HRSA 89-179
3606

MATHEMATICA
Policy Research, Inc.

HRSA 89-179
3606

Contract No.: HRSA 240-89-0011
MPR Reference No.: 7866

# EVALUATION OF THE AREA RESOURCE FILE (ARF) DATA SYSTEM

[redacted] FINAL REPORT:
VOLUME I

March 9, 1990


Authors:
Judith Wooldridge
David Edson
Joshua Rozen
Lisa Schopler


Prepared for:

Department of Health and Human Services
Health Resources and Services Administration
5600 Fishers Lane, Rm. 8C-02
Rockville, Maryland 20857

Project Officer:

Stuart Bernstein


Prepared by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ. 08543-2393

Principal Investiagtors:

Judith Wooldridge
David Edson

## CONTENTS


Chapter — Page

VOLUME 1

EXECUTIVE SUMMARY . . . . . viii

I. INTRODUCTION AND BACKGROUND . . . . . 1

A. PURPOSE OF THE EVALUATION . . . . . 1

1. Context of the Evaluation . . . . . 1
2. Criteria for Evaluating the Area Resource File . . . . . 3

B. EVOLUTION OF THE AREA RESOURCE FILE . . . . . 3

1. The Origins and Development of the ARF System . . . . . 4
2. Structure, Content and Maintenance of the ARF System . . . . . 6

C. APPROACH TO THE EVALUATION . . . . . 9

1. Interviews with Users . . . . . 9
2 Literature Review. . . . . 12
3 Review of the Technical and User Documentation . . . . . 13
4 Documentation of the ARF System . . . . . 15
5 Evaluation of Findings and Identification of Changes . . . . . 15
6 Development of Recommendations . . . . . 16

D. OUTLINE OF THE REST OF THE REPORT . . . . . 17

II. USERS AND USES OF THE AREA RESOURCE FILE . . . . . 17

A. USERS AND NON-USERS OF ARF . . . . . 17

1. ARF Acquisitions . . . . . 17
. Users . . . . . 19
3. Non-Users . . . . . 23
4. Response Rates . . . . . 26
5. Attitudes Towards and Use of the ARF in HRSA . . . . . 26

CONTENTS (continued)


| Chapter | | Page |
|---|---|---|
| II (continued) | | |
| | 6. Overall Attitudes | 29 |
| | 7. Summary | 32 |
| B. | USES | 32 |
| | 1. Overall Uses | 33 |
| | 2. Uses by HRSA Staff | 35 |
| | 3. Publications | 37 |
| | 4. Findings from the Literature Review | 39 |
| | 5. Contractors Who Have Used the ARF | 41 |
| | 6. Data Elements Used | 43 |
| | 7. Data Elements Desired | 47 |
| | 8. Summary.. | 54 |
| C. | DATA QUALITY | 55 |
| | 1. Accuracy | 56 |
| | 2. Timeliness and Time Series | 64 |
| | 3. Comment and Summary | 69 |
| D. | EASE OF USE | 72 |
| | 1. Obtaining the ARF | 72 |
| | 2. Using the Data File | 77 |
| | 3. Understanding the Documentation | 79 |
| | 4. MPR's Evaluation of the ARF documentation | 81 |
| | 5. Suggestions for Improving the ARF and its Documentation | 88 |
| | 6. Summary.. | 88 |
| III. | ARF ORGANIZATION AND MANAGEMENT | 93 |
| A. | INTRODUCTION | 93 |
| B. | AN OVERVIEW OF INFORMATION FLOW THROUGHTHESYSTEM | 93 |
| | 1. The Extract Program | 95 |
| | 2. The Update Program | % |
| | 3. Other Programs | 97 |

CONTENTS (continued)


| Chapter | Page |
|---|---|
| III (continued) | |
| C. DETAILED DESCRIPTION OF SYSTEM COMPONENTS | 97 |
| 1. Flowcharts | 98 |
| 2. System Components and Processes | 98 |
| 3. Components' Inputs and Outputs | 112 |
| 4. Example of Component Interrelationships: LAP1 Data | 112 |
| 5. Processing Costs | 116 |
| D. INFORMATION MANAGEMENT AND MAINTENANCE PROCEDURES | 117 |
| 1. Quality Control | 117 |
| 2. Data Storage | 119 |
| 3. Backup.. | 119 |
| E. SYSTEM DOCUMENTATION | 120 |
| F. THE ACCESS SYSTEM | 121 |
| 1. Overview | 121 |
| 2. Capabilities | 121 |
| 3. How It Works | 122 |
| 4. Limitations | 123 |
| G. OTHER SUBSIDIARY FILES AND SYSTEM | 124 |
| 1. AHA Hospital Data Files | 124 |
| 2. Enhanced Files | 125 |
| 3. Graduate Medical Education Files | 125 |
| 4. Creation of Documentation | 125 |
| 5. Profile Reports | 126 |
| VI. ARF IN THE 1990s: SUMMARY AND RECOMMENDATIONS | 203 |
| A. BHPr GOALS FOR ARF | 204 |

CONTENTS (continued)


Chanter Page

VI (continued)

B. RECOMMENDATIONS: CURRENT USER BASE ..... 208

1. Data Content ..... 208
2. Utilization ..... 210
3. Documentation ..... 211

C. ADDITIONAL RECOMMENDATIONS: EXPANDED USER BASE WITHIN THE FEDERAL GOVERNMENT ..... 212

D. ADDITIONAL RECOMMENDATIONS: EXPANDED USER BASE OUTSIDE THE FEDERAL GOVERNMENT ..... 213

E. RECOMMENDATIONS CONCERNING DELIVERY/ ACCESS SYSTEMS ..... 214

1. A Hybrid Delivery/Access System ..... 214
2. Other Recommendations ..... 215

F. RECOMMENDATIONS CONCERNING SYSTEM EFFICIENCY ..... 216

VOLUME 2

APPENDIX A

APPENDIX B

APPENDIX C

APPENDIX D

APPENDIX E

HRSA 89-179

# EXECUTIVE SUMMARY

## A. BACKGROUND

Program planners and analysts of the health care system in the United States frequently need information about the availability of health care resources in local areas. In an effort to meet these needs, the Bureau of Health Professions (then the Bureau of Health Manpower) has developed the Area Resource File, a data base of health care resources and other information,, primarily organized by county. The Area Resource File (ARF) system has become a massive data base with supporting files and systems. The central file in the system, the basic county file, contains over 8,000 data elements for each county. Over the years the file has been augmented as new data sources became available. The system's design has been modified to improve user. access and to make the data available in new formats, such as floppy diskettes for use on personal computers. The diskettes contain extracts of the basic county file.

The principal types of data included in the basic county file are the various geographic identification codes and the following types of data: descriptions of the numbers of health professionals and health facilities by type; use of hospitals; health care expenditures; population characteristics; morbidity, mortality, and natality; economic characteristics; environment; and health professions training.

The ARF system is used by the BHPr and its contractors to support the Bureau's mission of providing intelligence on the supply and distribution of the health professions in the United States. Other agencies of the Federal government and state and local agencies also have used the system.

## B. PURPOSE OF THE EVALUATION

Each year about $2 million of BHPr's budget is set aside for meeting the requirements of Section 708 of the Public Health Service Act. These activities include data collection, analytical studies, and computer time. The Area Resource File is maintained using between 10 and 15 percent of this pool of funds. As a major item in the BHPr budget, the Area Resource File has, not surprisingly, come under scrutiny from the Bureau and from others in the Agency. The Bureau wants to be certain that the expenditure on the file yields the maximum benefit both within the Bureau and within the Agency, while maximizing outside use. The Bureau of Health Professions has therefore decided that, after 15 years of growth, and in the context of burgeoning new computer technologies, now is an appropriate time to evaluate the uses, accuracy, and organization of the Area Resource File. The Bureau awarded a contract to Mathematica Policy

Research Inc. (MPR) to evaluate these aspects of the system and to develop (and cost out) recommendations for improvements to the system.'

The evaluation addresses the following questions:

- For what purposes is the file used and by whom?
- How accurate is the file?
- How is the file and its supporting system organized?
- Could the file's usefulness, accuracy, and organization be improved?

This report presents the results of an evaluation of these questions. The report also describes how the file and its supporting system and documentation might be modified to best satisfy user needs given the Bureau of Health Profession's resource constraints.

## C. APPROACH TO THE EVALUATION

### 1. User Interviews

The principal approach followed for collecting information about ARF usage was to interview users about their experiences. In addition, to provide background for these user interviews, we also interviewed key HRSA staff to find out how the ARF is used in different offices.

### 2. Literature Search

The contract called for a review of the appropriate literature for ARF citations to identify and evaluate the prevalence of studies utilizing the ARF and the nature of its use. We used several approaches for conducting a search for citations of ARF use:

- Formal search of computerized data bases of health literature
- Asking interviewees for citations of reports that they had prepared using the ARF
- Search of specific BHPr publications for citations

### 3. Review of Documentation

We reviewed the documentation provided to ARF users to establish whether the documentation provided sufficient information. Criteria for evaluating the documentation included whether the user could easily find a variable in the documentation, whether the user

---

'Contract No. HRSA 2404394011.

could easily determine how the variable was coded, and whether limitations of the data element (or source) were reasonably obvious to the user. Specifically, we looked at organization of the sections of the documentation, and within each section, at variable order. We looked in the technical documentation for consistency in both naming conventions and variable descriptions. We also looked for clarity in the explanations of variable content provided in the one line description in the technical documentation and in the longer explanations given in the user documentation. In addition, we looked for linkages and for consistency between the user and technical documentation.

### 4. Review of Svstem Efficiency

The ARF system was also evaluated with respect to the organization and efficiency of the programs and procedures used to create the existing ARF. Our approach to this task was to review the existing system documentation and discuss the ARF processing with the ARF contractor. The key components and processes of the ARF system were then documented. Potential modifications designed to promote processing efficiency were identified.

## D. RESULTS

### 1. User Interviews

ARF users are found in the Public Health service, other federal government agencies, state government, and outside of government. Each year, about 40 copies of the ARF file are distributed to users outside the Public Health Service, with nearly one third delivered to federal and state agencies and about half to academia and contractors. Among the federal health agencies that we spoke with (all of which had received the file) relatively little use has been made internally (only one current user was found among HCFA, PROPac, and PPRC). In HRSA, most direct use of the file is made in BHPr and BHCDA, especially in ODAM and ODM, the respective data management offices, where technical programming skills are more common.

With a few caveats, the initial attitudes (before we had asked any detailed questions) to the ARF among users and attitudes to ODAM (the office which supports it) were positive, with users concerned that ARF should be maintained in the future. Nevertheless, HRSA staff mentioned concerns about the high cost of the ARF contract, and questioned whether it could be reduced. Some users also commented that the distribution of the ARF data file and information about file contents could be improved. One or two HRSA users mentioned data problems.

Considering the size of the file, remarkably few errors were identified by users. Some of these were errors of labeling rather than content. A few fields need to be corrected while others need to be more clearly labelled. There are a few problems that relate to the manner in which data are aggregated that make the file awkward to use or uninformative. These problems could

be addressed by changes in policy toward data aggregation. Currently, the ARF is viewed as providing value because the user does not have to seek out and combine the separate data sources. Some users would like a little more value added in the form of more data checks on source data before they are added to the file.

It is clear from what respondents told us that errors in the file are remembered long after they have been corrected. To avoid this problem in the future, ODAM would have to undertake a more thorough review before adding data to the file, (including more discussions with the source or persons who had used the source data and review of the data). To assist in this process, a formal set of criteria for inclusion (or exclusion) could be developed and described in the documentation. Another way of handling the problem might be improved communication with users.

With respect to timeliness, it may never be possible for a secondary data file like ARF to be timely enough for activities related to program implementation. However the file is sufficiently timely for many descriptive and comparative analyses of the health care system. BHPr cannot control the collection of adequate, timely, and equivalent data across years, because it does not have the budget to do so. Nevertheless, this creates complexities for the users of the file. While we have focused on some of the best data in the file, there are greater problems of timeliness and dislocation of time series for many other variables (e.g., allied health professions). ODAM could assist the user by providing better documentation on exactly what is and is not available in the file across years.

Most users had no difficulty obtaining the ARF and its documentation, and most users obtained these in the basic county file format from ODAM. More than half of the users knew of the alternative formats and the source files of ARF, but few had used them.

Users tended either to use the ACCESS system available at the NIH data center or to use the Statistical Analysis System (SAS) software package to extract data from ARF. Some HRSA users suggested that they be provided with routine notice of updates to the file and changes in passwords. One user had a problem with a technical constraint of ACCESS.

2. Literature Search

Users were able to provide an extensive list of publications that drew on the ARE A review of ten of these publications showed that the ARF was not cited in four out of ten. Thus it is not surprising that an automated literature search found only three publications with ARF citations, of which one was a BHPr publication about the ARF. Lack of formal citation made it difficult to be sure how the ARF was used in some analyses, including those that are presented in BHPr's Sixth Reoort to the President and Congress on the Status of Health Personnel in the United States.

3. Review of Documentation

A thorough evaluation of the documentation uncovered some deficiencies in layout and consistency that may have contributed to users' lack of knowledge about what was in the file. For some variables, considerable effort is needed before the nature and content of the variable can be fully understood. The user documentation can be awkward to use due to the organization of the documentation and inconsistencies with the technical documentation.
We also found a small number of errors in the variable descriptions.

4. Review of Svstem Efficiency

Despite its large size and complexity, the current ARF system is a well-organized and generally efficient system. However, the review revealed several areas where relatively minor modifications would improve the functioning and maintenance of the system:

- The number and detail of system flowcharts in the system documentation could be expanded. This would help reduce future development and maintenance costs.
- A re-examination and streamlining of selected merge procedures offers savings in storage and processing costs.
- Increasing the use of routines written in procedural languages rather than SAS will reduce processing time.
- The use of data entry software packages could reduce development costs relative to the current approach which utilizes procedural languages to custom-built data entry routines.

## E. POTENTIAL CHANGES

Potential changes to the ARF system were identified from the user interviews and the review of the ARF data system and user documentation. These potential changes, plus some that were contractually mandated, were categorized into five types of issues: under-utilization of the ARF, content and structure of the data, content and structure of the documentation, difficulty in accessing the existing data, and the cost of the existing system. At this stage, each potential modification was reviewed with respect to criteria addressing the following: the goal; the need; the impact on the current system procedures; impact on end-users; development, implementation and operational costs; requirements for additional user training; requirements for additional software and hardware for the ARF contractor and the users; alternatives; and the feasibility of success. As a result of this review, a refined list of modifications was prepared and the modifications were labeled as critical, important, or useful. The modifications classified as critical or important were subject to further evaluation. Those classified only as useful were not

evaluated further. Table 1 shows the list of modifications with priority rankings. Cost estimates were made for the highest priority modifications.

## F. RECOMMENDATIONS

The wide range of suggested modifications reflects not only the complexity of the ARF system, but also the variety of opinions within BHPr and the user community about the purpose of the current ARF system and data files. Prior to committing additional resources to the ARF system, we feel that BHPr must make some fundamental decisions about the goals of the ARF in the 1990s. Specifically, BHPr must decide:

- Who is the ARF prepared for?
- What should the ARF be?
- Given these goals, is ARF being prepared as cost-effectively and efficiently as possible?

The specific recommendations we have made are appropriate only when considered within the context of the target population which BHPr wishes to serve with the ARF. We have stratified the potential target population into three groups:

- Current users
- Current users plus additional non-technical users within the Federal Government
- Current users plus additional non-technical users both within and outside the Federal Government

Within the context of serving the current user base, MPR feels that a number of relatively minor modifications to the ARF procedures and documentation would improve the ARF. Expanding the user base to include additional non-technical users within the Federal Government would require additional modifications--mainly a more aggressive attitude towards cleaning and editing data and the development of a documentation codebook. Expanding the user base to non-technical users outside the federal government would require the same changes plus additional outreach activities.

The recommendations for each group are summarized in Exhibit 1 and are discussed below.

### 1. Current Users

The ARF appears to be more or less what its current users need. It was clear from the interviews that most users were reasonably satisfied with the ARF's content. Naturally, some

TABLE 1

Overview of Proposals Addressing Problem Areas

| PROBLEM AREA | PROPOSED RESPONSE | CODE | SPECIFIC MODIFICATION TO BE CONSIDERED | REFERENCE NUMBER |
|---|---|---|---|---|
| Underutilization | Expand Outreach Activities | C | Expand Product Information Dissemination (Newsletter, Professional Meetings, etc.) | Al |
| | | | Develop Newsletter/Disseminate Information | |
| | | | Develop/Revise Brochures, Materials | |
| | | U | Organize User Group | A2 |
| | | I | Provide an Electronic Codebook | A3 |
| Data Problems | Correct Errors/Improve Timeliness | C | Correct Specific Data Errors on Existing File | B1 |
| | | I | Expanded Data | B2 |
| | | | Add New Data Sources | |
| | | | Focus on Obtaining Complete Time-Series | |
| | | | Add New Geographic Codes | |
| | | | Expand Population, Morbidity, and Mortality Data | |
| | | C | Data Editing Procedures (Front-End) | B3 |
| | | | Standardize Data Coding Procedures | |
| | | | Provide Explicit Criteria for Inclusion/Exclusion of Data | |
| | | | Standardize Missing Data Codes | |
| | | | Reevaluate Policy for Virginia Independent Cities | |
| | | | Improve Categorization of Data | |
| | | | Implement Better Procedures to Avoid Duplicates | |
| | | U | Speed Data Acquisition/Alter Distribution Cycle | B4 |
| Organization /structure | Modify Structure/Organization | U | Provide Data at Other Geographical Levels | C1 |
| | | U | Create Current Cross-sectional: Historical/Time-Series Files | C2 |
| | | U | Expand Availability of Profile Reports | c3 |
| Documentation | Modify Documentation | C | Improve Existing Documentation | Dl |
| | | | Include a Dump of Data with the Documentation | |
| | | | Expand Table of Contents | |
| | | | Correct Existing Identified Errors | |
| | | | Provide Complete Description of Source Data and References | |
| | | | Provide a Contact Person for Source Data | |
| | | | Document and Utilize a Consistent Set of Abbreviations | |
| | | | Explain Suppression Fields and Flags | |
| | | | Provide Tables of Available Time-Series Data | |
| | | I | Develop Codebook | D2 |
| | | | Provide Cross-Walk Between User and Tech. Documentation | |
| | | | Coordinate Variable Order in User and Tech. Documentation | |
| | | | Page References linking Tech. to User Documentation | |
| | | | Move Descriptions to User Documentation/Add Complete Variable Descriptions | |
| | | I | Add Summary Statistics to Codebook | D3 |

TABLE 1

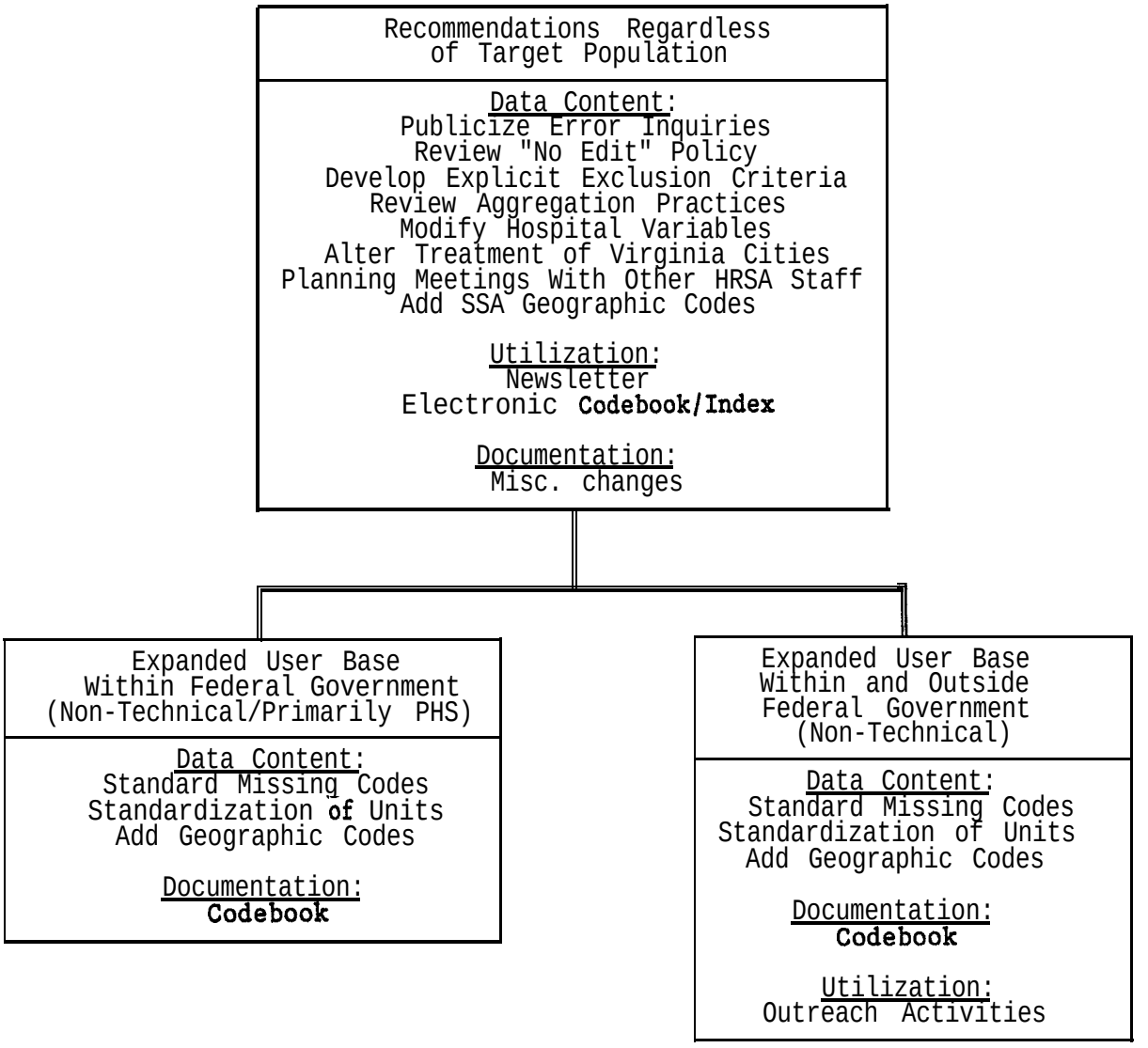
Overview of Proposals Addressing Problem Areas
(continued)

| PROBLEM AREA | PROPOSED RESPONSE | CODE | SPECIFIC MODIFICATION TO BE CONSIDERED | REFERENCE NUMBER |
|---|---|---|---|---|
| Access/Delivery | Alter Access/Delivery Methods | | Manual: | |
| | | I | Magnetic Tape (Alternative Formats) | E1 |
| | | I | Distribute on Alternative Media (Optical Disk) | E2 |
| | | | Computerized: | |
| | | U | Automated Access to Individual Data Elements/Observations | E3 |
| | | | Manual Access to Individual Data Elements/Observations | |
| | | | Mainframe: | |
| | | I | Flat File/Non-menu Driven Software | E4 |
| | | I | Flat File/Menu Driven Software | E5 |
| | | I | DBMS/Non-menu drive Software | E6 |
| | | I | DBMS/Menu-driven Software | E7 |
| | | | Microcomputer: | |
| | | I | Flat File/Non-menu Driven Software | E8 |
| | | I | Flat File/Menu Driven Software | E9 |
| | | I | DBMS/Non-menu drive Software | E10 |
| | | I | DBMS/Menu-driven Software | E11 |
| | | | Manual Access to ARF Data Files | |
| | | I | Microcomputer Bulletin Board | E12 |
| Efficiency/Cost | Modify system to reduce costs | U | Store Data in Compressed Format | F1 |
| | | C | Expand Flowcharts in System Documentation | F2 |
| | | I | Reexamine/streamline Merge Procedures | F3 |
| | | I | Utilize Procedural Languages | F4 |
| | | I | Computerized Data Entry | F5 |
| | | U | Artificial intelligence-based algorithms | F6 |
| | | I | DBMS | F7 |
| | | I | Microcomputer | F8 |

CODE: C - Critical: I - Important: U - Useful

**EXHIBIT 1**

**RECOMMENDATIONS RELATED TO ARF TARGET POPULATIONS**



users could think of improvements, and some had specific suggestions for new data that should be added. These suggestions may be considered as marginal improvements, inasmuch as there were no major changes in content requested.

