	13	312	35	302
Medical technologist	400	3	427	72	472
Nurse, registered	695	112	1,036	275	970
Nursing aide	875	2	1,200	86	961
Physical therapist	125	2	154	14	139

** indicates no shortage

TABLE 1

...nt and Production Data, Selected Health Occupations, Central HSA

...ccupation	Current Demand Employment	Vacancies	Expected Demand July, 1980	1976-80 Grads Locating in Central HSA	1980 Supply	1980 Shortage
	(1)	(2)	(3)	(4)	(5)	(6)
					(1)+(4)	(3)-(5)
...c technician	75	2	81	8	83	**
...an	275	50	361	47	322	39
...graph technician	18	2	24	6	24	**
...al nurse	3,300	167	3,801	512	3,812	**
...ry assistant	375	21	432	34	409	23
...technician	267	13	312	35	302	10
...gist	400	3	427	72	472	**
...d	695	112	1,036	275	970	66
	875	2	1,200	86	961	239
...st	125	2	154	14	139	15

** indicates no shortage

expected to locate in Central HSA during the next four years. In calculating the numbers presented in this column, it is assumed that recent trends in the location of graduates will continue over the next four years. If these trends change, the estimates presented in column (4) may overestimate or underestimate the actual numbers.

The numbers in column (5) are estimates of the supply in 1980 of persons in selected health occupations in Central HSA. It assumes that all persons presently employed in health occupations in hospitals and nursing homes will continue to be employed. It also assumes that new graduates locating in the HSA will seek employment.* A number of other factors obviously affect supply. These include migration of persons in and out of Central HSA, trained persons who re-enter the job market after periods of inactivity, and employment of persons in other settings such as physicians' offices. Data from this study do not, however, allow estimates of these components of supply to be made.

The numbers in column (6) are estimates of the shortage in 1980 of trained persons in selected health occupations in Central HSA. Obviously, a number of factors which cannot be predicted will influence whether a shortage actually develops. These include migration, employment in other health care settings and re-entry of persons into the labor market. The numbers in this column do indicate, however, that for six of the ten occupations, the current work force plus new graduates expected to settle in the HSA will be insufficient in number to meet the estimated demand.

Further analyses of the shortages indicated in Table 1 is possible through the use of vacancy rates. The vacancy rate is calculated as follows:

---

*It is also assumed that all persons entering these health occupations received appropriate training, and were counted in the survey of educational institutions. If this is not true, this estimate may be in error, e.g., if some institutions hire persons with no previous training as nursing aides, column (5) will underestimate the supply of nursing aides.

$$\text{vacancy rate} = \frac{V}{E + V} \times 100$$

where,

V = number of current vacancies /from column (2)/

E = current employment /from column (1)/

Among the various occupations employed in industrial and commercial firms, "vacancies up to 3 percent are regarded as norma.; 3 to 5 percent possibly serious, depending on the reason; and vacancies of 5 percent and greater are regarded as increasing evidence of possible critical shortage."[11] Table 2 presents these vacancy rates for the occupations included in Table 1. This table indicates that serious shortages of clinical dietitians, medical laboratory assistants and registered nurses currently exist in Central HSA and that current shortages of licensed practical nurses and medical records technicians may also be serious.* calculation of these vacancy rates provides both a method for ide ng shortages and a means for assessing the relative seriousness of shortages. For example, Table 2 clearly indicates that the shortages of dietitians and registered nurses are much more serious than the shortage of medical laboratory assistants. In interpreting these vacancy rates it is wise to avoid placing much confidence in rates computed on the basis of only a few cases. For example, vacancy rates based on total demand (E + V) of less than 50 are not calculated in Table 2. Rates based on 50 to 100 cases are calculated, but should be interpreted cautiously since filling one vacancy will decrease the vacancy rate by one to two percent.

Educatio al instit tions which train person the health occupations for which the shortage exists may use the inf rmation provided in Tables 1 and 2 in planning expansions or changes in their enrollments. Projected 1980 shortages are indicative of the extent to which additional persons may need to be educated in health occupations by these institutions in order to achieve balance between demand and the supply

*Vacancy rates do not conclusively indicate a shortage of trained personnel in the health service area. Vacancies may result from factors which are not addressed in this study, such as salary structure.

Table 2

Vacancy Rates for Selected th
Occupations, Central HSA

Health Occupation	Vacancy Rate
Clinical dietetic technician	2.6
Clinical dietitian	15.4
Electroencephalograph technician	*
Licensed practical nurse	4.8
Medical laboratory assistant	5.3
Medical records technician	4.6
Medical technologist	0.7
Nurse, registered	13.8
Nursing aide	0.2
Physical therapist	1.6

*Not calculated, E + V less than 50

of appropriately trained persons. Identification of shortages also suggest health occupations for which special recruiting or advanced training courses may provide the needed persons.