We make the following specific recommendations to increase the usefulness of the ARF for current users:

*BHPr should produce and distribute a newsletter about ARF to all the HRSA bureaus and known ARF users.*

*When BHPr receives a query about data errors, a marked effort should be made to respond directly to the query and to let other users know (via the newsletter) whether or not a change is needed*
*BHPr should review its current "no editing" policy (the policy is not, in any case, applied absolutely consistently, since data at the sub-county level have to be aggregated and implicit decisions get made about how to enter the data in the county file).*

*BHPr should set out explicit criteria for whether or not data should be included in the file.*

*BHPr should conduct an ongoing review of current aggregation practices. Specific changes should be made immediately to the coding of the hospital occupancy rate and hospital bed variables.*

*BHPr should consider adding a trailer file appended to the main file and tables showing for which years and variables Virginia cities are combined with the county and for which they are separate.*

*BHPr should consider looking for duplicates in source data before aggregating sub-county level data and adding it to the ARF.*

*We recommend more discussions between the ARF staff and other HRSA staff about which additional data might be included on the file. In particular, we recommend that at the start of every new cycle of ARF production, ODAM and other HRSA staff meet to discuss potential additions and changes for the next public version of the ARF.*

*BHPr should look at one area of data content: the addition of new geographic codes to the ARF. If it is administratively feasible to add the SSA state and county codes, we believe it would be useful to HCFA's contractors.*

*BHPr and the ARF contractor should develop an electronic codebook as a supplement to the printed ARF documentation*

BHPr *should make improvements to the existing documentation. Of the possible improvements that we reviewed, we rank the order of importance as follows:*

1. *Correct existing errors (an ongoing activity that should be maintained)*
2. *Include a dump of several records in the documentation*
3. *Provide a more detailed contents to the technical documentation*
4. *Prepare tables of time-series data available in ARF*
5. *Provide complete descriptions of source data with references*
6. *Provide a list and explanation of the suppression fields and flags*
7. *Provide the names of contact persons for source data whenever possible*

*BHPr should also include summary statistics as part of the ARF documentation, although we view this as less important than the other recommendations.*

Modifying historical data currently in the file would be quite expensive. Therefore, it may be a better use of BHPr resources to make the changes for future additions to the file only.

## C. ADDITIONAL RECOMMENDATIONS: EXPANDED USER BASE WITHIN THE FEDERAL GOVERNMENT

If the decision is made to expand the user base outside of BHPr, particularly if the target population is not technically trained in programming and research, we believe that a fully-integrated database would be advisable. We make the following recommendations:

*Missing values should receive a standardized and unambiguous code.*

*Standardization of units across comparable variables, such as population, should be introduced*

*BHPr* should *include other types of geographic codes like the New England Consolidated Metropolitan Area code in the ARF.*

*BHPr may also wish to send the newsletter to other health agencies as well as other regular users in the federal government (such as CBO and GAO) and possibly to regular users outside the federal government.*

*If BHPr decides to move to a fully integrated database, with the goal of increasing the number and type of users, a full codebook should be developed*

## D. ADDITIONAL RECOMMENDATIONS: EXPANDED USER BASE OUTSIDE THE FEDERAL GOVERNMENT

To increase the user base outside the Public Health Service would require more expensive approaches such as using a marketing organization to develop brochures and disseminate them

to new potential markets. We believe that BHPr must make a very careful assessment of the costs and advantages of expanding the user base outside of the Public Health Service. An increased user base will generate increased numbers of technical assistance questions from a larger base of nontechnical users. A wide expansion in the user base of the ARF would require changes in documentation and access and delivery methods. These changes would be costly. If there were a way of collecting a fee for distribution of the file, there would be some financial advantage to BHPr of increasing use among outsiders that might offset the increased costs. We are not aware of any methods that BHPr could now employ to collect and benefit from such fees.' Thus:

> *We do not recommend that BHPr undertake a major expansion of the ARF user base outside the federal government with the intent of generating a significant revenue stream It is not clear to us that there is a large market for a research file like ARF. A substantial expansion of users would likely increase costs without commensurate financial benefits. However, with some modifications and outreach activities, the ARF would provide a useful service to the research community and generate positive publicity for BHPr.*

## E. RECOMMENDATIONS CONCERNING DELIVERY/ACCESS SYSTEMS

The delivery and access methods selected for the basic ARF are tied closely to the fundamental question: Who is the ARF for?

### 1. A Hybrid Delivery/Access System

There are basic differences in the needs of researchers wishing to use the basic ARF file. Some analyses require processing large amounts of data while others require the use of a relatively small number of data elements. There is also a wide range in technical programing skills of researchers.

In developing the ARF system, ODAM has attempted to address the conflicting needs of researchers by implementing a hybrid delivery/access system. For those researchers with the need to process relatively large amounts of data, a magnetic tape containing the basic ARF is available. Use of the tape version of the basic ARF requires some technical programming skills.

For those researchers within the federal government wishing to access only a few variables, ODAM developed the ACCESS System. The ACCESS System is an on-line system designed to simplify the creation of data extracts and the use of the extracts with statistical, tabulation, and graphics software. The system is intended for both technical and non-technical users.

---

'Fees collected in this way would have to be returned to the U.S. Treasury.

The ACCESS System has not received strong user acceptance. The ACCESS System apparently is considered intimidating by novice computer users. Users must have NIH computer center accounts, therefore limiting users to federal agencies and their contractors. Many users are also interested in using their microcomputers for analyzing and displaying ARF data. Obtaining data from the ACCESS System for use on microcomputers is too cumbersome for many.

The two groups of users are both important. In order to satisfy both groups, BHPr should continue the hybrid approach to delivery of ARF data:

*We recommend that BHPr continue to distribute the basic ARF in character format on magnetic tape.*

Furthermore:

*We recommend that BHPr discontinue support for the ACCESS System and that BHPr replace it with a microcomputer-based Bulletin Board System.*

2. Other Recommendations

Other recommendations involving access/delivery methods are:

*We recommend that BHPr delay using optical disk technologies for distributing ARF data. We suggest that BHPr address the use of optical disks technologies at the time of the re-competition of the next ARF support contract.*

If BHPr is interested in providing users with direct access to specific records or data elements (as compared to the data extracts available on the Bulletin Board), there are a number of options available:

*We conditionally recommend that the ACCESS System be discontinued and replaced with a DBMS-based system. The recommended DBMS is ORACLE. MPR feels that maintaining the current ACCESS System, with its low usage rates, is not justified because of its costs.*

## F. RECOMMENDATIONS CONCERNING SYSTEM EFFICIENCY

The ARF system is generally efficient and significant reductions in processing costs are not likely. There are, however, several changes or issues which could be addressed in order to reduce processing costs and facilitate future maintenance and development:

*We recommend that the ARF contractor expand the number and detail of flowcharts contained in the ARF system documentation. The ARF contractor should use*

*microcomputer flowcharting software in preparing the flowcharts The expanded set of system flowcharts will be valuable for future maintenance and development activities.*

*The ARF contractor should provide BHPr with a report justifying the continued use of SAS rather than more efficient procedural languages such as COBOL. MPR believes that selected processes within the ARF system could be made more efficient through the use of procedural languages, especially when using microcomputers.*

*The ARF contractor should customize the EXTRACT and UPDATE procedures for selected source files in order to allow multiple variables per record This will result in substantial reductions in the size of selected transaction files.*

*The ARF contractor should justify the use of SAS screens and menus rather than data entry software such as QPL for data entry. MPR feels that development costs would be lower and that more flexible data entry routines would be produced using QPL.*

*The ARF contractor should continue the transfer of processing from the mainframe to a microcomputer environment.*

*The ARF contractor should provide BHPr with a system design report providing details of an ARF system based on a DBMS such as ORACLE. The design report should describe the file structures, storage requirements, processing efficiency, and other costs of using the DBMS as a tool in the development of the ARF* files. *Information in the report will enable BHPr and the ARF contractor to effectively analyze the implications of using new software technologies in the ARF system.*

# I. INTRODUCTION AND BACKGROUND

Program planners and analysts of the health care system in the United States frequently need information about the availability of health care resources in local areas. In an effort to meet those needs, the Bureau of Health Professions (then the Bureau of Health Manpower) developed the Area Resource File (ARF), a data base of health care resources and other information, primarily organized by county.

The Bureau of Health Professions has decided that, after 15 years of growth, and in the context of burgeoning new computer technologies, now is an appropriate time to evaluate the uses, accuracy, and organization of the Area Resource File. The Bureau awarded a contract to Mathematica Policy Research Inc. (MPR) that would evaluate these aspects of the system and that would develop (and cost out) recommendations for improvements to the system.'

## A. PURPOSE OF THE EVALUATION

### 1. Context of the Evaluation

The Bureau of Health Professions is part of the Health Resources and Services Administration (HRSA). HRSA's mission is "the support and delivery of health services to the disadvantaged and the underserved--those with limited access to health care--and the development of resources-qualified health professionals and facilities--to meet the needs of the nation." (HRSA, A Profile, April 1989.) Thus for HRSA to fulfill its mission, it must assess what resources the nation currently has, where they are located, and whom they serve. This process enables HRSA to determine which professions and geographic areas need additional resources.

---

'Contract No. HRSA 240-89-0011.

The Bureau of Health Professions "monitors and guides the development of health resources by providing leadership to improve the education, training, distribution, utilization, supply and quality of the Nation's health personnel" (ibid). Within BHPr, the Office of Data Analysis and Management (ODAM) provides support to the Bureau's analytical program, and also provides analytical and technical assistance to HRSA and outsiders. ODAM created the ARF in response to a critical need in the early 1970s for data to aid in establishing the Area Health Education Centers (AHECs). Recognizing the value of the file, it has since been augmented and maintained by ODAM and is intended to help BHPr and ODAM to meet their goals.

Each year about $2 million of BHPr's budget is set aside for meeting the requirements of Section 708 of the Public Health Service Act. These activities include data collection, analytical studies, and computer time. The Area Resource File is maintained using between 10 and 15 percent of this pool of funds. As a major item in the BHPr budget, the Area Resource File has, not surprisingly, come under scrutiny from the Bureau (and from others in the Agency). The Bureau wants to be certain that the expenditure on the file yields the maximum benefit both within the Bureau and within the Agency, while maximizing outside use.

The Bureau of Health Professions requested an evaluation of the following questions:

- For what purposes is the file used and by whom?
- How accurate is the file?
- How is the file and its supporting system organized?
- Could the file's usefulness, accuracy, and organization be improved?

This report presents the results of an evaluation of these questions. The report also describes how the file and its supporting system and documentation might be modified to optimize user needs and the Bureau of Health Profession's resources.

2. Criteria for Evaluating the Area Resource File

The evaluation of the Area Resource file is limited to answering the questions stated above. The evaluation followed several stages. The first stage was to gather data from a number of sources concerning the file's uses and structure. The information gathered from these sources was then evaluated in terms of who the ARF users were and what the uses were, and whether the file met reasonable standards of accuracy, accessibility, and technical organization and processing. Where standards appeared not to be met, potential changes that would meet these standards were identified. These potential changes were evaluated in terms of their importance and their approximate cost, in light of the requirement that changes should be evaluated as to whether they could be made under the current ARF contract (that is they would add little to the cost of that contract). Those recommendations for change that reached a certain level of importance and reasonable-cost (plus certain changes that the contract mandated be evaluated) were fully evaluated in terms of their cost and recommendations for adoption or not were made. Thus, changes that would require purchase of very expensive hardware, for example, would not be recommended following these criteria.

B. EVOLUTION OF THE AREA RESOURCE FILE

The ARF system is a massive data base with supporting files and systems. The central file in the system, the basic county file, contains over 8,000 data elements for each county, describing amounts of health care professionals and other health resources, training of health professionals,

utilization of health care, morbidity and mortality, and demographic, environmental and economic characteristics. The county file has been augmented over the years as new data sources became available. In addition, the system's design has been modified to improve user access and to make the data available in new formats, such as floppy diskettes for use on personal computers. The diskettes contain extracts of the basic county file.

The ARF system is used by the Bureau of Health Professions (BHPr) and its contractors to support the Bureau's mission of providing intelligence on the supply and distribution of the health professions in the United States. Many other agencies of the Federal government and state and local agencies also have used the system.

1. The Origins and Development of the ARF System

The current ARF system has developed since 1971 from a simple, primarily county-based file with limited cross-sectional data, to a complex system which retains at its center a comprehensive longitudinal data set of county level information. The system was originally designed to assist in the location of the Area Health Education Centers program, as well as functioning as a general purpose area data file that could be used for analyzing the distribution of health resources (Bureau of Health Manpower, 1979). The file was originally developed within the Bureau's Division of Manpower Analysis. However, starting in 1975, an external contractor was selected to maintain and develop the system. By 1979, the system had been expanded to include a wide range of health and socioeconomic data (about 3,000 data elements), and the stated primary purpose (ibid) was:

> To serve as a major tool to provide the data for the Division (of Manpower Analysis) to carry out its mandated program of:

1. Research and analysis of the geographic distribution and maldistribution of health manpower and the identification and designation of health manpower shortage areas;

2. Analysis of health manpower supply, utilization, requirements, costs, and related issues; and

3. Dissemination of health manpower data and analyses

Furthermore, the stated objectives of the ARF at that time (ibid) were:

- To provide the most timely and accurate health and health related data available at the county level; and
- To provide a flexible analytical capability for analysis of health and health related data

By 1988, these departmental purposes and ARF objectives had been slightly modified. Thus in describing the purposes for which the ARF was used (Bureau of Health Professions, Office of Data Analysis and Management, 1988) the third purpose reads "Development and refinement of long-range forecasts of health professions supply and requirements" and two additional ARF objectives had been distinguished:

- To provide compatible and consistent health related data
- To provide data in a variety of forms in order to be of use to the broadest possible range of users

The development of the ARF system has been of fundamental importance to the community of health care planners and analysts. The Bureau's policy in developing the file from its origins into its current structure (see below) was farsighted. In the early 1970s, before the file was developed for use by the community, any attempt to use data such as are found in the ARF was fraught with difficulty. The data had to be identified by source, rules had to be developed

for their use when there were problems of missing data or different definitions of areas, and moreover, the same work was duplicated repeatedly in different settings for different purposes. The Bureau's recognition of the importance of a tool like the ARF led it to develop the ARF system through a deliberate and careful policy, and make it widely available. Moreover, the architects of the ARF system, though facing the same problems in building the file that individuals had faced on their own, had the opportunity to develop consistent policies for handling problems across data sources and across time, and documented these problems in the user documentation.

2. Structure, Content and Maintenance of the ARF System

The ARF system has evolved into an elaborate system of files and functions. In addition to the basic county file that is available for use on mainframe computers, there are the following components:

- A State/National Time series data base that includes data not available at the county level, in addition to data extracted from the county file and aggregated to the state level. The focus of this file is on education and the health professions. It is, however, "moribund" according to ODAM staff, and has not been updated for about seven years.

- A Graduate Medical Education file that contains over 1,000 data elements at the level of the educational program. This file has not been updated since 1987.

- Four Microcomputer data series extracted from the basic ARF county file: the Demographic series, the Health Facilities series, the Health Professions series (all containing data through 1985) and the update series (containing data since 1985 on the other three series). All of these data series are available on floppy diskettes.

- Detailed files containing hospital data and support files on various health professions or disciplines. These files, organized by county, contain data that are too detailed to maintain on the basic ARF county file. However, these are the support files to the ARF and new data are first augmented to these files before being added the 'basic ARF county file. These detailed

support files are available to individuals outside of ODAM only with special permission.

In addition to these files, the file documentation is available on paper or on magnetic tape. Furthermore, a series of hard copy area profile reports is available for each state and county, based on data contained in the ARF on request to ODAM. These profiles can be created for multi-county areas, but the program that runs them is not available to persons other than ODAM and ARF contractor staff. Normally, a complete set of single county profiles is printed out approximately every 18 months.

A user-friendly access system was developed by the ARF contractor in the 1980s to make it possible for users who do not know SAS or other programming languages to retrieve data from the ARF in usable formats such as tabulations or graphics. This interactive front-end to SAS allows users to indicate the data required, then the front end creates the SAS instructions needed to extract, process, and print the data in the requested format.

The basic ARF county file, as its name suggests, is organized largely with the logical record consisting of the county, which is identified through the Federal Information Processing Standards (FIPS) code. The exceptions to this county-based organization are as follows:

- Alaska appears as one state-wide logical record on the file because of inconsistencies across data sources in the level of disaggregation of data within the state.
- Since 1984, five of the independent cities of Virginia appear on the file as individual logical records and whenever data is available for the city level it appears in the city record. Otherwise, data for these cities is incorporated into the counties in which they are situated.

The county records also include codes that indicate other geographic entities within which the county is contained, such as states and metropolitan statistical areas (both primary and standard),

and geographic codes used by other entities, such as the American Medical Association county codes. The county structure is also followed in many of the support files and the four microcomputer data series (demographic, health facilities, health professions, and update series). Each microcomputer data series is organized so that there is (1) a diskette for all the counties in one state, (2) a diskette containing all the counties for contiguous states with few counties, or (3) several diskettes for all of the counties within one state, when there are a lot of counties in the state. The State/National Timeseries database is organized into two national files and one state level file, each organized by year of the data.

The principal types of data included in the basic county file are the various geographic identification codes and:

- Health Professions Descriptors: current and historical numbers of health professionals in different disciplines, characteristics of physicians, health manpower shortage area status and presence of National Health Service Corps (NHSC) sites and personnel
- Health Facilities: numbers and characteristics of hospitals in the area and data on nursing home beds and level of care
- Utilization: use of hospitals
- Expenditures: hospital expenditures, Medicare enrollment and reimbursements, physician prevailing charge indexes
- Ponulation Characteristics: numbers and distribution of the population by age, sex, and race
- Morbidity and Mortality: social morbidity indicators (housing and crime), and mortality, infant and overall
- Economic Characteristics: employment, unemployment, and income
- Environment: various measures of the environment

- Health Professions Training: numbers of schools, enrollments, graduates for major health professions and some allied health professions

## C. APPROACH TO THE EVALUATION

The methods followed to gather data and to evaluate the findings are described in this section. We collected data from four sources: users of the file, the health research literature, the ARF technical and user documentation, and the ARF system designers and their documentation. Having collected information from these sources, we evaluated whether the file met certain standards, and identified changes that would help meet specified standards. The potential changes were themselves evaluated in terms of their importance and cost, and a final set of recommendations have been made.

### 1. Interviews with Users

The principal approach followed for collecting information about ARF uses was to interview users about their experiences. In addition, to provide background for these user interviews, we also interviewed key HRSA staff to find out how the ARF is used and supported in different offices. This section describes how we developed an interview guide, how we selected our respondents and how we interviewed them.

Guide Development. At our initial meeting with the project officer we clarified project objectives. We then developed a workplan for the project that provided a scheduling framework. A matrix of issues to be addressed in the interview (and in subsequent project tasks) was developed and delivered to the project officer. Finally, the issues were developed into a series of questions following key topics: attitude, purposes, data elements used, experience with accuracy, timeliness, and access. An interview guide was then drafted. Using this draft, an

interview was conducted with an ARF user on MPR's staff after which the guide was revised and submitted to the project officer for approval.

Mode of Interviewing. We employed three approaches to collecting information from users: in-person meetings, telephone conversations, and a mailed list of questions, to which we requested a written response. We scheduled meetings with key HRSA staff and with some of the federal agency staff in Washington, D.C. These meetings followed a planned format, using an interview guide. Three different guides were used, one for agency heads or their designees, a second for persons anticipated to be ARF users, and a third for ODAM staff responsible for the ARF. The guides followed are shown in Appendix A.

For the telephone conversations, two different formats were followed. One was the same as that used for in-person conversations with ARF users. An abbreviated version of this format was followed when we were trying to track down users and uses of the ARF in agencies where there was some indication of prior ARF use and when we felt that some information could be collected but that a full interview was not warranted. A mail version of the interview guide was developed based on the telephone and in-person formats. This was sent to several respondents.

Identifying Respondents. Our original intention was to focus our interviews solely on ARF users, principally in the federal government, but also among external users, such as state government employees, federal contractors, and academics. These users would be identified from four sources: self-identification by BHPr staff, the records of ODAM in BHPr and their ARF contractor, the sales records of the National Technical Information Service (NTIS), and individuals identified by persons interviewed. At the request of BHPr's deputy director, this group of users was augmented to include four office heads whose agencies use the ARF. There was no attempt to select a random sample of interviewees. Rather, we relied on the guidance

of our project officer to establish which were the most important agencies to contact. In addition, we were constrained by the requirements of OMB to keep the number of interviews with persons who were not federal employees to nine or fewer.

We anticipated that a large group of users would be from within BHPr, and identified this group with the help of our project officer. HRSA office heads also identified members of their staffs with whom we should talk

The remaining interviewees were identified initially from the ODAM, ARF Contractor, and NTIS records. In many cases the individual whose name was listed as requesting the ARF or its documentation was not the ultimate user and that person then identified the user (if there was one).[2] In addition to the HRSA interviews identified above, we also talked to Public Health Service (PHS) staff in the National Center for Health Statistics (NCHS), the Centers for Disease Control (CDC), and the National Center for Health Services Research (NCHSR). Among other health agencies, we attempted interviews with staff from the Physician Payment Review Commission (PPRC), the Health Care Financing Administration (HCFA) and the Prospective Payment Commission (PROPac).

We identified other federal government staff ARF users in the Congressional Budget Office (CBO), the General Accounting Office (GAO), the Appalachian Regional Commission (ARC), The United States Department of Agriculture (USDA), the Veterans Administration (VA), the Office of Management and Budget (OMB) and the Federal Trade Commission (FTC).

We also identified users among federal contractors and grantees, health professional associations, universities and state government from the NTIS, ODAM, and ARF contractor lists.

---

[2]These sources were not always a good guide to the major users of the ARF. Major current users tend to have few questions, and therefore may not appear on the query list. Those who had acquired the file frequently did so on behalf of others who would actually use it.

Evaluating Interviews. Two types of information were collected in interviews: how the file had been used and for what purpose, and the experience of the user with the file's accessibility, timeliness, and accuracy for their purpose. We present these experiences in Chapter II. We also evaluated criticisms about the accessibility, timeliness, and accuracy of the data in the ARF in the context of several questions:

- Did the users have any misperceptions about the file's contents or availability that affected their use of the file?
- Were the origins of the misperceptions something that ODAM could respond to and rectify (such as providing clearer documentation or regular information about the ARF)?
- Were the users being realistic (for example, did the data they want exist)?

2. Literature Review

The contract called for a review of the appropriate literature for ARF citations to identify and evaluate the prevalence of studies utilizing the ARF and the nature of its use. We used several approaches for conducting a search for citations of ARF use:

- Formal search of computerized data bases of health literature
- Asking interviewees for citations of reports that they had prepared using ARF
- Search of BHPr publications for citations

Computerized literature searches can be conducted using databases of abstracts of published articles. They are accessible by keywords that describe the subject of the article. We used the string "Area Resource(s) File" in Dialog to search the Medline data base of health literature.

All publications in the data base from 1970 to the present were searched. Three citations were found.

The most fruitful approach was asking ARF users for citations, a list of which is provided in Appendix B. We also reviewed the Sixth Renort to the President and Congress on the Status of the Health Professions for references to the Area Resource File. Chapter II describes the uses of the ARF identified from the citations found.

3. Review of the Technical and User Documentation

We reviewed the ARF documentation to establish whether the documentation provided sufficient information. This documentation consists of technical information in a file layout that shows variable positions, variable names, a brief description of the variable, the source, and the date the variable was added or updated; and user documentation, which provides additional information about some of the variables included in the file.

Our criteria for evaluating the documentation was based on ease of locating a variable in the documentation; whether it could easily be determined how the variable was coded; and whether limitations of the data element (or source) were reasonably obvious to the user. Specifically, we looked at organization of the parts of the documentation, and within parts at variable order. We looked for consistency in conventions for naming and describing the variables in the technical documentation and for clarity in the explanations of variable content provided in the one line description in the technical documentation and in the longer explanations given in the user documentation. In addition, we looked for linkages and for consistency between the two parts of the documentation.

4. Documentation of the ARF System

The ARF system was also evaluated with respect to the organization and efficiency of the programs and procedures used to create the existing ARF. Our approach to this task was first to establish the facts through review of existing system documentation and discussions with the ARF contractor. The key components and processes of the ARF system were then documented in flow charts. At that stage, potential modifications to the existing system were identified that would result either in improvements in efficiency, or would bring the quality of the processing up to industry standards.

5. Evaluation of Findings and Identification of Changes

The findings from the user interviews, literature review, user documentation review, and system documentation provided a basis for identifying problem areas and for making a list of potential modifications to the ARF that would address these problems.

A memo was prepared that listed all the areas in which problems were believed to exist. These areas included:

- Under-utilization of the file
- The content or the structure of the data in the ARF
- The content or structure of the documentation available to users
- The (technical) accessibility of the ARF
- The cost of operating the current ARF system

Then, for each of these possible problem areas, the potential solutions were arrayed by type and with respect to which problem it would ameliorate. The types of solutions included:

- Outreach activities

- Immediate improvements to the documentation
- Long range improvements to the documentation
- Modifications to the data in the ARF
- Re-organization of data
- Modified delivery systems and access techniques

At the request of the project officer, we then went on to develop recommendations concerning these potential modifications without any input from BHPr.

6. Develonment of Recommendations

The final step was to develop recommendations for the future of the ARF, and to provide a context for BHPr to make choices among those recommendations. Our approach was to establish a set of criteria that could be applied to the potential modifications identified from a review of the Endings. The criteria applied to the potential modifications included the following:

- How critical was the need for this modification?
- How would the users be affected by the modification?
- Would additional user training be required?
- How would the modification affect the ARF programs and procedures?
- How much would it cost?
- Would additional hardware or software be required by the ARF contractor or the users?
- Would the modification be likely to solve the problem?
- Were there alternative approaches to meet the same objectives?

After applying these criteria the modifications were ranked by importance of the problem addressed, likelihood that the modification would solve the problem, and likelihood that substantial cost savings or only small cost increases would result. The most high ranked modifications (those viewed as either critical or important and likely to be feasible) were costed out in detail. The final step was then to make recommendations to BHPr about which of the high-ranked modifications should be introduced.

## D. OUTLINE OF THE REST OF THE REPORT

Chapter II of this report presents the findings from the interviews with users, the literature review, and the review of the user documentation. Chapter III presents the findings from the review of the ARF system. Chapter IV brings the findings together under a framework for evaluation: what are the problems, and what modifications could potentially solve them; and continue-s with an evaluation and ranking of these possible changes. Chapter V contains the evaluation of the most important potential modifications in terms of their costs and likely outcomes. Chapter VI lists those recommendations for the ARF system that we feel that BHPr should consider.

# Il. USERS AND USES OF THE AREA RESOURCE FILE

This chapter draws upon several sources of information to describe the range of users and uses of the Area Resource File. These sources are: ODAM and NTIS records, interviews of users of the Area Resource File, a review of literature for citations of ARF usage, and a review of the documentation provided to users of the file. While these sources do not provide a complete history of ARE usage, they are representative of the set of uses. These sources indicate widespread use of the ARF in the health research, policy and planning community, despite some surprising gaps in the user community. On the whole users were satisfied with the ARF though they offered suggestions which they felt would improve the ARF Data System.

The chapter is organized in five sections. Section A discusses who uses the ARF and who does not, and overall attitudes towards the ARF. Section B describes the purposes for which the ARF has been used and the most often used data elements, as well as the findings from a literature search MPR conducted. Section C discusses the accuracy and timeliness of ARE data, as perceived by users. Section D discusses the ease with which the ARF can be used, drawing on the interviewee responses and a formal review of the technical and user ARF documentation.

## A. USERS AND NON-USERS OF ARF

### 1. ARF Acquisitions

We attempted to discover the number of ARF users based on acquisitions of the ARF (standard magnetic tape version) either from ODAM or from the NTIS. NTIS sales records were available only for a six month period; eight copies of the ARF were sold in the six month

period prior to May 1989.[4] ODAM provided 26 copies of ARF during 1987, 19 copies in 1988, and 7 copies in the first five months of 1989. Assuming that NTIS annually sells 16 copies, the average number of acquisitions a year from NTIS and ODAM combined is between 35 and 42. Based on the 60 NTIS and ODAM sales records that were documented (for 1987, 1988, and part of 1989), acquisition was distributed across the following types of users:

- Federal government: 22 percent(NTIS 0, ODAM 13)
- State government: 8 percent(NTIS 1, ODAM 4)
- Academics: 38 percent(NTIS 4, ODAM 19)
- Contractors: 13 percent (NTIS 2, ODAM 6)
- Insurance: 2 percent (NTIS 1, ODAM 0)
- Professional Associations: 12 percent (NTIS 0, ODAM 7)
- Other: 5 percent (NTIS 0, ODAM 3)

We also requested from NTIS records of purchases of the ARF on floppy diskettes. Their records showed that for the six months preceding the request there were no purchases. ODAM, by contrast, has supplied the diskette version to several Public Health Service Regional offices and to at least one contractor. Nevertheless, the diskette version is rarely acquired relative to the standard ARF.

---

'During approximately the same six month period ODAM provided ARF to 16 users free of charge.

As we show below, acquisition of the ARF does not necessarily imply its use. Some federal government acquisitions are for contractors. Furthermore, acquisitions by one organization do not necessarily imply only one use or user of the ARF.

2. Users

For the user component of the ARF Evaluation, MPR interviewed individuals from the Federal government, the state government, contractors, professional associations, and academic institutions. These individuals were selected (as described in Chapter I) either because they appeared to be major users, or to represent a class of users, Table II.1 shows the number of persons interviewed by agency type and mode of interview and Table II.2 shows the federal agencies and outside institutions represented among respondents. All persons in HRSA whom we identified and approached were interviewed (see Figure II.1). Additional interviews within the Public Health Service included formal interviews with two persons, one affiliated with NCHS and one from CDC, as well as three informal interviews with individuals at the National Center for Health Services Research (NCHSR). Among HCFA, PROPac, and PPRC we succeeded in finding a current user only at PPRC.

Other federal agencies in which users were interviewed included CBO, GAO, VA, FTC, USDA, and the Appalachian Regional Commission (seven interviews). OMB, though a past user of the ARF, did not have any current users during the time-frame in which we were interviewing individuals.

TABLE II.1

DISTRIBUTION OF USER INTERVIEWS BY TYPE OF RESPONDENT AND MODE

| | Bureau of Health Professions' | Other PHS[b] | Other Federal[c] | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| In-person | 4 | 3 | 4 | | | | 11 |
| Telephone | | | | | | | |
| Full-length | 1 | 2 | 4 | 1 | 2 | 1 | 11 |
| Shortened | | 2 | 1 | 2 | | | 5 |
| Mail | | | 1 | 3 | 2 | 2 | B |
| Number of interviews: | 5 | 7 | 10 | 6 | 4 | 3 | 35 |

[a]The interviews with Bureau heads or designees and the interview with ODAM staff are not included in this table. In addition, the three individuals contacted in the BHPr's Division of Disadvantaged Analysis are shown as one interview in this table and are discussed as one interview throughout the text.

[b]The three individuals interviewed fran the Office of Data Management of HRSA's Bureau of Health Care Delivery and Assistance are shown as one interview in this table, and are discussed as one interview throughout the text.

[c]Contacts with HCFA, OMB, and PROPac are not included in this table.

TABLE II.2

AGENCIES AND INSTITUTIONS CONTACTED

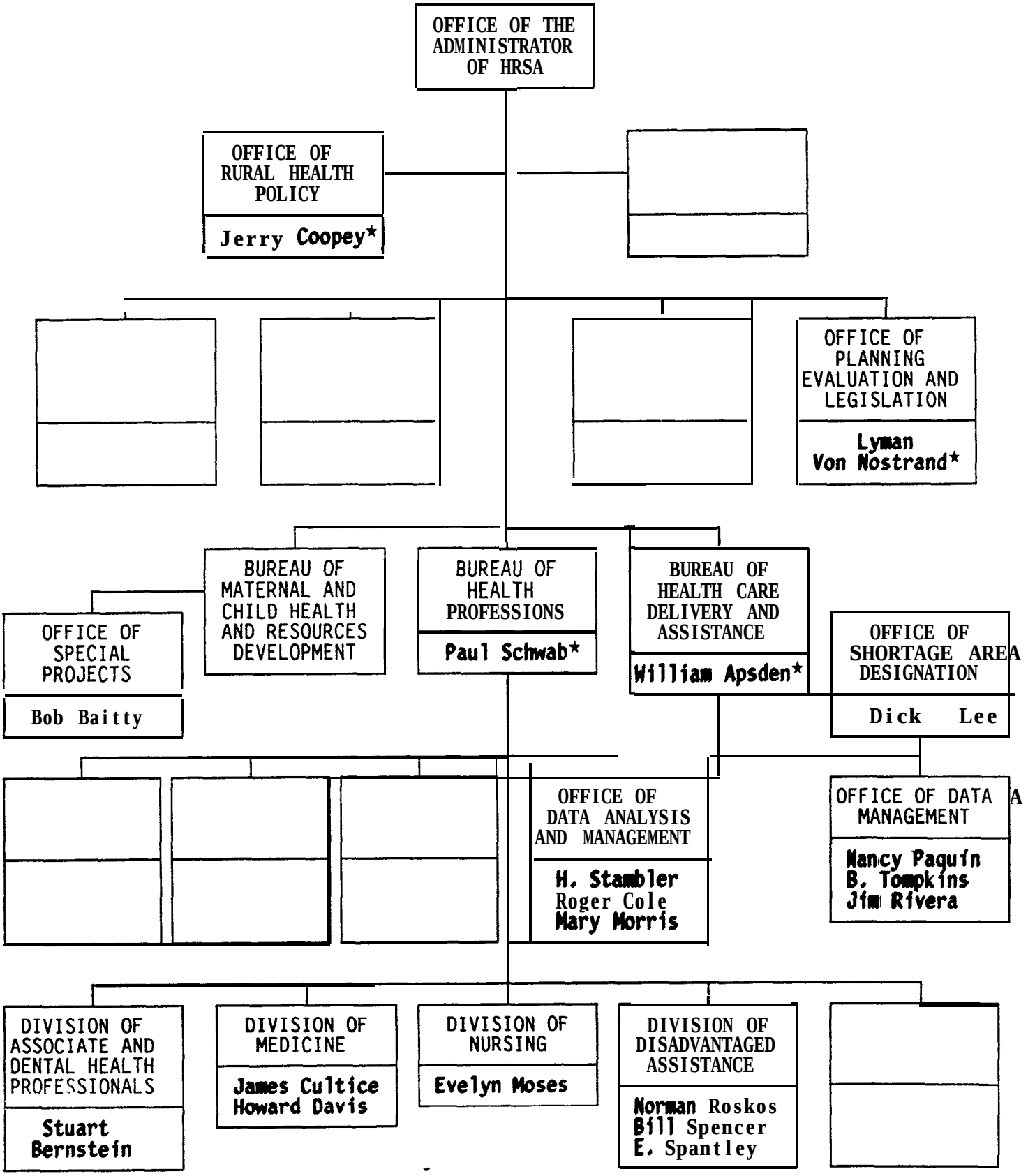
| Federal | | | | Professional | |
|---|---|---|---|---|---|
| DHHS | Other | State | Contractor | Association | Academic |
| Public Health Service: | CBO | Michigan[b] | HER | AMA | Johns Hopkins |
| HRSA | GAO | | Abt | ADA | U. Illinois |
| NCHS | ARC | | sss | AHA[a] | U. Michigan |
| CDC | VA | | SysteMetrics | Michigan Primary Care | |
| NCHSR | USDA | | | | |
| HCFA[a,b] | FTC | | | | |
| PROPac[a,b] | OMB[a] | | | | |
| PPRC | | | | | |

[a]No current users could be traced.

[b]Future use planned.

FIGURE II.1

INTERVIEWS WITH HRSA STAFF




* denotes agency head or designee of head

In addition, we interviewed or received mailed responses from staff of federal government contractors,[5] representatives of health professional associations,[6] academics,[7] and state government.'

3. Non-Users

Although we only approached persons who had acquired or asked questions about the ARF, some of these persons had not used the file. Reasons for not using the ARF are perhaps as important as reasons for using it. This section focuses on interviews in HRSA and federal agencies of individuals whom we had expected to be users, but who had never or scarcely ever used the file.[9]

HRSA Respondents in BHPr's Division of Disadvantaged Analysis had not used the ARF in their work largely because they need detailed race or ethnic data on health professionals. ARF collects little of this type of data. One of the respondents recommended that the ARF should be accessible through a user-friendly, menu-driven system; probably because he had never

(a) ARF is not primary for data. (b) this type of data is not available at county level.

[5]Health Economics Research, Abt Associates, Social and Scientific Systems, Rand, RTI and SysteMetrics.

[6]American Dental Association, American Medical Association, and the Michigan Primary Care Association.

[7]Johns Hopkins University, the University of Illinois, and the University of Michigan.

[8]Kent County Health Department in Michigan, which indicated that the ARF was a proposed data file for acquisition in their 1990 fiscal year budget.

[9]A potential contact at the National Institute of Drug Abuse had left two years prior to our call.

used the file he was unaware of the ACCESS system which has been designed for use by nontechnical staff.

Another respondent, in the Bureau of Maternal and Child Health and Resources Development, had not used the ARF basic county file, but had requested and received access to the hospital support file.[10]

Despite strenuous efforts and the appearance of numerous HCFA contacts on the ARF contractor's list of persons with queries about the file, we traced no HCFA staff who had used the ARF in recent years. Our initial list included three individuals, one of whom had apparently left HCFA since making a query about the ARF. Neither of the two remaining individuals were ARF users although one had acquired the file for another HCFA staff member who had in turn made the file available to HCFA data center users." These two individuals suggested other contacts at HCFA. We spoke to a total of 13 HCFA staff with the following results:

- Two had used the file, one 10 years ago and another 8 to 9 years ago
- Three had several contractors who were using the file, often at their recommendation

Why was it here.

[10]Gaining access to one of the ARF's support files can take a long time, as indicated by this respondent. However, in accordance with the contract which has been negotiated between the American Hospital Association (AHA), who is the distributor of the facility-level data, and BHPr's ODAM, ODAM must seek formal permission before releasing the facility-level data outside of BHPr.

"The other individual on our list from HCFA had not used the file. At one time, some of the HCFA data, specifically county level Medicare reimbursements, came through his office and was supplied on an annual basis for inclusion on the ARF. This data 'now goes through a different office. According to the ODAM contact listing provided to MPR, this individual called in regards to replacing Medicare reimbursement data since HCFA's originals had been destroyed.

- Two (from the Office of Research and the Office of Peer Review) had considered using the file at one time but had not actually used it
- One was planning to use the file in the coming months for an analysis of variations in hospital utilization

Based on the numerous contacts we had with HCFA, it appears that HCFA generally procures the ARF for use by its contractors, but does not often use the file internally. This practice reflects the role of external contractors in the HCFA research program.

The person at OMB we identified for interview had been a summer intern who had used the ARF during her internship at OMB. The staff member for whom she had worked for had also left OMB. We contacted three current staff members who work on health issues in an attempt to identify current ARF users. Two staff members were familiar with the file but had not used it recently, nor were they aware of anyone else who was using it then or had recently used it.

PROPac. The person identified for interview had left PROPac by the time of our call. We spoke with three staff members, including the executive director, none of whom is using or has recently used the file. The commission had acquired the file recently for future use.[12]

[12] At the time of the interview with PROPac, MPR spoke with a Dr. Donald A. Young, who is the Executive Director, and Ms. Debra Williams, who is a programmer. They had received the ARF data and documentation, but had not yet had the opportunity to examine either item. They plan to merge socioeconomic and Medicare admissions and reimbursement variables from the ARF with data on closed and open hospitals. The ARF variables will be used as proxies, such as Medicare admissions as a proxy for access to care in counties where hospitals have closed.