B. Assessing Hospital and Nursing Home Staffing Patterns

In addition to identifying shortages in manpower supply, the data from this study may be used to describe staffing patterns in the health service area's health care institutions. By combining the data from this study with data on the number of hospital and nursing home beds in the health service area, an index of staffing patterns can be calculated. This index is the number of full-time equivalents in a particular health occupation per 100 hospital or nursing home beds. This index may also be calculated for individual institutions. Comparisons of staffing patterns between individual institutions and Central HSA or the nation may help to indicate problem areas in terms of understaffing or overstaffing of particular health occupations. Table 3 presents an example of this type of comparison. It indicates that City Home's staffing pattern is somewhat different than that of the nation or Central HSA. City Home employs fewer registered nurses per 100 beds than Central HSA or the nation. At the same time, City Home employs more licensed practical nurses and nursing aides per 100 beds. A number of possible explanations exist for these differences. Staffing patterns in City Home may be the result of the previously noted shortage of registered nurses in Central HSA. This staffing pattern may also be the result of the type of care provided. By combining information on manpower from this study with information on the type of facility from the nursing home inventory described in A Guide to the Development of Health Resource Inventories, this second explanation can be addressed in more detail.

Staffing patterns for particular types of nursing homes may be calculated for Central HSA. For example City Home is listed in the inventory as a skilled nursing facility. Table 4 displays staffing patterns for skilled nursing facilities in Central HSA along with the previously calculated patterns for City Home. When City Home is compared with

Table 3

Number of Full-time Equivalent Staff in Selected Health Occupations Per 100 Beds in Nursing Homes: United States, 1973-74; Central HSA, 1976; City Home, 1976

Health Occupation	Staff Per 100 Beds		
	United States**	Central HSA	City Home
Nurse, registered	4.5	4.30	3.60
Licensed practical nurse	5.3	5.50	5.72
Nursing aide	28.0	29.70	31.52
Physical therapist	*	.58	1.24
Audiologist	*	.26	.00
Occupational therapist	*	.37	.66
Clinical dietitian	*	.52	.61

*Not available.
**Calculated from: National Center for Health Statistics, Division of Health Resources Utilization Statistics, Long Term Care Statistics Branch, "Preliminary Data from the 1973-1974 National Nursing Home Survey" Tables A and T.

Table 4

Number of Full-time Equivalent Staff in Selected Health Occupations Per 100 Beds in Skilled Nursing Facilities: Central HSA, 1976; City Home, 1976

Health Occupation	Staff Per 100 Beds	
	Central HSA	City Home
Nurse, registered	5.20	3.60
Licensed practical nurse	5.50	5.72
Nursing aide	30.10	31.52
Physical therapist	.75	1.[illegible]
Audiologist	.50	.00
Occupational therapist	.64	.66
Clinical dietitian	.34	.61

other homes of the same type the apparent understaffing of registered nurses is even more pronounced. Differences in staffing levels of audiologists and physical therapists also are indicated in these tables. However, because relatively few of these health occupations are employed in nursing homes, small changes in the actual number of employees can cause large changes in the data displayed in Tables 3 and 4. Data representing small numbers of personnel, therefore, require more caution in interpretation than data representing large numbers of personnel.

As a result of examining these staffing patterns, it is clear that City Home presently has fewer registered nurses per bed than is the case for Central HSA or the United States. Furthermore, City Home's response to the survey questionnaire indicated that they were not presently attempting to fill any vacant positions for registered nurses.

These descriptive data may be particularly useful when assessing competing proposals for new or expanded services. Institutions with staffing patterns which differ greatly from those of similar institutions in the HSA's health service area may be asked to explain these differences as part of the review and comment process.

Data on staffing patterns can also be used to estimate staffing need for proposed new institutions. For example, the number of people in various health occupations needed to staff a proposed new 250 bed skilled nursing facility can be estimated on the basis of the data obtained from this study. The data on staffing patterns in Central HSA's skilled nursing facilities presented in Table 4 can be used to estimate these staffing needs. Table 5 demonstrates the calculation of staffing needs for this proposed facility.

In Table 5 the staffing patterns in Central HSA as a whole are used to determine the number of employees needed in each health occupation in the proposed new facility. For example, 5.2 registered nurses per

Table 5

Calculation of Staffing Needs in Selected Health Occupations for Proposed 250 Bed Skilled Nursing Facility

Health Occupati	(1) Central HSA Staffing Per 100 Beds	(2) Number of Proposed Beds (100's)	St For Nur
Nurse, registered	5.20	2.5	
Licensed practical nurse	5.50	2.5	
Nursing aide	30.10	2.5	
Physical therapist	.75	2.5	
Audiologist	.50	2.5	
Occupational therapist	.64	2.5	
Clinical dietitian	.54	2.5	

Table 5

Calculation of Staffing Needs in Selected Health Occupations for Proposed 250 Bed Skilled Nursing Facility