4. Response Rates

We set out to interview 16 individuals in person and 14 individuals by telephone. We succeeded in covering the full set of issues in-person with 15 individuals and by telephone with 11 individuals. In addition, we had telephone conversations yielding more limited information with a further 20 individuals. We received eight fully or partially completed responses to our mailing to 17 individuals representing federal, state, contractor, academic, and other users. We followed up all non-respondents by telephone. Some did not wish to answer the questions, because their experience was so limited that they did not feel they could be helpful. Five of the respondents were no longer with the organization. The remainder stated they did not have time to respond or said that they intended to respond but then did not do so.

We collected responses from users in 35 interviews from all sources, and had helpful conversations with many other individuals. Table II.1 shows only the completed responses, and the tabulations of user responses in subsequent sections include only the complete responses. We augment these responses with narrative information from our informal conversations with additional users and persons who had not yet used the file but planned to do so.

It should be reiterated that MPR did not interview every individual who has used the ARF nor who is currently using the ARF, but rather interviewed a sample of these individuals. In addition, we were further constrained by the OMB restriction of conducting formal interviews with only nine individuals not associated with the federal government.

5. Attitudes Towards and Use of the ARF in HRSA

We spoke with the deputy directors, or deputy designees of the director, for four HRSA offices: BHPr, the Office of Planning, Evaluation and Legislation (OPEL),' the Office of Rural

Health Policy (ORHP), and the Bureau of Health Care Delivery and Assistance (BHCDA).[13] Each of these offices has a distinct mandate, different data needs, and different approaches to meeting these needs. We discussed how these offices met their data needs and the ARF's role (and ODAM's) in meeting these needs.

Because of drastic cuts in BHCDA's programs, sophisticated analyses are not currently of great importance to this bureau BHCDA has its own equivalent to ODAM, the Office of Data Management (ODM), which provides analysis support to that bureau. Hence, BHCDA does not often require ODAM support although it does use the ARF regularly. ODM coordinates with ODAM staff in the collection of new data for the ARF.

OPEL's mandate is to prepare policy and options papers on future directions HRSA programs might take. To fulfill this mandate some analyses are needed, for which ODAM frequently supplies help. The ORHP is a new office of HRSA, primarily an advocate for rural health issues, for which data are sometimes needed. ODAM is one source of such data. Neither OPEL nor ORHP staff are very familiar with the data sources ODAM uses to meet their data needs, and thus have limited knowledge of the extent to which ODAM uses the ARF or its support files in supplying requested data.

We may characterize OPEL as a frequent user of ODAM help, ORHP as an occasional user of ODAM help, and BHCDA as a rare user of ODAM support. The two offices which use ODAM as a source of data to answer policy questions were impressed by the responsiveness

[13] In addition, we talked to the director of BHCDA's Office of Data Management (ODM) and the Director of BHPr's ODAM, whose comments are also incorporated in this section.

(especially timely response) of ODAM to their requests. ORHP also mentioned the offer by ODAM staff to meet and discuss assistance that ODAM could provide.

The bureau chiefs (and proxies) were not familiar with ARF use, but some of them commented on problems that staff in their respective bureaus had mentioned to them. Two respondents had concerns pertaining to the mortality data. We believe that the origins of these concerns are (1) the absence of 1981 data for infant deaths by sex and race for infants under one day old, one to six days old, seven to twenty-seven days old, and twenty-seven to three hundred and sixty-four days old; and (2) the lack of annual infant mortality rates on the file, although these can be calculated. In addition, small area (sub-county) analyses cannot be conducted using the ARF. This comment was also made frequently by users (see below). The respondents recognized that lack of data at the sub-county level makes such a file hard to create.[14]

All three of the Bureau chiefs outside of BHPr suggested that their offices might make more constructive use of the ARF if they were better informed about ARF contents, when the file had been updated, and what new data had been added. (This was also a noted by many HRSA users).

At the Bureau level, the major limitation of ARF use within OPEL, ORHP, and BHPr seems to be a lack of staff with analytical and programming skills, in contrast to a lack of knowledge of the file or doubt as to its accuracy. BHPr/ODAM and BHCDA/ODM, where analytical and technical programming skills are available, make major use of ARF.

---

"OPEL mentioned that ARF potentially may be used in small area analyses planned by that office and is currently being investigated for use in a contract to Lewin and Associates.

The importance of ARF as a resource was recognized by all HRSA respondents. None would like to see an interruption in ARF maintenance because of the high start-up costs resulting from an interruption. However, the cost of the ARF contract was of general concern. The ARF contract tends to absorb between 10 and 15 percent of the $2 million set aside for Section 708 activities. Of this allotted amount, some is for analytical studies, but the majority is for file maintenance and development. Suggestions for reducing or offsetting the ARF maintenance costs included:

- Requiring outside users to pay for the file
- Bringing the ARF maintenance in-house
- Finding a contractor who could perform the current work at the same level, but at a lower cost

Despite these limitations, the ARF was regarded as a generally satisfactory tool by both OPEL and ORHP, and BHCDA considers it satisfactory for "macro" strategic planning with evident limitations for program targeting to small areas.

6. Overall Attitudes

Inevitably, when users are asked to evaluate a large data file almost all of them will have areas in which they would like to see changes made. However, the desire for changes in certain aspects of the file does not necessarily mean that users are generally dissatisfied. Thus, in our evaluation of the ARF, we wished to distinguish between users' overall attitudes and perceptions and their detailed concerns about elements of the file.

To this end, each interview in our user survey began with two broad attitudinal questions addressing how users felt about the ARF; specifically, did they have strong feelings about the adequacy of the file, and were these feelings positive or negative? In addition, each interview concluded with another general question about users' overall experience in working with the ARF. Responses to these questions are shown in Table II.3.[15]

As can be seen, all the individuals interviewed felt positively about the ARF to some degree, and characterized their experience with the file as satisfactory or very good (though one individual also indicated frustration). All six of the respondents who expressed strong feelings about the ARF were somewhat positive about the file.

While the overall user appraisal of the ARF is undoubtedly positive, some broad, general concerns were also expressed before the in-person and telephone interviews formally started. Users offered the following suggestions to enhance the accessibility of the file:

- Some data elements on the file are regarded as erroneous, incomplete, or misleading
- Users would like component data to be more current
- An improved distribution system for information about the ARF is needed.

These general concerns and the reasons for them are discussed in detail in subsequent sections of Chapter II, where we also to try to clarify perceptions and facts about the file.

[15]The interview guide which was used for interviews conducted in-person and by phone worded the questions at the beginning of the interview in a slightly different manner than the Interview Guide mailed to respondents. Because of these differences, the two groupings of respondents (in-person/by phone versus mail) are presented separately in Table II.3.

TABLE II.3

PERCEPTIONS OF ARF

| | Bureau of Health Professions | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| IN-PERSON INTERVIEWS | | | | | | | |
| Strong feelings regarding the adequacy of ARF? | | | | | | | |
| Yes | | 3 | 2 | 1 | | | 6 |
| No | 2 | 1 | 5 | | 2 | 1 | 11 |
| Both | 2 | | 1 | | | | 3 |
| Generally positive or negative about ARF? | | | | | | | |
| Positive | 2 | 4 | 8 | 1 | 2 | 1 | 18 |
| Negative | | | | | | | 0 |
| Both | 2 | | | | | | 2 |
| How would you characterize your experience with ARF? | | | | | | | |
| Very good | 1 | 4[b] | 7[c] | 1[d] | 1 | 1 | 15 |
| Satisfactory | 3[a] | | 1 | 1 | 1 | | 6 |
| Frustrating | 1 | | | | | | 1 |
| Unacceptable | | | | | | | 0 |
| Number of interviews | 4 | 4 | 8 | 1 | 2 | 1 | 20 |
| NAILED INTERVIEWS | | | | | | | |
| Strong feelings (either positive or negative) regarding the adequacy of ARF? | | | | | | | |
| Positive | | | 1 | 2 | 1 | 1 | 5 |
| Negative | | | | | | | |
| No opinion | | | | 1 | 1 | 1 | 3 |
| Row would you characterize your experience with ARF? | | | | | | | |
| Very satisfactory | | | | 1 | | | 1 |
| Satisfactory | | | | 2 | 1 | 2 | 5 |
| Unsatisfactory | | | | | | | |
| Very unsatisfactory | | | | | | | |
| Did not respond | | | 1 | | 1 | | 2 |
| Number of interviews | | | 1[e] | 3 | 2 | 2 | 8 |

[a]One user characterized experience with the ARF as both satisfactory and frustrating.

[b]One of these responses is between very good and satisfactory.

[c]Two of these responses are between very good and satisfactory.

[d]One interview is represented twice for this question. with the experiences of the researcher and programmer recorded separately.

[e]The interviewee responding by mail only answered these two questions of the Interview Guide.

7. Summary

ARF users are found in the Public Health service, other federal government agencies, state government agencies, and outside agencies. Each year, about 40 copies of the ARF file are distributed to users outside the Public Health Service, of which nearly one third is to federal and state agencies and about half is to academics and contractors. Among the federal health agencies that we spoke with (all of which had received the file) relatively little use has been made internally (only one current user was found among HCFA, PROPac, and PPRC). In HRSA, most direct use of the file is made in BHPr and BHCDA, especially in ODAM and ODM, the respective data management offices, where technical programming skills are more common.

Before asking any detailed questions during the interviews, users had positive attitudes to the ARF and to ODAM (the office which supports it), but were concerned that ARF should be maintained in the future. Nevertheless, HRSA staff mentioned concerns about the high cost of the ARF contract, and questioned whether it could be reduced. Some users also commented that the distribution of the ARF and information about it could be improved and one or two mentioned data problems.

## B. USES

Section 1 is a discussion of the overall uses of the ARF (based on the interviews), and Section 2 continues with a discussion of uses by HRSA staff. Section 3 describes publications that have used ARF data, followed by a discussion of the findings of the literature review conducted for formal citations of ARF usage. In Section 5 we describe how contractors use the

file. We describe in Section 6 the most frequently used data elements, and in Section 7 the data elements that users would like to see included in future versions of the ARF.

1. Overall Uses

In our interviews with ARF users, we identified a wide range of health services policy and research issues for which the file is being used (shown in Table II.4). (The table shows that a single study using the ARF could have multiple purposes, all of which are included in the table.) The three purposes for which the ARF was used most frequently are: demographic analyses; estimates of health professional demand or needs; and estimates of health professional supply. Examples of the types of evaluations being conducted include:

- Examining rural areas' health status, health care use, and shortage of physicians
- Evaluations of the impact of Medicare and Medicaid policies
- Rural hospital closures and the effect of this on access to care
- The openings, closures, and mergers of short term acute hospitals.

Only one of our interviewees had used the ARF for educational or training issues, and use of the file to address service and program planning issues was also mentioned infrequently. Although education and training of health professionals are major policy issues for BHPr, the

TABLE II.4

ISSUES IN WHICH THE ARF HAS BEEN USED

| | Bureau of Health Professions' | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| Service planning | | | | 1 | 1 | | 2 |
| Demographic analyses | | 2 | 2 | 2 | 2 | | 8 |
| Estimating health professional demand or needs | | 2 | 2 | 2 | 3[b] | | 9 |
| Estimating health professional supply | 1 | 1 | 1 | 3 | 3[b] | 1 | 10 |
| Educational or training issues | | 1 | 1[b] | | | | 2 |
| Population health care requirements | | 3 | 2[b] | 1 | 1 | | 7 |
| Shortage area designation | | 2 | 1 | | 1 | | 4 |
| Program planning | | | 2 | 1 | | | 3 |
| Program evaluation | | 1 | 3 | 1 | | 1 | 6 |
| Respond to Congressional inquiries | | | 5 | | | | 5 |
| Other | 3 | 3 | 2 | 1 | I | 2 | 12 |
| Number of interviews: | 3 | 4 | 8 | 4 | 4 | 3 | 26 |

[a]One interview is not included in the table because the issue cited pertains to what the ARF could be used for.

[b]Interviewees stated that several issues might be examined in the future. This occurred once for estimating health professional demand or needs, once for estimating health professional supply, once for educational or training issues and twice for population health care requirements.

limited amount of good data on these topics appears to have contributed to the lack of use of this portion of the file.[16]

We were surprised that the federal government employees were not using the file for more program planning and placement purposes. There appear to be three reasons contributing to this lack of use (aside from the possibility that those who have used the file for these purposes were not interviewed). One reason is that programs were scaled back heavily in the 1980s and hence there is less planning and placement activity. Another reason is that the data on the ARF may not be appropriate for planning.[17] Finally, for those conducting plan&g activities at the sub-county level (for example, medically underserved area), the ARF's county-based organization is not the most useful.

2. Uses bv HRSA Staff

The ARF data system is maintained by ODAM, which is part of BHPr in HRSA. The ARF would seem to be a readily available data source for BHPr personnel, as well as for HRSA personnel in general, for use in conducting their analytical research. MPR conducted five interviews, including seven individuals, within BHPr and three interviews, including five

---

[16]The health professions training portion of the ARF (based on the August 1989 documentation) concentrates on the educational background of individuals, with variables on the number of schools, total enrollment, enrollment by year (i.e. first year, second year), and total graduates. Enrollment by year and sex are available for schools of medicine, dentistry, pharmacy, and optometry, and weighted average tuition amounts are available for schools of medicine and dentistry. There are thirteen training variables, such as anesthesiology and radiology, on the ARF, where each type of training appears for both short-term general hospitals as well as short-term non-general and long-term hospitals. The training data are currently available only for 1985.

"Although the ARF contains the most recent data available for most subject areas, the "oldness" of some of the data renders the use of it for planning purposes inappropriate (such as information on the total number of active dentists and the number of registered nurses).

individuals, from various HRSA divisions in order to assess for what purposes these individuals were utilizing the ARF.[18]

There were three ways in which the ARF was used within HRSA:

- As an accessible tool in answering research questions posed either from within HRSA or from the research community outside of HRSA (3 of 5 interviews)[19]
- To augment an already existing data set, such as the BHCDA net system and the nursing survey data collected by the Division of Nursing (3 of 5 interviews)
- As the primary data source in conducting an analysis (1 of 5 interviews).

One individual from the Division of Medicine in BHPr has not utilized the ARF in a number of years, but has a project in mind where the ARF would play an integral role. This individual would like to create a real time system which could be employed in answering questions received via telephone concerning a particular county and its available health resources. At present, the individual is receiving calls from a mix of consulting companies, hospitals, and local and state planning agencies and has to refer the callers to ODAM. Rather than route the calls to ODAM,

---

[18]Of the seven individuals in the BHPr, two were from the Division of Medicine, one was from the Division of Nursing, three were from the Division of Disadvantaged Analysis (discussed as one interview throughout the text), and one was from the Education and Resources Section of the Division of Dentistry and Allied Health Professions. Of the five individuals within HRSA: one was from the Office of Special Projects in the Bureau of Maternal and Child Health and Resources Development; one was from the Office of Shortage Area Designation, and three were from the Office of Data Management (discussed as one interview throughout the text), both offices within the Bureau of Health Care Delivery and Assistance. The three individuals from the Division of Disadvantaged Analysis and the individual from the Office of Special Projects had not actually used the file although they had been identified as users. Their responses were discussed in an earlier section of this report.

[19]One interview is not included in these counts because the ARF file had not been used recently.

an extract of the ARF would be created and placed onto a Personal Computer. In placing the extract on the PC, the individual could quickly and easily relay information about a county to the caller; the only information needed from the caller would be either the FIPS county code or the name of the county of interest.[20]

3. Publications

Analyses utilizing the ARF have been published in a wide range of reports and articles. As shown in Table II.5, Federal government interviewees indicated that ARF-based analyses were published most frequently in reports to Congress, and next most often in internal agency reports. The funds for the maintenance of the ARF data system are derived from Section 708 of the Public Health Service Act. Four of the six respondents within HRSA indicated having worked on a project which was funded under Section 708 and which required the submission of a report to Congress. Three of these four individuals utilized the ARF in completing their work. A different picture was seen outside the federal government: for nongovernmental interviewees, the most frequently cited publication vehicle was professional journals.[21]

---

[20]The county profile reports, which are available in hard-copy only, would seem to be an alternative data source for this type of project. However, they are very voluminous, and would have to be reproduced for the particular counties of interest to the caller. If the profile reports could be produced on demand by nontechnical users, this proposed development of an entirely new subsystem of the ARF might not be needed.

[21]A list of references in which the ARF was used to complete analyses can be found in Appendix C. A number of respondents have forthcoming articles, for which references could not be obtained.

TABLE II.5

REPORTS WHERE ARF ANALYSES INCLUDED

| | Bureau of Health Professions' | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| BHPr/OTHER PHS/OTHER FEDERAL | | | | | | | |
| Reports to Congress | 2 | 1 | 6 | | | | 9 |
| Reports to your agency | | 3 | 4 | | | | 7 |
| Other reports | 2 | 2 | 1 | | | | 5 |
| Can you provide a list of reports, with authors? | | | | | | | |
| Yes | | 1 | 3 | | | | 4 |
| No | 3 | 3 | 5 | | | | 11 |
| CONTRACTORS/PROFESSIONAL ASSOCIATIONS/ACADEMICS | | | | | | | |
| Reports to Congress | | | | 1 | | | 1 |
| Reports to your client | | | | 2 | | 1 | 3 |
| Reports prepared for your organization | | | | | 3 | | 3 |
| Articles in published journal? | | | | 2 | 2 | 2 | 6 |
| Other | | | | 3 | 1 | 1 | 5 |
| Can you provide a list of reports, with authors?[b] | | | | | | | |
| Yes | | | | 2 | 2 | 1 | 5 |
| No | | | | 2 | 2 | 2[c] | 6 |
| Number of interviews: | 3 | 4 | 8 | 4 | 4 | 3 | 26 |

"One interviewee is not represented in the table, due to the absence of any published reports.

[b]A listing of references in which the ARF was used can be found in Appendix C.

[c]The version of the Interview Guide mailed to respondents did not ask for references.

4. Findings from the Literature Review

A literature review was conducted to determine the extensiveness of ARF use and the specific analyses conducted. MPR's study identifying literature that cited ARF use was outlinedin Chapter I of this report. The findings of the review of this literature are presented here. MPR examined two of the three publications found in the automated literature search,[22] as well as ten of the fifty-nine citations given by users. The twelve articles examined were published in the following journals:[23] American Journal of Public Health, Health Care Financing Review, Health Services Research, Inquiry, Journal of Chronic Diseases, Medical Care, and Vital and Health Statistics. Each of these journals is well known in the health care arena. In addition, Bulletin 915 and Special Report 134 from the Arkansas Agricultural Experimental Station, in cooperation with Economic Research Service of the United States Department of Agriculture, were reviewed.

Each of the twelve publications dealt with some aspect of health care, either from an operational or a policy perspective. The ARF was mentioned as a source in eight of twelve articles. Variables used included patient care nurse anesthetists, per capita income, hospital beds per capita, number of nursing home beds, physicians per capita, metropolitan/non-metropolitan indicators, and Health Service Areas codes. ARF variables were also used as proxies, such as the number of emergency rooms relative to the county's land area as a proxy for emergency

[22]One article was a description of the Area Resource File written by Howard Stambler in 1988, and was thus not reviewed.

[23]Extracts of the twelve articles we examined, including the journal name, the title of the article, the main research question addressed, the primary data source used, and the secondary data sources used, are presented in Appendix D. Six of the fifty-nine citations are papers given at conferences or association meetings; two are unpublished or of unknown origin; and four are forthcoming.

which was used to complement or expand upon the primary data set. The authors of one article stated that the variables from the ARF "are critical in the analysis because they represent major influences on the demand for and supply of physicians' services and thus on local medical prices" (Muller and Otelsberg, Medical Care, 1979).

Ten of the twelve had been selected for review because their authors told us they had used the ARF in preparing the article. Nevertheless, four of these ten articles did not cite or reference the ARF. However, all of these articles cited data sources which are included on the ARF. Some examples of these data sources are: Bureau of the Census for the Censuses of Population (1970 and 1980) and the Census of Agriculture (1974), American Hospital Association data, and American Medical Association data. It is likely that the authors cited the source data found on the ARF in place of the ARF itself. It is thus not clear how the ARF was used.[24]

The difficulty of identifying how the ARF is used was not restricted to articles. In addition to the 12 articles selected for review, MPR examined BHPr's Sixth Report to the President and Congress on the Status of Health Personnel in the United States. The Report presents information on the present and future supply of health personnel in the United States. Each

[24] It may be instructive to indicate the potential for hidden uses of the ARF by describing the uses MPR has made of the copy of the ARF provided for the evaluation. Four federal government contracts have used or will shortly use the file. These are as follows:

| Agency | Contract | Use |
|---|---|---|
| BHPr | ARF Evaluation | Review of Data |
| HCFA | Rural Transition Grant Evaluation | Description of Areas in which grantees are located (including health facilities and professionals data) |
| FNS | WIC/Medicaid Evaluation | Description of areas with WIC enrollees (including health professional data) |
| FNS | Quick Response Project | Used area economic characteristics to model placements after employment training |

chapter discusses a specific health profession. Although BHPr staff had cited this publication as having drawn on the ARF for the data presented, we could find no direct references to the ARF. However, several publications written by ODAM and Applied Management Sciences were cited in which presumably ARF was used. References to data sources were made throughout each chapter, and usually pertained to one of four sources: the division itself, a government agency (such as the Census Bureau), an appropriate association, and unpublished data sources. In addition, several of the divisions created estimates based on data that they had compiled. Presumably ODAM and the ARF had a role in creating the estimates. In fact, ODAM has stated that they had direct involvement in creating the nursing supply forecasts by state through the year 2020 for the Division of Nursing.

### 5. Contractors Who Have Used the ARF

In addition to investigating direct use of the ARF during our interviews, we wished to determine whether the federal government respondents had contractors who had used the file within the last five years. Only four of sixteen respondents had contractors who used the ARF. At HCFA, where none of the staff were currently using the file, three people we spoke with had contractors who were using the file (see Table II.6).

TABLE II.6

ARF EXPERIENCE OF CONTRACTORS TO BHPr, OTHER PUBLIC HEALTH SERVICED AGENCIES, AND OTHER FEDERAL AGENCIES

| | Bureau of Health Professions | Other PHS | Other Federal | Total |
|---|---|---|---|---|
| Have your contractors or grantees used ARF in the last five years? | | | | |
| Yes: Contractors have used ARF | | | 2 | 2 |
| No: Contractors have not used ARF/no contractors | 2 | 4 | 6 | 12 |
| Unsure | 2 | | | 2 |
| Types of projects working on: | | | | |
| Service planning | | | | |
| Demographic analyses | | | 1 | 1 |
| Estimating health professional demand or needs | | | 1 | 1 |
| Estimating health professional supply | | | 1 | 1 |
| Educational or training issues | 2 | | | 2 |
| Population health care requirements | | | | |
| Shortage area designation | | | 1 | 1 |
| Program planning | | | 1 | 1 |
| Program evaluation | | | 1 | 1 |
| To respond to congressional queries | | | | |
| Other | | | 1 | 1 |
| Reports for: | | | | |
| Congress | | | | |
| HRSA administrator | | | | |
| Your Agency | | | 2 | 2 |
| Others | | | 1 | 1 |
| Unsure | 2 | | | |
| Number of interviews: | 4 | 4 | 8 | 16 |

We also discovered whether the respondent knew of other users of the file, with the results shown in Table II.7. About half of the respondents were aware of other users of the ARF. Two individuals emphasized that they are promoting the use of the file within their respective organizations.

6. Data Elements Used

One goal of the ARF interview was to ascertain which types of variables are used. The ARF is an extremely large data set, which could be significantly reduced in size if rarely or never used variables were deleted. Table II.8 presents a summary of the types of data elements respondents have used. While this is obviously not a comprehensive listing of all variables found on the ARF nor of all variables used, it covers fields from each of the major groupings: Health Professions, Health Facilities, Utilization, Expenditures, Population, Environment, and Health Professions Training.

Table II.8 shows that several variables appear to be used by a large number of the respondents. These variables include:

- The number of physicians (23 of 27 respondents)
- The characteristics of physicians (12 of 27 respondents)
- The number of hospitals in an area (20 of 27 respondents)
- The characteristics of hospitals in an area (16 of 27 respondents)
- Utilization of hospitals (12 of 27 respondents)
- The number and distribution of the population (20 of 27 respondents) by: age (19 of 27 respondents), sex (15 of 27 respondents), and ethnic background (13 of 27 respondents)
- Infant mortality rates (11 of 27 respondents)

TABLE II.7

OTHER USERS MENTIONED

| | Bureau of Health Professions | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| Are you aware of any other users?[a] | | | | | | | |
| Yes | 2 | 2 | 4 | 1 | 1 | 1 | 11 |
| No | 2 | 2 | 4 | | 1 | | 9 |
| Number of interviews: | 4 | 4 | 8 | 1 | 2 | 1 | 20 |

[a]The mailed version of the Interview Guide did not ask this question.

TABLE II.8

DATA ELEMENTS USED

| | uureau of Health Professions | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| **DATA ELEMENTS USED** | | | | | | | |
| HEALTH PROFESSIONS DESCRIPTORS | | | | | | | |
| Number of Health Professions | 4 | 4 | 7 | 4 | 3 | 3 | 25 |
| Physicians | 4 | 4 | 7 | 3 | 3 | 2 | 23 |
| Nurses | 1 | | 5[a] | 2 | 1 | | 9 |
| Other | | | | 1 | | 1 | 2 |
| Characteristics of physicians | | 1 | 4 | 3 | 1 | 3 | 12 |
| Health manpower shortage areas | | | 4 | 2 | | 1 | 7 |
| HEALTH FACILITIES | | | | | | | |
| Number of hospital in an area | 1 | 2 | 7 | 4 | 3 | 3 | 20 |
| Characteristics of hospitals in an area | 1 | | 6 | 4 | 2 | 3 | 16 |
| Number of nursing home beds | 2 | | 3 | 1 | 1 | 2 | 9 |
| Indicators of nursing home level of care | | | 2 | | 1 | 1 | 4 |
| **UTILIZATION** | | | | | | | |
| Use of hospitals | | 1 | 5 | 3 | 1 | 2 | 12 |
| EXPENDITURES | | | | | | | |
| Hospital Expenditures | | | 2 | 1 | 2 | 3 | 8 |
| Medicare Reimbursements | | | 3 | 2 | | 2 | 7 |
| Physician Prevailing charges | | | 2 | 1 | 1 | 3 | 7 |
| POPULATION CHARACTERISTICS | | | | | | | |
| Number and distribution of the population by: | | 2 | 6 | 3 | 2 | 3 | 20 |
| Age | | 2 | 6 | 3 | 2 | 2 | 19 |
| Sex | | 2 | 3 | 2 | 2 | 2 | 15 |
| Ethnic background | | 1 | 3 | 3 | 1 | 2 | 13 |
| MORBIDITY AND MORTALITY | | | | | | | |
| Infant mortality rates | | 3 | 4[a] | 2 | | 2 | 11 |
| Overall mortality rates | | 2 | 1[a] | 1 | 1 | 1 | 7 |
| ENVIRONRENTAL VARIABLES | | | | | | | |
| Housing/crime numbers as indicators of social morbidity | | | | 1 | | | 2 |
| Climate or other environmental variables | | 1 | 2 | 2 | | 1 | 6 |
| Acres of land used to estimate population density | | | 3 | 3 | 2 | 2 | 10 |

TABLE II.8 (continued)

| | Bureau of Health Professions | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| ECONOMIC CHARACTERISTICS | | | | | | | |
| Employment numbers | 2 | | 4 | 3 | 3 | 3 | 14 |
| Unemployment numbers | 1 | | | | | 3 | 15 |
| Income levels | 2 | 2 | 7 | 4 | 2 | 3 | 20 |
| Poverty status | 1 | 2 | 5 | 4 | 1 | 3 | 16 |
| HEALTH PROFESSIONALS TRAINING | | | | | | | |
| Numbers of schools for specific fields | | | 1 | 2 | | 1 | 4 |
| Enrollments for specific fields | | | | | | 1 | 1 |
| Graduates in specific fields | 1 | | | | 1 | 1 | 3 |
| GEOGRAPHIC MATCHING CODES PROVIDED OR FILE | | | | | | | |
| FIPS codes | 4 | 3 | 7 | 4 | 3 | 2 | 23 |
| MSA | | | 2 | 3 | | 2 | 7 |
| Other county codes | | 2 | 1 | 1 | | 2 | 3 |
| Contiguous county indicator | | 1 | 2 | 2 | 1 | 2 | 7 |
| Other | | | 3 | | | 1 | 7 |
| OTHER | 1 | 1 | 2 | 1 | | 1 | 2 |
| Number of Interviews: | 4 | 4 | 8 | 4 | 4 | 3 | 27 |

[a]One person expressed interest In using this variable in the future.

- Employment rates (14 of 27 respondents)
- Unemployment rates (15 of 27 respondents) *Income levels (20 of 27 respondents)
- Poverty rate (16 of 27 respondents)
- FIPS codes (23 of 27 respondents).

Respondents indicated two basic ways in which information from the ARF is treated:

- ARF data is used to examine issues of interest, e.g. the relationships among supply of primary care physicians, poverty rates, and infant mortality; variations in hospital use by Medicare beneficiaries across geographic regions. These essentially descriptive and cross-sectional analyses make use of cross tabulations prepared straight from the file.
- Selected ARF variables are added to an independent data set to permit analysis of the effect of geographic variations in the health care environment on the primary variables of interest (e.g. national survey data on insurance coverage patterns of low income individuals could be augmented by data on health environment characteristics of respondents' counties of residence to permit examination of the effects of health care markets on the demand for insurance). Both cross-sectional and time-series analyses occur.

One can see in Table II.9 that twenty-four of twenty-seven the respondents merge data from the ARF with another data set.

7. Data Elements Desired

In addition to finding out which data elements were most frequently being used from the ARF, we asked respondents whether the ARF contained the data elements they needed, whether they had ever supplemented the ARF with other data sources, or had ever decided against using the file for some purposes. Table II.10 presents users' replies.

TABLE II.9

HOW INFORMATION FROM THE ARF IS USED IN DATA ANALYSIS

| | Bureau of Health Professions | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| How do you use information from ARF? | | | | | | | |
| Directly | 1 | 1 | | 1 | | | 3 |
| Another data set | 1 | 3 | 8 | 3 | 3 | 3 | 21 |
| Other | 1' | | | | | | 1 |
| Non-response | 1 | | | | 1 | | 2 |
| Number of interviews: | 4 | 4 | 8 | 4 | 4 | 3 | 27 |

'This respondent indicated a potential use of the ARF data system.

TABLE 11.10

ADEQUACY OF DATA ELEMENTS

| | Bureau of Health Professions | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| Did ARF contain the data elements you needed? | | | | | | | |
| Yes | 3 | 3 | 4 | 3 | 3 | | 16 |
| No | 1 | 1 | 3 | | | 3 | 8 |
| Both | | | 1 | 1 | 1 | | 3 |
| Have you used other data sources in addition or instead of data on ARF which are available at the county level? | | | | | | | |
| Yes | 2 | 3 | 3 | 3 | 2 | 3 | 16 |
| No | 2 | 1 | 5 | 1 | 2 | | 11 |
| Would you like the file to be available for other geographic areas other than the county? | | | | | | | |
| Yes | 3 | 3 | 5 | 3 | 4 | 3 | 21 |
| No | 1 | 1 | 3 | 1 | | | 6 |
| Have you decided against using ARF because data were not adequate? | | | | | | | |
| Yes | 1 | 1 | | 1 | 2 | 2 | 11 |
| No | 3 | 3 | 2 | 3 | 2 | 1 | 16 |
| Number of interviews: | 4 | 4 | 8 | 4 | 4 | 3 | 27 |

Clearly, since the respondents had used the file, they must have found the data acceptable or adequate for their purpose. Thus it is not surprising that over half of the users interviewed (16 of 27) felt that the file contained the data elements they needed. The three individuals who indicated that the file contained only some of the data elements explained that the problem was with the currency of certain data elements.[25] However, 11 of 27 users had decided against using the file on at least one occasion because the data were not adequate, mostly because the data they wanted was not on the ARF (nine respondents).

More than half of the users have used county-level data sets other than the ARF to augment the ARF (see Table II.10). Three reasons for using other data with ARF were indicated: (1) the other data set at the county level was more extensive in a specific subject area; (2) the other data set at the county level was more current in that it could be obtained before being available on the ARF; or (3) the other data completed a time-series. The additional data sets used, as well the information used from the data set and the reason(s) given for doing so, are shown in Table II.11. Users indicated that certain types of variables already present on the ARF could be expanded and made suggestions on how this might be done. These comments included:

- Health Professions: There is a lack of information on Allied Health and nursing professions. The possibility of obtaining data from various lesser known health associations was suggested.

[25] These three individuals said: "sometimes the problem was that information was out of date;" "The variables are on the file, but they are not always current enough;" "It is not the scope of the information on the ARF but rather the timeliness of some of the data that presents a problem."

TABLE II.1 1

ADDITIONAL DATA SOURCES USED BY ARF USERS

| Source | Type of Data | Reason |
|---|---|---|
| NCHS and Information Sciences | Annual infant mortality rates | Not present on the ARF. For the years in which total births and deaths in all categories are provided, annual mortality rates can be constructed. |
| National Planning Research | Population counts | For one user, use of these data was prior to the inclusion of population counts on the ARF. However, these data are proprietary, and cannot be included on the ARF. |
| InterStudy Edge | Number of HMOs | Data is more current, in that Interstudy publishes data every quarter and the ARF data only has information as of one quarter (for example, the number of HMOs for 1987 is as of March 1987, and as of June 1988 for 1988). |
| University Data Bank | A wide array of data elements | The ARF is just one component of the Data Bank. The Data Bank has some of the same source data as the ARF, such as the Censuses of Population (1970 and 1980). |
| AOA | Physician counts by specialty | Data were more recent. Interviewee stated that once the data were received, they would be made available to ODAM for inclusion on the ARF. |
| Census Bureau | Population estimates | The Census Bureau adjusts the population estimates they publish. The user knew the exact definition the Census Bureau used in constructing the estimates as a result of going to the source. |
| AMA | Physician data | User received the data at the same time ODAM receives the data for inclusion on the ARF. |

TABLE II.1 1 (continued)

| Source | Type of Data | Reason |
|---|---|---|
| AHA | Facility-level data | 1) User receives the data at the same time ODAM receives that data for inclusion on the ARF. 2) Several users were processing this data themselves based on rumors from colleagues that the AHA data were not "clean". In processing the data on their own, users know exactly how variables are defined (such as the criteria used for inclusion/exclusion, double counting). |
| Center for Diseases Control | Morbidity | User receives the data at the same time ODAM receives the data for inclusion on the ARF. |
| Health Care Financing Administration | Medicare claims with pay locality codes | Medicare claims data have recent pay locality codes. |
| City County Data Book | Prevailing charges | Several users created their own prevailing charge indexes from this source[1]. The City and County data book is also available in CD-ROM format, which makes retrieving the data easily accessible for one user. |
| Department of Labor (BLS) | Employment and wage data | Information obtained was not present on the ARF. |
| State Health Departments | Infant Mortality, climate, physicians by specialties | For users' purposes, data were more current and easily obtainable. |
| CoSTAT3 | unknown | |

'Prevailing charges for general practice and specialties by specialty for 1978 and 1984 were deleted from the ARF in August 1986. Currently, there are three overall prevailing charge indices (1984 data) on the ARF: an index for general practice, for specialties, and for general practice and specialties.

- Census Data: Users said that the Census data dealing with population numbers should have included population estimates, as well as projections.*" Population estimates of the persons aged 65 and over is an example. In addition, users felt that it would be useful if the economic variables currently on the ARF could be provided annually and not limited to the Census years.[27]

- HMO data: Users stated that HMO industry data was lacking on the ARF. At present, there are two HMO variables on the ARF: the number of HMOs in a county and the number of federally qualified HMOs in a county. Users felt that these two variables were an inadequate portrayal of the HMO industry, and that more current data were available.[28]

- Mortality Data: Several users wanted annual mortality rates included in the ARF, particularly for infants.[29]

- Price Data: Hospital, nursing home, and physician prices.

- Geographic Codes: Social Security Administration state and county codes, and the New England Consolidated Metropolitan Area (NECMA) codes.

The user comments reflect both valid problems as well as the following concerns:

- For specific purposes, when timeliness is critical, it may always be faster to update a personal data base from the source than to wait while ODAM updates the ARF from that source and makes it available

- Users of the ARF who used the ACCESS system could have difficulty creating a variable (such as annual infant mortality) because the component data must be

---

[26]These users were apparently unaware that the ARF does contain Census population estimates.

[27]These users were apparently unaware that personal income by year had been added to the ARF from the Local Area Personal Income tape produced by the Bureau of Economic Analysis.

[28]There are no reliable comprehensive sources on HMO enrollment by county. Table II.11 provides a further comment on the available data.

[29]Currently, five year infant mortality rates are provided as well as numbers of births and numbers of deaths by race, sex, and length of life from which annual infant mortality rates may be computed.

added, and this could exceed the technical limits of ACCESS or the user's technical understanding.

- Some users remembered problems that existed in the past but which had been corrected
- Some users wanted data that do not exist at the county level.

## 8. Summary

This section reviewed the uses of the ARF. The principal findings were that HRSA is a major user of the ARF and within HRSA, BHPr's ODAM and BCHDA's ODM appear to be the major users. Users outside of the federal government were able to provide an extensive list of publications that drew on the ARF (see Appendix C). A review of ten of these publications showed that the ARF was not cited in four out of ten. Thus it is not surprising that an automated literature search found only three publications with ARF citations, of which one was a BHPr publication about the ARF. Lack of formal citation made it difficult to be certain how the ARF was used in some analyses, including those that are presented in BHPr's Sixth Report to the President and Congress on the Status of Health Personnel in the United States.

The major data elements used appear to be physician, hospital, population, income, infant mortality, and employment and unemployment. Users sometimes use other data sources available at the county level instead of the ARF, including those that are normally included in the ARF. Some users would like data added to or removed from the ARF, although we found that some of the data they would like added are already in the file or are not available, and some of the problem data they cited have already been removed from the file.

## C. DATA QUALITY

In addressing the question of whether the ARF is sufficiently accurate and timely, it must be recalled that the ARF is a secondary data source that draws on other data sources available at the county level of aggregation or below and provides a composite of sources for the user. Thus ODAM's responsibility for accuracy and timeliness is limited. Nevertheless there are several areas in which ODAM can control accuracy and timeliness. With respect to accuracy these areas include:

- Choosing whether to add a data source, based on recommendations from the source of the file, nonresponse rates, or other measurable problems
- Choosing whether to delete a data source based on problems encountered by users or additional information from the source
- When a source file is at the sub-county level, choosing how to aggregate data to the county level and choosing how much data review to conduct in that process of aggregation
- Choosing how to present missing data

With respect to timeliness, ODAM can time publication of a new ARF to coincide with the addition of major data sources, and can minimize the time taken to add such sources to the file; we recognize that ODAM attempts to maintain flexibility in the file production to meet these objectives.

Two major considerations when deciding whether to use a particular data set are the users' perceptions of the accuracy and timeliness of the data elements. If a user believes data to be either inaccurate or out of date, then the user will avoid using that particular data set unless no better alternative exists. This section addresses users' perceptions of the accuracy and timeliness

of data elements on the ARF. In addition, we have incorporated our findings from the evaluation of the ARF's content where appropriate. In our discussion of data accuracy we focus on the topics of nonresponse rates of the source data, zero-filling of fields, and problems cited concerning specific data elements and geographic coding, as well as the validity of these statements.

1. Accuracy

In general, the ARF users we spoke with understood that the file can be no more accurate than the component data sources. Some were aware of problems with particular source files, while others assumed that the ARF would contain only those data which reach some minimally acceptable level of completeness and accuracy. Table II.12 shows that the majority of the respondents indicated that they were satisfied with or had never questioned the accuracy of the data contained on the ARF. Several interviewees stated that they had verified some of the data elements and found that they corresponded with either the primary data source or another independent data source.

a. Nonresponse Rates in Source Data

A major facet of the accuracy of data elements is the extent to which high nonresponse rates in the data are avoided and nonresponse is explained. Only three individuals, all employees of the federal government, stated that the accuracy of the data on the ARF was impaired by nonresponse rates. Two BHPr respondents both were concerned with the nonresponse rates in the source, and thus the accuracy, of the health professions data. While these respondents

TABLE II.12

ACCURACY OF DATA ELEMENTS

| | Bureau of Health Professions' | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| Are you satisfied with the accuracy of the data? | | | | | | | |
| Yes | 3 | 4 | 6 | 3 | 3 | 3 | 22 |
| No | | | 1 | | | | 2 |
| Unable to answer | 1 | | 1 | 1 | 1 | | 3 |
| Are you satisfied with the missing values of data elements? | | | | | | | |
| Yes | 2 | 4 | 6 | 4 | 3 | 3 | 22 |
| No | 1 | | 1 | | | | 2 |
| Unable to answer | 1 | | 1 | | 1 | | 1 |
| Have you ever had a problem matching different types of geographic codes to the county codes? | | | | | | | 2 |
| Yes | | 1 | | 2 | 1 | | 4 |
| No | 4 | 3 | 8 | 2 | 2 | 3 | 22 |
| Unable to answer | | | | | 1 | | 1 |
| Number of interviews: | 4 | 4 | 8 | 4 | 4 | 3 | 27 |

[a]The Interviews with Bureau heads or designees and the interview with ODAM staff are not included in this table.

perceived that the information received from AMA and AHA are acceptable and accurate, the data on other health professions were not perceived to be of the same standard.[30] Many health professions have collected data on members only. Some have attempted to collect information on nonmembers, but have not received good response rates. In deciding which data to put on the ARF, ODAM has made tradeoffs between accuracy and timeliness.[31] The other respondent expressing dissatisfaction with the accuracy of the data stated that, based on reading through the documentation, there appeared to be some variables with inadequate response rates. This individual recognized that the inadequate response rates were a result of the data collection process used by the source data, and beyond the control of the ARF.

In examining the technical documentation, MPR found two problems with the way in which nonresponse rates are documented. The first problem MPR found was that no explicit criterion of an acceptable nonresponse rate is stated in the documentation. Second, MPR found that there is great variability in the way that nonresponse and unallocated data are documented.[32]

---

[30]In point of fact, the AHA data, though the best available source of hospital information, may be quite poor for rural areas. Some observers have pointed out that only about 50 percent of rural hospitals respond to the AHA questionnaire. Furthermore, one respondent was critical of the AMA data for physician counts, on the grounds that residents are included in the counts although they ought to be distinguished from other physicians. This user mentioned that AMA probably provides the data for the ARF in this manner. It should be noted that the ARF user documentation (March and August 1989) states that AMA physician counts for 1940, 1950, and 1960 exclude residents, although there is no indication that this holds true for the remaining physician counts from AMA on the file.