Occupation	(1) Central HSA Staffing Per 100 Beds	(2) Number of Proposed Beds (100's)	(3) Staffing Needs For 250 Bed Skilled Nursing Facility
			(1) x (2)
ed	5.20	2.5	13.0
cal nurse	5.50	2.5	13.8
	30.10	2.5	75.3
ist	.75	2.5	1.9
	.50	2.5	1.3
erapist	.64	2.5	1.6
ian	.54	2.5	1.4

100 beds is the pattern of staffing in the HSA's health service area. Thus, the number of registered nurses needed in a 250 bed skilled nursing facility is $5.2 \times \frac{250}{100}$ or $5.2 \times 2.5 = 13$.*

Information on staffing needs presented in Table 5 is useful to both the persons proposing the new facility and the planner reviewing this proposal. In reviewing proposals of this type, P.L. 93-641 requires HSAs in their review process[12] to consider the availability of health manpower necessary to provide the new services. In examining the availability of manpower, the HSA must consider a number of questions, such as "Does a sufficient supply of appropriately trained persons needed to staff the facility presently exist?" "Will a sufficient supply of these persons exist when the proposed facility or expansion is completed?" Comparing the number and type of health occupations needed (Table 5) with current shortages (Table 2) indicates that it will be difficult to obtain at least two needed health occupations (i.e., clinical dietitians and registered nurses) to staff this facility. Table 2 also indicates an apparent, though less critical, shortage of licensed practical nurses. The applicant proposing this new skilled nursing facility should be required, therefore, to demonstrate the availability of registered nurses, licensed practical nurses and clinical dietitians to staff the proposed facility or to provide contingency plans for staffing should these health occupations not be available in sufficient numbers.

*In this example, current staffing levels are used as the standard or planning norm. HSAs should, however, develop manpower staffing standards on a more meaningful basis than the patterns currently existing within their service area's health facilities.

FOOTNOTES - SECTION III

Chapter 6

1. Public Law 93-641, Section 1513 (a).
2. *Ibid*., Section 1513 (b) (1) (D).
3. *Ibid*., Section 1513 (b) (1) (B).
4. *Ibid*., Section 1532 (c) (3).
5. *Ibid*., Section 1502 (1).
6. *Ibid*., (3).
7. *Ibid*., Section 1513 (a).
8. P.L. 93-222, Section 1306 (b) (5).
9. P.L. 93-641, Section 1532 (c) (8).
10. *Federal Register*, Vol. 40, No. 170, September 2, 1975, pp. 40315-415.
11. Bureau of Health Manpower, *Criteria and Procedures for Identification of Critical Health Manpower Shortage Areas*. Report No. 76-29, September 1975, p. 3.
12. Bureau of Health Planning and Resources Development, *CHP Program Letters* 74-10 (April 18, 1974), 75-09 (January 21, 1975), 75-11 (January 27, 1975); and Bureau of Community Health Service's letter (March 10, 1975) to areawide CHP agencies.
13. P.L. 93-641, Section 1513 (a).
14. *Ibid*., Section 1532 (c) (6).

Chapter 7

1. American Hospital Association, *Hospital Statistics*, Chicago: American Hospital Association, 1975, p. xviii.
2. *Ibid*.
3. *Ibid*.
4. *Ibid*., p. xix.

5. American Hospital Association, Guide to the Health Care Field, Chicago: American Hospital Association, 1975, p. 13.

6. Ibid.

7. Hospital Statistics, op. cit., pp. xvii-xviii.

8. J. H. Mueller, K. F. Schuessler, and H. L. Costner, Statistical Reasoning in Sociology (Second Edition), Boston: Houghton Mifflin Company, 1970, p. 180.

9. Ibid., p. 181.

10. L. S. Reed and H. Hollingsworth, How Many General Hospital Beds are Needed? Washington, D.C., U.S. Public Health Service Publication Number 309, 1953.

11. Avedis Donabedian, Aspects of Medical Care Administration: Specifying Requirements for Health Care, Cambridge: Harvard University Press, 1973, pp. 431-485.

12. Ibid., pp. 324-336.

13. Ibid., p. 323.

14. Odin W. Anderson and Paul B. Sheatsley, Hospital Use--A Survey of Patient and Physician Decisions, Chicago: Center for Health Administration (Research Series 24), 1967.

15. Ibid., p. 33.

16. Daniel L. Drosness, et al., "Uses of Daily Census Data in Determining Efficiency of Units," Hospitals, Vol. 41, December 1, 1967, pp. 45-48, and December 16, 1967, pp. 65-68; Morris London and Robert M. Sigmond, "How Week Ends and Holidays Affect Occupancy," The Modern Hospital, Vol. 97, August, 1961, pp. 79-83.

17. Anderson and Sheatsley, op. cit., p. 23.

18. Ibid., p. 24.

19. Ibid.

20. Robert Y. Garrett Jr., "Seven-Day Work Week Improves Service," The Modern Hospital, November 1964, pp. 114-116.

21. Donabedian, op. cit., p. 282.

22. Ibid.

23. Ibid., p. 283.