[31]Some sources of more recent data with very low response rates are not included in the file. For example, the American Podiatric Medical Association Survey data of 1984 are not included in the file.

[32]The description in the user documentation on the merging of the source data onto the ARF indicates the number of responses that could not be allocated to a county.

For some of the health professions, the total numbers in the source and the numbers successfully allocated by county are explained. For some of the health facility data, numbers of unallocated facilities are given, but no totals, so that no sense of the magnitude of the problem can be gained. Two examples of this are:

- Veterinarians for 1977 and 1979: The number of nonmerged cases is given but the total number of cases is not, hence the user has no sense of the scale of the nonmerge problem.
- Pharmacists for 1980: The total number in the source, the numbers deleted, and the total included in ARF are all given but they do not add up.

b. Zero-Filling of Fields

A user pointed out that, at times, it is difficult to distinguish between a zero indicating that the field has been suppressed or is missing and a zero representing a true value for a particular data element. This is highlighted for variables which have a one byte field indicating the suppression of data. When such fields from the ARF are being used, the user must also look at the byte indicator; if they do not, they cannot distinguish whether a zero value for a data element is an indication of suppressed and/or missing data or is a genuine value for the variable. For example, when the number of Hispanic families with an income of greater than $50,000 has a value of zero it may mean that the data were suppressed, were missing, or that no Hispanic families with this income range were present in the county.

c. Data Problems

The respondents who said they were dissatisfied with the accuracy of data in the ARF mentioned data that had been in the ARF but have subsequently been removed because they were created problems. Three problem areas were mentioned by users:

- Prevailing Charges
- HMO Data
- Nursing Data

Prevailing Charges. According to some users, the specific prevailing charges for general practitioners and specialists present on the file were incorrect and the values did not make sense upon examination of the raw data. These data were removed from the file in August of 1986.[33] A federal employee stated that the prevailing charge region and carrier codes appeared to be incorrectly merged onto the ARF. These two codes were adjusted in the March 1988 version of the ARF, and now appear to be correctly merged onto the file.

HMO Data. A number of respondents criticized the HMO data in the ARF, on the grounds that the data were inaccurate and misleading. The problem with the data (enrollees per county from InterStudy) centered around the fact that the InterStudy counts placed enrollments in the county where the HMO had its headquarters rather than by the county of residence of the enrollee. In fact, ODAM removed the HMO enrollment information from the file in May 1986

[33]ODAM had the ARF contractor develop an algorithm which created consolidated prevailing charge indices based on the specific prevailing charges provided by HCFA. The three overall indices, for general practitioners, specialists, and all physicians combined, replaced the specific prevailing charges.

because users were interpreting the information incorrectly. Currently, the ARF contains information from InterStudy on the number of HMOs in a county, as well as the number of federally qualified HMOs in a county, where these counts also are based on the location of the headquarters of the HMO. This is not clearly stated in the user documentation, and if these fields are included in future versions of the documentation, they should be clarified. In fact, Interstudy itself states: "While the headquarter location is generally representative of the service area of a particular HMO, it may not completely reflect the service area served" (InterStudy, Summer 1987).

Information on the number of HMOs which are federally qualified could be obtained from HCFA's Office of Prepaid Health Care (OPHC). However, federally qualified HMOs represent approximately 55 percent of all HMOs in operation which indicates that this source would not be comprehensive.

Nursing Data. The data describing the number of registered nurses in the ARF were considered by Division of Nursing staff to be so obsolete as to constitute a problem in terms of use of the data by other individuals (there were also problems in general with the registered nursing data which came from the 1980 Census). Our respondent was seriously concerned that when data are so old, most uses of these data will be misuses.[34] Despite the age and possible inaccuracy of the data, no other data are collected at the county-level which could replace those data elements currently included on the ARF. The data are clearly of use for historic analyses.

[34]The respondent does not use the nursing data found on the ARF file (although demographic and socioeconomic variables are used) because information on the nursing profession is available through sample surveys by the Division of Nursing. However, the sample survey data collected by the Division of Nursing would not be appropriate to add to ARF.

Other Problems. Two additional types of data problems were described by users:

- Coarse coding of some variables that limits the information they provide
- Misleading data

An example of coarse coding is found in short term general hospital bed-size. These variables (which are also documented in a way that makes it hard to know how they are defined until the data are actually used) indicate the number of hospitals in the county with a specified bed size range in a given year. In 1987, 1,862 counties (60.5 percent) had one or more hospitals with 6 to 99 beds (representing 2,634 hospitals). Thus, of the universe of counties, almost two thirds have been effectively classified in one cell. Since these small hospitals are those that are struggling financially and since they provide a large component of care in many areas, it would be helpful from a policy perspective if the classification used could provide more detail, for example 6 to 24 beds, 25 to 49 beds, 50 to 74 beds and 75 to 99 beds. A similar problem exists for occupancy rates where rates of 0 to 69 percent are included in one category, while the average occupancy rate is now less than 69 percent.

Another example of misleading data (HMO data was discussed earlier) is that two sets of total population data appear to be provided in the file, with sources being cited as P-26 and LAPI tape. In fact the LAPI data is not total population since it includes businesses (as the user documentation indicates), hence it is mislabeled in the technical documentation. A further issue is that the P-26 population estimates provided in the file would appear to be inconsistent. The Census Bureau reestimates the previous decade's estimates at the time of the succeeding Census. Hence the P-26 estimates of the 1970s should have been replaced when intercensal

estimates became available in the early 1980s. Instead the file includes P-26 data added during the 1970s. The time-series is not consistently estimated from decade to decade unless the updated estimates are used.[35] Neither The P-26 population estimates nor their origins are described.[36]

d. Geographic Code Problems Cited by Users

Three individuals stated that ARF's handling of the independent cities of Virginia has hindered their use of the data. The coding of these cities is unique to the ARF, and all three users felt that this presented obstacles to extracting historical data. In addition, this coding could lead to inconsistencies when users try to link variables from external data sources. We understand that BHPr has attempted to code the Virginia independent cities in the way most helpful to users, and that some like the current coding and others do not. Two possible solutions to the problem are:

- A trailer file could be created which would give available data for FIPS county codes only
- Tables of county codes which are appropriate to use depending on which year of data is being used, could be provided in the documentation with a clear explanation

Some users were concerned about the absence of the Social Security Administration state and county codes, which is the principal geographic coding used by the Health Care Financing

[35] Telephone conversation with Bureau of Census staff 11/89.

[36] It should be noted that though the P-26 population estimates may be the best available, data are derived from the states which may use slightly different methodologies.

Administration. While some users have obtained a FIPS to SSA crosswalk (of which ODAM has a copy), the addition of the SSA codes to the file would make it more accessible.

In addition, two individuals were concerned about MSA codes for the New England region. They felt that the MSA was not a correct representation for this region, because the New England MSAs are based on townships rather than counties. They suggested that the New England Consolidated Metropolitan Areas (NECMA), which are similar to the MSA codes, would be a more proper geographic coding for this area because these codes are truly at the county level for the New England region. A trailer file containing data by MSA/nonMSA for split counties in the New England Census division could improve the comparability of the data with other areas.

2. Timeliness and Time Series

Another major consideration in determining the extent to which a data set will be used is the timeliness of the data elements. It is understandable, however, that individuals using data for different purposes will have different perceptions of what they consider to be "timely" data. Accordingly, in MPR's interviews, respondents were asked their attitudes towards the timeliness of the data components of the ARE

As shown in Table II.13, one third of the twenty-four individuals responding to the question on data timeliness felt that some of the variables on the ARF were not sufficiently timely (for their purposes). The opinions expressed by these users focused on two distinct issues:

- The length of time which elapses between when a data source is made available to the general user community and when it is incorporated into the ARF data system.

TABLE II.13

TIMELINESS OF THE DATA

| | Bureau of Health Professions' | Othe PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| Are the data elements you use sufficiently timely? | | | | | | | |
| Yes | 2 | 3 | 6 | 1 | 2 | 2 | 16 |
| No | | | 1 | 1 | 1 | 1 | 4 |
| Both | 2 | 1 | 1 | | | | 4 |
| Unable to respond | | | | 2 | 1 | | 3 |
| Have you tried to find more timely data? | | | | | | | |
| Yes | 2 | 1 | 3 | 2 | 1 | 2 | 11 |
| Unable to respond | 2 | 3 | 5 | 2 | 2 | 1 | 15 |
| | | | | | 1 | | 1 |
| Of those trying to find more timely data, were you successful? | | | | | | | |
| Yes | 2 | 1 | 3 | 2 | 1 | 1 | 10 |
| No | | | | | | | |
| Sometimes | | | | | | 1 | 1 |
| Number of interviews: | 4 | 4 | 8 | 4 | 4 | 3 | 27 |

[a]The interviews with Bureau heads or designees and the interview with ODAM staff are not included in this table.

- The timeliness of the data elements themselves which are contained on the data source used as input into the ARF.

As we indicated above, ODAM, as the producer of the ARF Data System, has limited control of these issues since the timeliness of the data elements is largely a function of how often the source updates the data elements. For instance, some of the data elements included on the ARF from the Census are only updated every ten years; therefore, at present 1980 is the most timely data available for the field representing the number of housing units which are occupied year round. It is likely that the user's version of the file would have a significant role in the perception of the timeliness of data elements; some nonfederal users indicated that they obtain a new ARF only once a year.

Some respondents handled timeliness problems on ARF by obtaining more recent data from other sources and merging them with the ARF. Eleven respondents indicated that they have tried and succeeded in finding more timely data than those contained on the ARF. Of these eleven respondents: two found information at a unit of observation other than the county (individual Medicare beneficiaries and state Medicaid rates for nursing homes); two from data sources which are not being used as input into the ARF (allied health and nursing staff); and six from the source of the data used for the ARF (Census, BLS, BEA, AMA, AHA). The fact that individuals are going outside of the ARF for more timely data may be indicative of two potential problems: 1) the length of time it is taking to incorporate data elements into the ARF data system once the data are made available to the general user community; and 2) primary data sources which potentially have more timely data are not always incorporated into the ARF.

However, an alternative explanation is that the user did not require nationwide county data, and could use a source that ARF could not.

In general, respondents believed that a lag of one to two years was reasonable for the data elements on the ARF. This means that in the 1989 versions of the ARF, the most recent data elements should be for 1988 or 1987. Table II.14 shows the most recent data for key health professions, facilities, vital statistics, and demographic data. (Table II.14 is based on the August 1989 version of ARF.) As we see, the types of data meeting the suggested 1987 or 1988 cutoff are hospital data from the AHA (1987), population estimates from Census (1987), the number of doctors of osteopathy (1987), the number of active MDs, and the number by specialty (1988) and per capita income (1987). Data not meeting the 1987 or 1988 cutoff are total births, deaths, and infant deaths (which were available for 1986 in the August 1989 version of the tile) and other types of health professionals counts (which are from 1980 or 1981, and hence are usable only for historical analyses). As mentioned earlier, some respondents handled this by obtaining the source data at roughly the same time ODAM receives the data for inclusion on the ARF.[37]

Two users were concerned about the discontinuities that occur in timeseries data for some variables: one user stated that gaps in annual data required the use of other data sources to obtain a complete ten year period of data, while a second was not able to match all data for the same year. They wanted to use the ARF to examine data elements over a span of time--ten years, for instance--or to examine the data at two points in time. They were unable to do so

[37]The 1987 population estimate was published in September 1988 and first included on the file in March 1989. This is the earliest possible date at which the data could have been included. However, six months elapsed during which users could obtain the data because it was not available on the ARF.

TABLE II.14

MOST RECENT YEAR DATA ARE AVAILABLE FOR KEY FIELDS
(AUGUST 1989 EDITION OF ARF)

| Data Element/Type | | Year | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Prior | 1980 | 1981 | 1984 | 1985 | 1986 | 1987 | 1988 |
| M.D.s | Total Active | | | | | | | | X[a] |
| | by specialty | | | | | | | | X |
| | by sex | | | | | X | | | |
| | by age and specialty | | | | | X | | | |
| D.O.s | Total Active | | | | | | | X | |
| | by specialty | | | X | | | | | |
| Dentists | Total Active | | | X | | | | | |
| Veterinarians | Total Active | | X | | | | | | |
| Pharmacists | Total Active | | X | | | | | | |
| Optometrists | Total Active | | | X | | | | | |
| Podiatrists | Total (Census) | | X | | | | | | |
| RNs | Total (Census) | | X | | | | | | |
| LPNs | Total (Census) | | X | | | | | | |
| Physician Assts. | Total Active | | | X | | | | | |
| Hospitals | Total | | | | | | | X | |
| Hosp. Beds | Total | | | | | | | X | |
| Hosp. Admissions | Total | | | | | | | X | |
| Population | Total | | | | | | | X | |
| | Estimate | | | | | | | X | |
| | by age and sex | | | | | | | | |
| Mortality | | | | | | | | | |
| Infant by race and sex | | | | | | | X | | |
| All, by major disease | | | | | | | X | | |
| 5 year infant deaths | | 1974-1978 | | | | | | | |
| 5 year infant mortality rates | | | | | | | X (1982-1986) | | |
| Natality | | | | | | | | | |
| Total births, by sex and race | | | | | | | X | | |
| 5-year live births | | 1974-1978 | | | | | | | |
| Income | per capita | | | | | | | X | |
| | family median | | 1979 | | | | | | |

[a]1988 MD data are for 198711988.

because of the absence of one or more years of data. For the key fields shown in Table II.14 uninterrupted time series data in the August 1989 ARF is shown in Table II.15 For these key fields (the most used), some of the time series are reasonably complete. However, some of them are misleading. For example, the Census (P-26) population estimates used a different methodology in the 1970s than in the 1980s, and use of estimates from both decades could be misleading. Presumably, though the documentation provides no explanation, 1984 and 1987[38] physician data were either uncollected or unusable. It would be helpful to have D.O. data on the file, both total and specialty, at regular intervals to use with the MD total and specialty data (preferably years in which MD data also were available), if these data were available. Population data by age and sex for interim and regularly repeating years would be particularly helpful, again if these data were available. The lack of timely allied health professions data is repeated in the lack of time-series data.

3. Comment and Summary

Is the ARF viewed as accurate and timely? Considering the size of the file, remarkably few errors were identified, and some of these were errors of labeling rather than content. A few fields need to be corrected; some fields need to be more clearly labelled or described. Otherwise, there are problems related to the manner in which data are aggregated. These problems could be solved by changes in policy concerning data aggregation. Currently, the ARF is seen as valuable because it relieves the user of the burden of seeking out and combining

[38]However, ODAM staff indicated that 1988 data cover the 1987-1988 period.

TABLE II.15

TIMESERIES DATA AVAILABLE FOR KEY FIELDS
(AUGUST 1989 EDITION OF ARF)

| Data Element or Type | | Years in Which Available | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 75 | 76 | 77 | 78 | 79 | 80 | 81 | a2 | 83 | 84 | 85 | 86 | a7 | 88 |
| MDs | Total | X | X | X | X | X | X | X | X | X | | X | X | | X[a] |
| | Specialty, total | (X) | | | | (X) | | | (X) | (X) | | X | (X) | | X |
| | by sex, total | X | | | | | | | | | | X | | | |
| | by age and specialty | X | | | | | | | | | | X | | | |
| DOs | Total active nonfederal | | b | | | | | X | | | | | | X | |
| | by speciality | | b | | b | | | X | | | | | | | |
| Dentists | total and by specialty | | | | | | b | X | | | | | | | |
| Veterinarians | total and by practice | b,c | | b,c | | X | b | | | | | | | | |
| Optometrists | total | | | | b | X | X | X | | | | | | | |
| RNs | graduates of foreign programs | | | X | | | X | | | | | | | | |
| Hospitals | total # hospitals | X | X | X | X | X | X | X | X | X | X | X | X | X | |
| | total # admissions | | | | | | X | X | X | X | X | X | X | X | |
| | total # beds | X | | | | | X | X | X | X | X | X | X | X | |
| Population | Census estimate (100s) | X | X | X | X | d | b | d | X | X | X | X | X | X | |
| Population by age/sex | NCI | X | | | | | b | | | | X | | | | |
| **Natality** | | | | | | | | | | | | | | | |
| Total births | | X | X | X | X | X | X | X | X | X | X | X | X | | |
| Births by sex/race | | X | X | X | X | X | X | X | X | X | X | X | X | | |
| 5 year births | | (69-73)[b] | | | 74-78 | | | | | | | | | | |
| **Mortality** | | | | | | | | | | | | | | | |
| Infant by race and sex < 1 year | | x | x | X | X | X | X | X | X | X | X | X | X | | |
| < 1 month (7-27 days old) | | | | | | | | | | X | X | X | X | | |
| All, by major disease | | x | x | X | X | X | X | X | X | X | X | X | X | | |
| Per capita income (LAPI Tape) | | x | x | X | X | X | X | X | X | X | X | X | X | X | |

NOTES: 1. Only series across a single source are given.
2. Parentheses indicate that data are not present for all categories.
3. Some health professionals are not shown because no time-series are available.

[a]1988 MD data are for 1987/1988.

[b]Available from a different source (including census) or an enhanced file.

[c]By type available, but no totals given.

[d]Not available. Also, note that 1980 is census, not estimate. and that methodology changed from 1970's to 1980's.

separate data sources. Some users would like more value added in the form of more data checks on source data before they are added to the file.

It is clear from what respondents told us that errors in the file are remembered long after they have been corrected. To avoid this problem in the future, ODAM would have to undertake a more thorough review before adding data to the file, a review that could include more discussions with the source or persons who had used the source data, and more extensive review of the data itself. To assist in this process, a formal set of criteria for inclusion (or exclusion) could be established and included in the documentation. Another way of handling the problem might be improved communication with users.

For some purposes extreme accuracy is essential, and in those cases the ARF may not be the best data source. For analytical purposes where the ARF is being used to supply a set of area variables as control variables in a regression analysis, minor errors in the data are unlikely to be serious, in other words, the information provided by the data adequately represents the variability across counties and provides useful information.

With respect to timeliness, it may never be possible for a secondary data file like ARF to be timely enough for activities related to program implementation. However, the file is sufficiently timely for many descriptive and comparative analyses of the health care system and only one third of respondents were concerned about timeliness. Users recognize that BHPr cannot control the collection of adequate, timely, and equivalent data across years, because it does not have the budget to do so. Nevertheless, this creates complexities for the users of the file. While we have focused on some of the best data in the file, there are greater problems of timeliness and dislocation of time series for many other variables (allied health professions).

ODAM could aid users by providing better documentation on exactly what is available in the file across years and the consistency of those data.

## D. EASE OF USE

We reviewed three aspects of access to the ARF: (1) the process through which users obtained the ARF and the related documentation; (2) how the respondents accessed the file; and (3) the comprehensibility of the file and its documentation.

### 1. Obtaining the ARF

Table II.16 summarizes users' experiences in obtaining the ARF data file and documentation. The ARF tile is obtained through (1) ODAM or their ACCESS system and (2) NTIS.[39] In Table II.16, it can be seen that all but two of the individuals we spoke with received the ARF through these regular channels. ODAM was the most common source (16 users), the ACCESS system was the second source of access (8 users) and NTIS was the source for two users. This distribution (in favor of ODAM as the source) reflects what we found from ODAM and NTIS distribution records (see Section A of this chapter). Of the other two users, one contractor indicated receiving the ARF through ODAM as well as the Health Care Financing Administration; another respondent associated with a professional association obtained the file through an affiliated organization.[40]

[39] ODAM terminated the NTIS contract and is looking for an alternative vendor for the ARF products. The new ARF file contract appears to make it possible for the ARF contractor also to supply the file and documentation.

[40] As indicated in Section B, it is not uncommon for staff at an organization to use one copy of the ARF for multiple purposes.

TABLE II.16

OBTAINING FILE AND DOCUMENTATION

| | Bureau of Health Professions' | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| Where did you get most recent copy of ARF? | | | | | | | |
| ODAM | | | 7 | 3 | 3 | 3 | 16 |
| NTIS | | 1 | | 1 | | | 2 |
| Access System | 4 | 3 | 1 | | | | 8 |
| Other | | | | 1 | 1 | | 2 |
| When did you obtain your copy of the ARF file? | | | 4 | | | 1 | |
| 88 | 2 | 4 | 3 | 2 | 2 | 2 | 15 |
| 87 | | | 1 | | 1 | | 6 |
| 86 | 1 | | | | | | 1 |
| 85 | 1 | | | | | | 1 |
| | | | | | | | 1 |
| Did not indicate | | | | 2 | 1 | | 3 |
| Had you ever obtained the file before? | | | | | | | |
| Yes | 4 | 3 | 4 | 2 | 3 | 3 | 19 |
| No | | 1 | 4 | | 1 | | 6 |
| Did not indicate | | | | 2 | | | 2 |
| Did you have problems getting a copy of the ARF file or documentation? | | | | | | | |
| Yes | 1 | | | | 1 | 1 | 3 |
| No | 3 | 4 | 8 | 4 | 3 | 2 | 24 |
| Did you request the accanpanying documentation for the file? | | | | | | | |
| Yes | 4 | 4 | 7 | 4 | 3 | 3 | 25 |
| | | | | | | | 1 |
| &not recall | | | 1 | | 1 | | 1 |
| Number of interviews: | 4 | 4 | 8 | 4 | 4 | 3 | 27 |

[a]The interviews with Bureau heads or designees and the interview with ODAM staff are not included in this table.

In general, individuals had no problem obtaining a copy of the ARF and its documentation. Within BHPr, there appears to be some problem with the internal distribution of the documentation. While three of the four respondents had no problems in obtaining the documentation, one user felt that having to continually request an updated version each time the ARF was to be used was a hindrance. Of the other two users indicating a problem, one had received the tape through an affiliate and stated that availability was a problem,"' while the other claimed that NTIS was uncooperative and had sent the incorrect tapes on a number of occasions. This respondent succeeded in obtaining the ARF from ODAM.

HRSA (nonBHPr) interviewees suggested that an update memo, outlining the newly added data associated with each version of the ARF, could be distributed internally by ODAM at the semiannual update. This would be feasible for two reasons: (1) there is a relatively small number of active users who would receive the memo, and (2) Section III of the ARF user documentation lists adjusted, newly added, and deleted variables for the past year.[42] This listing could be duplicated in another format.

Six of twenty-five respondents stated that they had never obtained the ARF file prior to the current version being used. All of these users can be classified as "first-time" or one-time-only users. In addition, over half of the respondents thought that they had received their version of the ARF file during 1989 (we particularly selected recent users to interview).

---

[41]This respondent, who received the mail version of the interview guide, merely mentioned "availability" as a problem, and it is uncertain whether this pertained to the ARF data or the ARF documentation. This respondent indicated that the documentation for the ARF had not been requested.

[42]There is no explanation of why a particular variable was adjusted or deleted. ODAM may want to consider the possibility of providing a brief reason for why the adjustments and deletions were implemented.

Although MPR's interviews centered upon respondents' use and perceptions of the ARF Basic County file, it also was important (particularly for HRSA users and their contractors) to determine whether ARF users were aware of the other ARF formats available, since the ARF Basic County file is just one component of the entire ARF data system. There are in addition the micro-computer diskette series and the source files.[43]

Table II.17 presents the distribution of individuals who were aware of other ARF formats. More than half of the respondents were aware of the various formats in which the ARF is available and the government users were the only users of other formats. This table highlights the fact that even when users are aware of the various formats, they are rarely using any but the Basic County file.

In order to get a sense of how often the diskette versions of the ARF were being used, we requested information from NTIS on recent purchases of the file.[44] As reported in Section A there were no diskette purchases within the last six months (although ODAM reports that the diskette riles have been distributed to PHS regional offices that requested them). The reason for the lack of purchases ODAM suggests may be the cost: $50 per diskette and $675 for the entire health professions data series, demographic data series; $775 for the entire health facilities data series and update series; $1,975 for the complete series (on 114 diskettes). Indeed, ODAM staff explained that they have terminated the NTIS contract and are looking for an alternative

[43]The state and national time series and the graduate medical education file also exist, although they have not been updated recently.

[44]Because of the very large number of files available, we requested information on a random number of the diskette files from NTIS. The files selected were the update series for the following states: (1) Alabama, (2) Arizona, California, Hawaii, New Jersey, and Nevada, (3) Connecticut, New Jersey, New York, and Rhode Island, and (4) Illinois; the update series for State/National, and the complete update series on diskette.

TABLE II.17

FORMATS

| | Bureau of Health Professions* | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| Are you aware of the various ARF formats? | | | | | | | |
| Yes | 3 | 1 | 5 | 2 | 4 | 2 | 17 |
| No | 1 | 3 | 3 | 2 | | 1 | 10 |
| Which formats have you used: | | | | | | | |
| Basic County File | 4 | 4 | 8 | 4 | 4 | 3 | 27 |
| State/National Timeseries Data Bases | 1 | | | | | | 1 |
| Area Profile Reports | 2 | 1 | 1 | | | | 4 |
| Graduate Medical Education File | 2 | | 2 | | | | 4 |
| Microcomputer Data Series | | | 1 | | | | 1 |
| Support or Source Files | 1 | | | | | | 1 |
| Number of interviews: | 4 | 4 | 8 | 4 | 4 | 3 | 27 |

'The interviews with Bureau heads or designees and the interview with ODAM staff are not included in this table.

vendor for the product. Another possibility is to make the data available through a computer bulletin board. In either approach significantly lower prices for the user are likely.

Source files are generally available only to BHPr users or their contractors. Table 11.17 shows that source files are only rarely used by the users we talked to (which may partly be related to the difficulty ODAM has experienced in getting permission to access some of these files).

2. Using the Data File

The most common computer package used to access the ARF was the Statistical Analysis System (SAS) which users run on a mainframe (Table II.18). Due to the large size of the ARF data set, it is not surprising that most individuals are using a mainframe. In addition, approximately half of the respondents do their own programming, while the other half have access to programmers.

There is a user-friendly (ACCESS) retrieval system for the ARF data available to federal government employees with accounts at the National Institute of Health Computer facilities:[45] With this ACCESS system, an individual without technical programming skills is able to generate and execute a SAS program through a series of queries. All of the respondents from the Bureau of Health Professions, three of the four Public Health Service interviewees, and one of the eight respondents in the "Other Federal" category have made use of the ACCESS system. In general, individuals felt this was a useful, if rather simplistic, approach to accessing the ARF. Two of the eight individuals who have used the ACCESS system have programmers who work with the ARF, while the other six do their own programming. Some of the BHPr respondents believe that more training classes for the ACCESS system might encourage BHPr staff to make more use of

[45] Users must request the documentation and password from ODAM.

TABLE II.18

ACCESSING FILE

| | Bureau of Health Professions" | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| Do you do programing yourself or do you have programmers? | 4 | 1 | 1 | 1 | 1 | | 8 |
| Programmers | | 3 | 4 | 3 | | 3 | 13 |
| Contract programing outside of organization | | | 2 | | | | 2 |
| Both self & programmers | | | 1 | | 2 | | 3 |
| Did not indicate | | | | | 1 | | 1 |
| What programming languages are used? | | | | | | | |
| ACCESS System | 4 | 1 | 1 | | | | 6 |
| SAS | 3 | 1 | 5 | 4 | 3 | | 19 |
| PC/SAS | | | | | | | 1 |
| SPSS | | | 1 | | | | 2 |
| PC/SPSS | 1 | | 1 | | | | 3 |
| COBOL | | | | | | | 0 |
| FORTRAN | 2 | | 1 | | | | 3 |
| PL/1 | | | 1 | | | | 2 |
| C | | | | | | | 0 |
| DBASE | | | | | | | 0 |
| Other | | | | | | | 2 |
| Did not indicate | | 2 | 1 | | | | 4 |
| What system do you use to run ARF? | | | | | | | |
| Mainframe | 4 | 4 | 6 | 4 | | | 23 |
| Microcanputer | | | | | | | 0 |
| Both | | | 1 | | | | 2 |
| Did not indicate | | | 1 | | | | 2 |
| Have you had any technical problems accessing data or using documentation? | | | | | | | |
| Yes | | | | 1 | | | |
| No | 4 | 4 | 8 | 3 | 3 | 3 | 2: |
| Did not indicate | | | | | 1 | | 1 |
| Number of interviews: | 4 | 4 | 8 | 4 | 4 | 3 | 27 |

[a]The interviews with Bureau heads or designees and the interview with ODAM staff are not included in this table.

it. In contrast, ODAM staff indicated that the training courses they have held in past years were not well attended. Furthermore, by the time class attendees were ready to use the ACCESS system, the majority had forgotten what they had learned. Interviewees mentioned several suggestions for the improvement of the ACCESS system:

once in six years

- BHPr employees suggested that they should routinely be informed each time the password is changed,
- A HRSA user suggested that it would be beneficial to create a cross-index of variables associated with a particular issue. If this were done, then the index could be previewed and necessary variables identified prior to running reports using the ACCESS system.
- The ACCESS system allows only a limited number of variables to be handled at a time. This constraint was bothersome to one user.

3. Understanding the Documentation

As indicated in Table II.19, most users felt they understood the data elements on ARF based on the documentation. However, users mentioned two types of problems with the documentation: the user could not understand how the data element was coded, constructed, or modified (insufficient or confusing documentation); the user was familiar with the source data and did not feel that problems were adequately explained. For instance, two users knew that a number of problems existed with the AI-IA data, based on discussions with colleagues, and felt that the ARF documentation did not fully explain these problems.[46] A similar problem was described by *one* NCHSR user MPR spoke with briefly on the telephone (not shown in Tables). This user works with the Inventory of Long-Term Care Places file, a source data file for nursing home data included on the ARF. The respondent stated that duplicate cases existed in the

[46]They also felt that the data had been included on the ARF without sufficient editing for their purposes. Therefore, these users were processing the AHA data on their own.

TABLE 11.19

DOCUMENTATION

| | Bureau of Health Professions' | Other PHS | Other Federal | Contractors | Professional Associations | Academics | Total |
|---|---|---|---|---|---|---|---|
| Can you understand data elements from reading the documentation? | | | | | | | |
| Yes | 3 | 3 | 7 | 2 | 3 | 3 | 21 |
| Not always | 1 | | 1 | 1 | | | 2 |
| Did not indicate | | 1 | | | 1 | | 2 |
| Are problems with data fully explained in the documentation? | | | | | | | |
| Yes | | 3 | 7 | 1 | 3 | 2 | 18 |
| No | | | | 2 | | 1 | 4 |
| Not always | | | 1 | | | | 1 |
| Did not indicate | | 1 | | 1 | 1 | | 4 |
| Who would you seek advice from if you had a problem with the documentation? | | | | | | | |
| OOAM | | 1 | 4 | 2 | 1 | 3 | 14 |
| NTIS | | | | | | | 0 |
| ANS | | 1 | 1 | 1 | | | 3 |
| Source of ARF data | | | | 2 | | 1 | 3 |
| Co-worker | | | | 1 | 2 | 1 | 4 |
| Other | | | 1 | | | | 1 |
| None | 1 | 1 | 2 | | | | 4 |
| Did not indicate | | 1 | | | 1 | | 2 |
| Number of interviews: | 4 | 4 | 8 | 4 | 4 | 3 | 27 |

[a]The interviews with Bureau heads or designees and the interview with ODAM staff are not included in this table.

source file, which should be deleted before use. This problem is not documented in the technical documentation of the ARF as occurring or as having been corrected.[47]

In seeking advice on a problem associated with the documentation or a particular data element, users were most likely to contact ODAM, or deal with the problem on their own. Only a small number of users indicated that in encountering a problem would they seek the advice of a co-worker, communicate with the ARF contractor, or contact the original source of data on the ARF.

4. MPR's Evaluation of the ARF documentation

While users were generally happy with the ARF documentation, some did not like the way it is organized, and had problems with the lack of explanation for particular variables. We reviewed the documentation closely with respect both to overall principles of organization and for particular types of problems, such as consistency in documentation.

Although we began our review with the sense that the ARF is a straightforward file, clearly documented, it became clear to us as we reviewed the documentation why some users found it awkward. A considerable effort is needed before the nature and content of some variables can be fully understood. In general, we believe that the documentation should provide more complete, more consistent explanations, and should be supplemented with additional types of explanation not now given.

a. Layout

There are two sections to the documentation: a technical section and a user section. Exhibit II.1 is a page from the technical documentation. Exhibit IL2 is a page from the user

"Although NCHSR uses stricter criteria in defining a facility as a nursing home, the person MPR spoke with stated that there should be approximately 25,379 nursing homes for 1986. Due to the differences in criteria, the ARF count would be higher.

EXHIBIT II.1

ARF TECHNICAL DOCUMENTATION

AREA RESOURCE FILE



| FIELD | COL-COL | YEAR OF DATA | NAME | CHARACTERISTICS | SOURCE | DATE ON |
|---|---|---|---|---|---|---|
| F08857-74 | 00621-00626 | 074 | Total Active M.D.s Non-Federal | Non-Fed; Excl Inact | AMA Dis: Physicians | 03/77 |
| F08857-73 | 00627-00632 | 1973 | lotrl Active M.D.s Non-Federal | Noo-ted; Exclinact | AMA Dis Physicians | 11/84 |
| F08857-72 | 00633-00638 | 1972 | Total Active M.D.s Non-Federal | Non-Fed; Excl Inact | AMA Dir Physicians | 12/75 |
| F08857-71 | 00639-00644 | 1971 | Total Active M.D.s Non-federal | Nor-fed; Exclinact | AMA Dis Physicians | 12/75 |
| F08857-70 | 00645-00650 | 1970 | Total Active M.D.s Non-Federal | Non-Fed; Exclinact | AMA Dir Physicians | 07/72 |
| F08857-69 | 00651-00656 | 1969 | Total Active M.D.s Non-federal | Non-Fed; Exclinact | AMA Dis Physicians | 08/71 |
| F08857-68 | 00657-00662 | 1968 | Total Active M.D.s Won-Federal | Non-fed; Exclinact | AMA Dis Physicians | 12/75 |
| F08857-60 | 00663-00668 | 1960 | Total Active M.D.s Non-Federal | Non-f cd; Exclinac t | DMI,67- 'l | 11/71 |
| F08857-58 | 00669-00673 | 1958 | Total Active M.D.s Nom-federal | Non-Fed; Exc l Inac t | AMA Phy s Pop | 02/85 |
| F08857-50 | 00674-00679 | 1950 | Total Active M.D.s Non-Federal | Non-Fed; Exc l Inac t | DMI,30- 4 | 11/71 |
| F08857-40 | 00680-00685 | 1940 | Total Active M.D.s Non-Federal | Act. Non-Fed, Except AK, HI | DMI,58- 2 | 11/71 |
| F04602-86 | 00686-00690 | t986 | Total M.D.'s,Total Non-fod | Non-Fed;Incl Inact/Not Class | AMA Dis Physicians | 09/88 |
| F04602-85 | 00691-00695 | 1985 | Total M.D.'s,Total Non-Fod | Non-fed: Incl Inact/Not Class | AMA Dir. Physicians | 03/87 |
| F04602-83 | 00696-00700 | 1983 | Total M.D.'s, Total Non-Fed | Non-Fed; Incl Inact/Not Class | AMA Dist Physicians | 02/86 |
| F04602-82 | 00701-00705 | 1982 | Total M.D.'s,Total Uon-Fod | Non-Fed;Incl Inact/Not Class | AMA Dist Physicians | 05/85 |
| F04602-81 | 00706-00710 | 1981 | Total M.D.'s,Total Non-Fod | Non-Fed; Incl Inact/Not Class | AMA Phys Master File | 11/84 |
| F04602-79 | 00711-00715 | 1979 | Total M.D.'s,Total Non-Fod | Non-Fed; Incl Inact/Not Class | AMA Phys Master File | 01/82 |
| F04602-75 | 00716-00720 | 1975 | Total M.D.'s,Total Non-Fod | Non-Fed;Incl Inact/Not Class | AMA Phys Microdata File | 03/89 |
| f04602-68 | 00721-00726 | 1968 | Total M.D.'s,Total Nan-Fod | Non-Fed;Incl Inact/Not Class | AMA Dist Physicians | 12/75 |
| F04602-58 | 00727-00731 | 1958 | Total M.D. 's, Total Non-Fed | Non-Fed; Incl Inact/Not Class | AMA Phys Pop | 02/85 |
| F11215-86 | 00732-00736 | 1986 | M.D.'s,Total Ptn Care Non-Fod | Non-Fed;Patient Care | AMA Dist Physicians | 09/88 |
| F08858-85 | 00737-00741 | 1985 | M.D.'s,Total Ptn Care Non-Cod | Non-Fed;Patient Care | AMA Dist Physicians | 03/87 |
| F08858-83 | 00742-00746 | 1983 | M.D.'s,Total Ptn Care Non-Fed | Non-Fed;Patient Care | AMA Dist Physicians | 02/86 |
| F08858-82 | 00747-00751 | 1982 | M.D.'s,Total Ptn Care Non-Fed | Non-Fed;Patient Care | AMA Dist Physicians | 05/85 |
| F08858-81 | 00752-00756 | 1981 | M.D.'s,Total Ptn Car. Non-fad | Non-Fed;Patient Care | AMA Dist Physician, | 02/84 |
| F08858-80 | 00757-00761 | 1980 | M.D.'s, Total Ptn Care Non-Fed | Non-Fed;Pat iont Care | AMA Dist Physicians | 07/82 |
| F08858-79 | 00762-00766 | 1979 | M.D.'s, Total Ptn Care Non-Fed | Non-Fed;Pat ient Care | AMA Oist Physicians | 07/81 |
| F08858-78 | 00767-00771 | 1978 | M.D.'s,Total Ptn Care Non-Fed | Non-Fed;Pat ient Care | AMA Dist Physicians | 06/80 |
| F08858-77 | 00772-00776 | 1977 | M.D.'s,Total Ptn Care Non-Fed | Non-Fed;Pat ient Care | AMA Dist Physicians | 12/79 |
| F08858-76 | 00777-00781 | 1976 | M.D.'s, Total Ptn Care Non-Fed | Non-Fed;Pat ient Care | AMA Dirt Physicians | 12/79 |
| F08858-75 | 00782-00786 | 1975 | M.D.'s,Total Ptn Care Non-Fod | Non-Fed;Pat ient Care | AMA Phys Microdata fi Jo | 03/89 |
| F08858-74 | 00787-00791 | 1974 | M.D.'s,Total Ptn Care Non-Fod | Non-Fed;Pat ient Care | AMA Dist Physicians | 12/79 |
| F08858-73 | 00792-00796 | 1973 | M.D.'s,Total Ptn Care Non-Fed | Non-Fed;Pat ient Care | AMA Dist Physicians | 12/79 |
| F08858-72 | 00797-00801 | 1972 | M.D.'s,Total Ptn Care Woo-fed | Non-Fed;Pat ient Care | AMA Dirt Physicians | 12/79 |
| F08858-71 | 00802-00806 | 1971 | M.D.'s,Total Ptn Care Non-Fed | Non-Fed;Pat ient Care | AMA Dist Physicians | 12/79 |
| F08858-70 | 00807-00811 | 070 | M.D.'s,Total Ptn Care Non-Fed | Non-Fed;Pat ient Care | AMA Dist Physicians | 03/80 |
| f08858-69 | 00812-00816 | 1969 | M.D.'s,Total Pln Care Non-Fed | Non-Fed;Pat ient Care | AMA Dist Physicians | 03/80 |
| F08858-68 | 00817-00821 | 1968 | M.D.'s, lotrl Ptn Care Non-Fed | Non-Fed;Pat ient Care | AMA Dirt Physicians | 12/75 |
| F08858-58 | 00822-00826 | 1958 | M.D.'s, Total Ptn Care Non-Fed | Non-Fed;Pat ient Care | AMA Phys Pop | 02/85 |
| F04603-86 | 00827-00831 | 1986 | M.D.'s, PC, Off Based Now-Cod | Non-Fed;Paln Care Office-Based | AMA Phyr Master File | 09/88 |
| F04603-85 | 00832-00836 | 1985 | M.D.'s, PC. Off Based Non-Fed | Non-Fed;Paln Care Off ice-Based | AMA Phys Master File | 03/87 |
| F04603-83 | 00837-00841 | 1983 | M.D.'s, PC. Off Based Non-Fed | Non-Fed;Paln Care Off ice-Based | AMA Phys Master Fi lo | 02/86 |
| f04603-79 | 00842-00846 | 1979 | M.D.'s, PC, Off Bared Non-Fed | Non-Fed;Paln Care Office-Based | AMA Phys Master File | 01/82 |
| F04603-75 | 00847-00851 | 1975 | M.D.'s, PC. Off Based Non-Fed | Non-Fed;Paln Care Office-Based | AMA Phys Microdata Fi lo | 03/89 |
| F05156-79 | 00852-00856 | 1979 | MDs, Primary Care lot Non-Fed | Non-Fed,Paln Care,Office-Based | AMA Phys Master File | 01/82 |
| F05156-75 | 00857-00861 | 1975 | MDs,Primary Care lot Non-Fed | Non-Fed,Paln Care,Of f ice-Based | AMA Phys Microdata File | 03/89 |
| F05157-79 | 00862-00866 | 1979 | MDs, Not Prim Car., lot Non-Fed | Non-Fed,Paln Care,Office-Based | AMA Phys Master File | 01/82 |
| F051 57-75 | 00867-00871 | 1975 | MDs, Not Prim Care, lot Non-Fed | Non-Fed,Paln Care,Office-Based | AMA Phys Microdata File | 03/89 |
| FJ1216-86 | 00872-00875 | 1986 | MDs, let Hosp-Based PC Non-Fed | Non-Fed;Res & Ft Staff | AMA Oirt Physicians | 09/88 |

EXHIBIT 11.2

B. HEALTH PROFESSIONS

ARF USER DOCUMENTATION

B-l ) Physicians

M.D. Physicians:

1968-1983, 1985 and 1986 Total Active Non-Federal M.D.'s were obtained from "Physician Characteristics and Distribution in the U.S.", published by the Center for Health Services Research and Development of the American Medical Association in each respective year.

**NOTE: 1) Data for all years are not reported for the following county:
30113 - - Yellowstone Park, Montana

1974 Total Active Non-Federal M.D.'s were obtained from the American Medical Association County Summary Tape. The American Medical Association county codes were converted to the ARF Modified FIPS County Codes.

**NOTE: 1) Data for 1974 are not reported for the following county:
30113 - - Yellowstone Park, Montana

Estimates for 1940, 1950, and 1960 Total Active Non-Federal M.D.'s come from AMA punch cards. AMA estimates have been'used rather than Census estimates because Census estimates occasional1y included interns and residents as well as chiropractors, dentists, and veterinarians in the physician counts when enumerators failed to properly differentiate the response of "doctor". 1960 data for Alaska (02001), and New York City Counties (FIPS 36005 Bronx, 36047 Kings, 36061 New York, 36081 Queens, 36085 Richmond1 are taken from the Health Manpower Source Book, Section 10, DHHS, Public Health Service Pub. No. 263.

**NOTE: 1) 1940 and 1950 M.D.'s are not available for Alaska (02001), Weston County, Wyoming (56045), or any of the Hawaii counties (Hawaii (15001), Honolulu (15003), Kauai (15007), Maui (15009)),.

2) 1960 data for Alaska, Hawaii, and the New York City counties of Bronx, Kings, New York. Queens, and Richmond are as of the middle of the calendar year 1959.

3) 1960 data for the state of Hawaii are not broken out by county; they are all contained in Honolulu (FIPS 15003); the other Hawaii counties (FIPS 15001 Hawaii, 15007 Kauai, 15009 Maui ) are zero-filled.

4) The 1960 data for Bronx, N.Y. (FIPS 36005) are included in New York County (FIPS 36061); Bronx, N.Y. is zero-filled.

5) 1960 M.D.'s in Alaska and Hawaii include all Non-Federal M.D.'s. Active Non-Federal M.D.'s are not available for these states for 1960. 1960 M.D.'s are not available for Weston County, Wyoming (FIPS 560451.

4: Total Active Non-Federal M.D.'s for St Louis City (FIPS 29191) for 1960 are from "Health Manpower Source Book, Section 10, Physicians' Age, Type of Practice, and location", DHEW, Public Health Service, Table A, 1959.

8) 1940, 1950 and 1960 data for the independent cities of Hampton and Newport News, Virginia are included in York County (51199). Hampton (51650) and Newport News (51700) are zero-filled for these fields.

documentation. The technical documentation is a file layout, with supplemental information, while the user documentation provides information about the origins and coding of particular variables and comments on problems that arose when the data were added to the ARF. Both parts are supplied with a contents that allows the user to find the start of any series of data they are interested in. The following problems arise with respect to each part separately and in relation to each other.

User Documentation. The user documentation is very awkward, for several sources:

- Variables are not discussed in the order in which they appear in the technical documentation
- Not all variables are discussed
- Variables are not discussed under the name given in the technical documentation

Examples of varying order in the two parts of the documentation include (but are not limited to):

- Health professionals: veterinarians, optometrists, pharmacists and podiatrists are in different orders in the two parts
- Demographics: labor force, mortality, and income data are in different orders in the two parts
- Hospital data

Because not all variables are discussed in the same order in the user documentation as inthe technical documentation, users sometimes assume that there is no discussion of a particular set of data elements. This increases the possibility of misunderstanding. One solution to this problem of order (which arises presumably because the user documentation is not rewritten but

added at every update) would be to indicate in the technical documentation which variables are discussed further in the user documentation.

Another problem with the user documentation layout is that it is not always possible to tell which variables are being discussed since the variables are not referred to by unique names.

Technical Documentation. Some variables are not found where one would expect them (e.g., hospital birthing rooms are with births and not with facilities, and hospital admissions and discharges are with facilities rather than utilization). Another position problem is that variables change column position or even relative position from one file to the next depending on additions and removals of data elements. Some users found this very frustrating but it is inevitable as long as the ARF is not accessed through a database management system or a system like SAS that accesses through variable names.

The source information provided as an appendix to the documentation is inadequate. Source information should include the names and dates of publications or their machine-readable equivalent files, and should provide references to documents that describe the source data.

The characteristics column is used inconsistently, sometimes adding no new information, sometimes appearing like a code book. (See further comments under consistency below.)

A code book indicating how a variable is coded would be desirable. (The user documentation does not supply this.) For example, "short-term general hospital 006499 beds" shows four codes in the characteristics column. However, that is not how the variable appears on the file and the user documentation does not explain the codes. It would be desirable for the documentation to show the mean, maximum, and minimum values for continuous measures. It would be helpful to have a record dump as part of the documentation if a code book is not feasible.

Some users indicated that they would like to have the documentation available on diskette as well as in hard copy (and magnetic tape). This request appears to reflect the fact that if the documentation were on diskette, PC software could be used to scan it for key words, creating a cross-index. Such an electronic codebook might consist simply of a diskette, or it could be associated with available retrieval software to make it more user-friendly. A cross-index might be a useful tool, although there are alternative approaches; for example, a more detailed contents list for the technical documentation, and tables showing time series data available and inconsistencies across apparent series (such as in Table II.15). Such tables could save users time and errors.

Exhibit II.3 illustrates potential changes to the technical and user documentation suggested by users.

b. Consistency

We indicated above that there are some inconsistencies with layout, naming, and the manner in which data elements are documented. Examples of problems of inconsistency in naming and use of the characteristics column in the technical documentation include (based on the March 1989 version):

- Sometimes, the name and characteristics column have the same information; for example columns 732-736, 827-831, 50885091, and 15401-15406 all repeat the name column in the characteristics column.

- Confusion results from some entries. For example, the name for column 852-856 states "MDs, Primary Care Tot Non Federal" but the characteristics column states "Non Federal Patn Care, Office-Based." Presumably the characteristic overrides or is a subset of the name in this case.

- Abbreviations are sometimes used, sometimes inconsistently without full explanations in the user documentation. It would seem perfectly acceptable to use abbreviations if they are used consistently everywhere and included in

EXHIBIT II.3

Proposed Structure of the technical documentation:

| Column Name | Column Description |
|---|---|
| FIELD | Variable name - F-number reference |
| COL-COL | Start and end position of variable |
| YEAR OF DATA | Year of data |
| DESCRIPTION | A maximum 50 character description of the variable which will incorporate the name and characteristics columns currently present on the ARF. The DESCRIPTION column should only contain information pertinent to the subject matter of the variable. |
| USER DOC PAGE REF | The page references in the user documentation for each variable |

Proposed Structure of the user documentation:

The proposed structure of the user documentation will include three sections: a comprehensive, alphabetical listing of variables; a listing of subject matters; a listing of source files.

ALPHABETIC LISTING OF VARIABLES A complete description of the variable, including information currently found in the CHARACTERISTICS column of the technical documentation. This will also include the following aspects: the source file, the year of the data, special editing procedures, unit of measure, missing cases, and any known problems. The description of special editing procedures should contain variable specific information that might not pertain to the entire source file.

SUBJECT MATTER: A listing of related variables and a brief description providing information to the user. The brief description of a variable should aid the user in recognizing the appropriate uses of similar variables.

SOURCES A description of editing procedures, such as the merging procedure, used in processing source files. This will list all of the variables derived from the same source file.

a glossary. "PC" is used sometimes to mean "Patient Care" and sometimes to mean "Primary Care."

## 5. Suggestions for Improving the ARF and its Documentation

During the course of the interviews, individuals made a number of suggestions for improving the accessibility as well as the usability of the file and its documentation. These suggestions are presented in Table II.20. Comments relating to ARF accessibility are grouped into four categories in the table:

- Revisions to the structure of the ARF
- Approaches to handling geographical coding of the data
- Enhancements to the data elements
- General improvements in accessibility and usability.

For example, one user suggested that all population data on the file should be scaled comparably, because most population estimates are in whole numbers, but some are in hundreds. This is confusing, and it might be sensible to convert all population data to whole numbers.

## 6. Summary

Most users had no difficulty obtaining the ARF and its documentation, and most users obtained these in the basic county file format from ODAM. More than half of the users knew about the alternative formats and the source file-s but few had used them.

Users employed either the ACCESS system to ARF available at the NIH data center or to use SAS to extract data from ARF. Some HRSA users suggested that they be provided with a routine notice of updates to the file and changes in passwords. One user had a problem with a technical constraint of the ACCESS system.

TABLE II.20

SUGGESTIONS FOR IMPROVING THE USABILITY AND ACCESSIBILITY OF THE ARF AND ITS DOCUMENTATION

---

1) ARF Usability and Accessibility

a) Modifications to the structure of the ARF file:

- SAS version of the ARF.
- Conversion of the ARF and accompanying data retrieval software to CDROM format.
- A coordinated data file, incorporating the current ARF, to which all federal agencies contribute data should be created. Any user should then have access to this data set.
- Creation of an historical version of the ARF. The variables which are dropped from the ARF when a new version is released could be collected into this historical data set. Alternatively, a Cross Directory could be created describing the years for which the ARF has specific annual data, e.g. annual series 1970-1986. At present this has to be pieced together from yearly ARF files.
- Variables should remain in the same positions from version to version. Newly added variables should be placed at the end of the file or put onto a separate file. If a separate tile were created containing only updated variables, then users could purchase a basic set of variables and then the yearly updates. The yearly update file might be small enough to enable storage on diskettes.

b) Alternative approaches to handling geographic coding of the data:

- Working with the ARF and the modified FIPS codes for independent city data can lead to inconsistencies with external data sets. A trailer file should be created which gives available data for FIPS county codes only. In addition, a trailer file containing available data by MSA/non-MSA for split counties in the New England region would improve the comparability of the data.
- Due to the recoding of county codes (VA independent cities), there should be a set of county codes which are appropriate to use depending on which years of data are being used. These sets of county codes should be clearly indicated and explained in the documentation.

**TABLE Il.20** (continued)

b) Alternative approaches to handling geographic coding of data: (continued)

- SSA geographic codes should be added co make work for HCFA staff and contractors less problematic. (A crosswalk file exists at ODAM.)

c) Enhancements of the data element?

- Decrease turn around time to insure timely data. Eliminate the data elements which are out of date.
- Economic and employment data that are estimated at the county-level should be documented and flagged as potentially unreliable.
- The possibility of obtaining data from various lesser known health associations should be explored.
- A series of variables describing the same area characteristic should be consistently coded.

d) General improvements in accessibility and usability:

- The ARF should be more readily available to the user community.
- The formation of an ARF Users Group, or a publication, where users could exchange ideas and see how others are treating problematic data elements.
- Within BHPr, training classes should be developed. In addition, there should be improved documentation on the different ways in which the ARF system can be accessed and used in research by BHPr employees. One way to encourage use of the ARF within BHPr could be to have ODAM internally distribute a memo describing updated fields each time a new version of the ARF is produced.

2) ARF Documentation

- There should be an indexing system, or crosswalk, between the user and technical documentation. In addition, an index for the user documentation, specifying major categories and sub-categories of variables and where the cited variables are located, is needed in light of its cumbersome nature. If an index is not possible, then headings of major information categories should be printed at the top of each page of the technical documentation.

TABLE II.20 (continued)

2) ARF Documentation (continued)

- For some variables, a more comprehensive explanation is needed. For example, what types of deaths are included in the infant mortality rates and is the variable for sunlight hours for each month, week, or day?

- The documentation should be available on diskette as well as in paper format.

- The ARF technical documentation should include the telephone number of a contact person from each of the organizations which provide source data to the ARF. This individual would be responsible for handling technical questions concerning the data.

- A set of tables displaying summary statistics, such as minimum and maximum values, for each data element should be included in the documentation for the file.

[a] Specific data elements which could possibly be added to the file are addressed in Chapter IV.

Although most users said they understood the documentation, a number of suggestions were made for improving its usability. A thorough evaluation of the documentation uncovered some deficiencies in layout and consistency that may have contributed to users lack of knowledge about what was in the file. Finally, some users provided specific suggestions for changes to the file structure and documentation.

# III. ARF ORGANIZATION AND MANAGEMENT

## A. INTRODUCTION

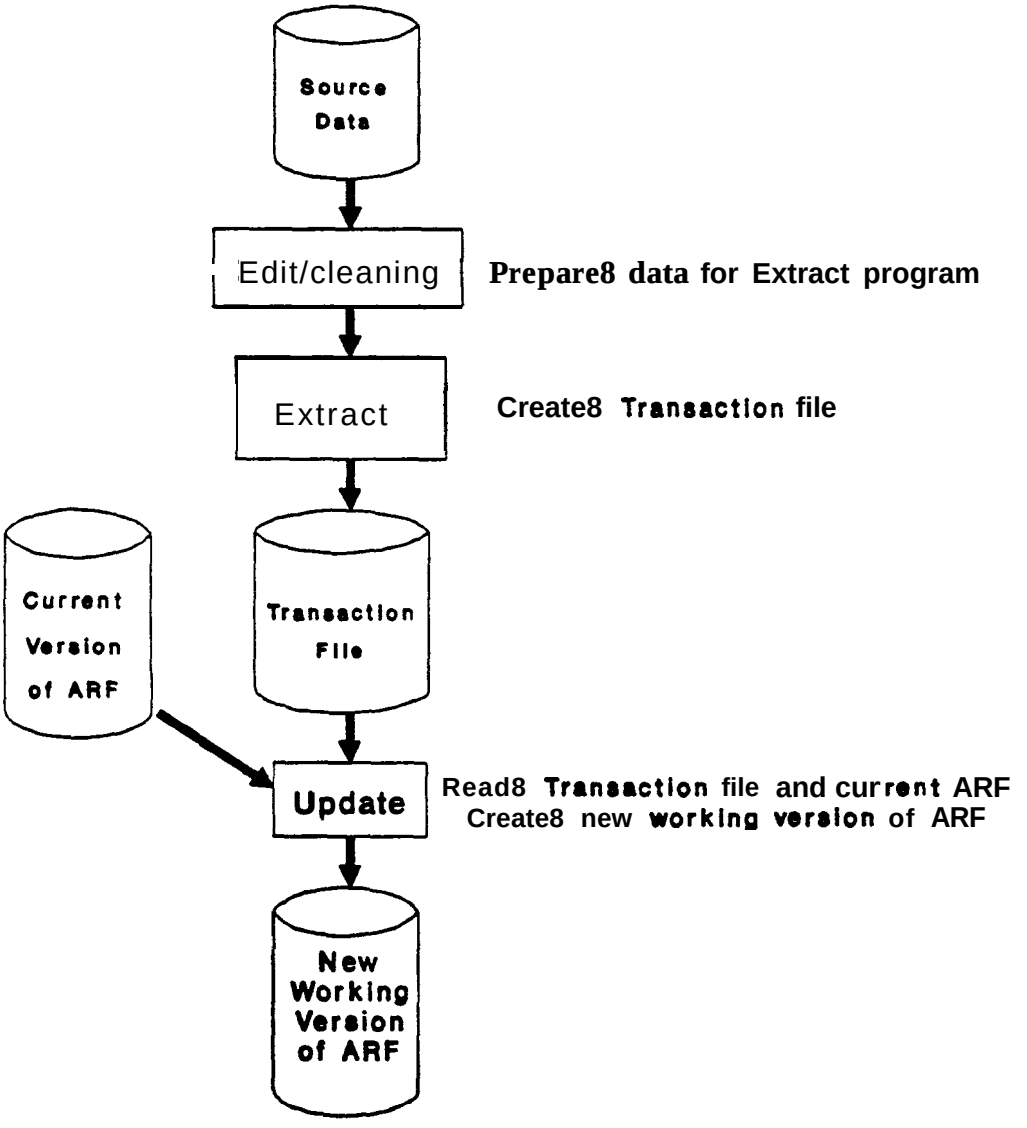
This chapter documents the organization and management procedures incorporated in the current ARF system. It is a technically-oriented survey of the current ARF system and provides a baseline with which proposed modifications to the ARF system will be compared. The next section of this chapter (section B) provides an overview of the flow of information from the source data files through the processing of the final ARF. Section C contains a detailed discussion of the components which make up the ARF system. Information management and maintenance procedures are described in section D. Section E reviews ARF computer system documentation. The next section (F), describes the ARF ACCESS system. The final section briefly discusses other files and processes not previously detailed. An appendix detailing the processing of typical source data has also been included.

## B. AN OVERVIEW OF INFORMATION FLOW THROUGH THE SYSTEM

The ARF system is a collection of many software programs which manipulate a number of source and intermediate data files in order to create the basic county-level Area Resource File. Often, a number of source-specific programs are used to clean, edit, and reformat source data. Once the data editing/preparation phase is completed, desired data for each county is retrieved from the source data file and placed in a transaction file. A final program then reads the current county-level ARF, adds data from the transaction file, and generates a new county-level ARF. Exhibit III.1 illustrates this process (Refer to Appendix E for a more technical description of the entire process).

EXHIBIT **III.1**

OVERVIEW OF TYPICAL PROCESSING




NOTE: Every rix months, the latest working version of ARF is reorganized to produce a public version of ARF.

A new county-level ARF is created each time a new source file is processed. New variables from the source are generally appended to each record in the ARF. Periodically, usually every six months, each record is reorganized in order to group similar data elements. The "organized" version of the ARF is the version available for public dissemination.

Two key programs, the EXTRACT and the UPDATE program, perform the bulk of the data processing. The discussion below provides a general overview of how these two programs integrate a typical county-level file into the basic ARF. (For many data sets, some initial processing must be performed before running these two programs. A detailed discussion of the initial processing performed on typical source files is contained in later chapters of this report.) At present, the bulk of the processing described below is performed on the IBM mainframe system available at the Computer Center of the Division of Computer Research & Technology, National Institute of Health.

1. The Extract Program

The EXTRACT program uses the source data to create a transaction file which is then used in the UPDATE program. The EXTRACT program requires three inputs:

- the source data - The extract program will correctly recode counties and Virginia independent cities as necessary.
- an edit table - The edit table is an especially important part of the process. This file describes all the variables on the ARF (descriptions include Fnumbers (variable references), year, starting position, length, description, characteristics, variable type, and source). Before running the EXTRACT program, this table must be manually edited to insert the new variables at the end. A separate program, the RECALC program, is used to determine all the start positions.
- a "card" file - The "card" data is a list of the Fnumbers to be added to the ARF.

The EXTRACT program uses these three input files to perform the following tasks:

1. Sorts the edit table by starting position.
2. Checks this new edit table for any errors.
3. Sorts the edit table by Fnumber.
4. Uses the COBOL program PREEXT to reformat records from the edit table to be used in a later step.
5. Recodes county codes in the source file to the modified FIPS codes. This step will determine which of the unusual counties have to be recoded (Virginia independent cities, Alaska, Lapaz, Cibola, Yuma, Valencia, etc).
6. Sorts the output from step 5 by modified FIPS code.
7. Defines the file layout of the source data that will be added.
8. Creates a transaction file by using the sorted source data, the modified edit table, and the specified Fnumbers.

This transaction file includes one record per county per variable to be changed (i.e., the program writes out one transaction record for each variable for each county that is to be modified or added). The transaction file describes exactly which records to update and which positions in each record to modify. Each transaction file record contains the modified FIPS code, the start position, the length, and the actual data value.

2. <u>The Undate Program</u>

The UPDATE program reads each record from the transaction file created by the EXTRACT program, updates the corresponding record on the current ARF, and generates a new ARF. By changing a parameter in UPDATE, the programmer may control the listings produced by the program. UPDATE generates either a listing of all the counties that matched or a listing

of all the counties that did not match. The appropriate type of listing depends on which source is being processed.

3. Other Programs

In addition to these two programs, various quality control programs are utilized. Some of the programs simply produce a listing of records from the original source to double check the file layout and determine the manner in which problem counties have been handled (i.e. Alaska, Virginia independent cities, etc). Other programs are used to confirm that state totals from original source data match state totals derived from the newly updated ARF. These various quality control programs will be covered in more detail in Section D.

This processing (the execution of the EXTRACT and UPDATE programs) is performed frequently, once for each time a source data set becomes available. As data from each source is added to the ARF, new variables are placed at the end of each record. Every six months, the ARF contractor runs a rearrange program (REARR) which reads and writes the ARF, changing the order of the variables so that variables are grouped by subject area. The product of this program is the current basic ARF for public use.

## C. DETAILED DESCRIPTION OF SYSTEM COMPONENTS

This section describes the ARF system components. Since the ARF system is a very large and complex collection of programs and data files which integrate data from over fifty separate sources, we have concentrated on the components used to process three typical examples of source data:

- The 1986 Local Area Personal Income (LAPI) data - a small, county-level file requiring minimal pre-processing/cleaning prior to inclusion in the ARF
- The 1987 MEDICARE data - a county-level file requiring some pre-processing

- The 1987 American Hospital Association (AHA) Hospital data - a large, facility-level file requiring substantial pre-processing

The descriptions of the system components are contained in the following: 1) a series of flowcharts, 2) a list of main system components and processes, and 3) a list of component inputs and outputs. Subsection 4 will help the reader understand the relationship between these three sections by describing, step by step, the processing of the LAPI data.

1. Flowcharts

Exhibits III.2A through III.2C provide a pictorial representation of the data flow for the three selected source files. Exhibit III.2A depicts the flow of the LAPI data, Exhibit III.2B the Medicare data, and Exhibit III.2C pertains to the AHA Hospital data. (Note that the first four pages of the Hospital data, Exhibit III.2C, cover the creation of the Hospital Facility-Level File).

The shapes used in the flowcharts are standard data processing flowchart symbols representing tapes, programs, disk files, printouts, and "card" input. Circles with a single letter are used to depict a continuation on a preceding or subsequent page. For each of the charts, the main flow of the data, from the original source to the final product, is in a straight line down the middle of the chart. Generally, test printouts are depicted on the right side and additional inputs and processing are shown on the left.

2. System Components and Processes

Tables III.1A through III.1C list all the main system components and processes, both manual and automated. The first column provides a name for the component, either the software program name (in UPPERCASE letters) or a general name. In many cases, a single component includes a number of similar programs. For instance, after many of the major

EXHIBIT III.2a

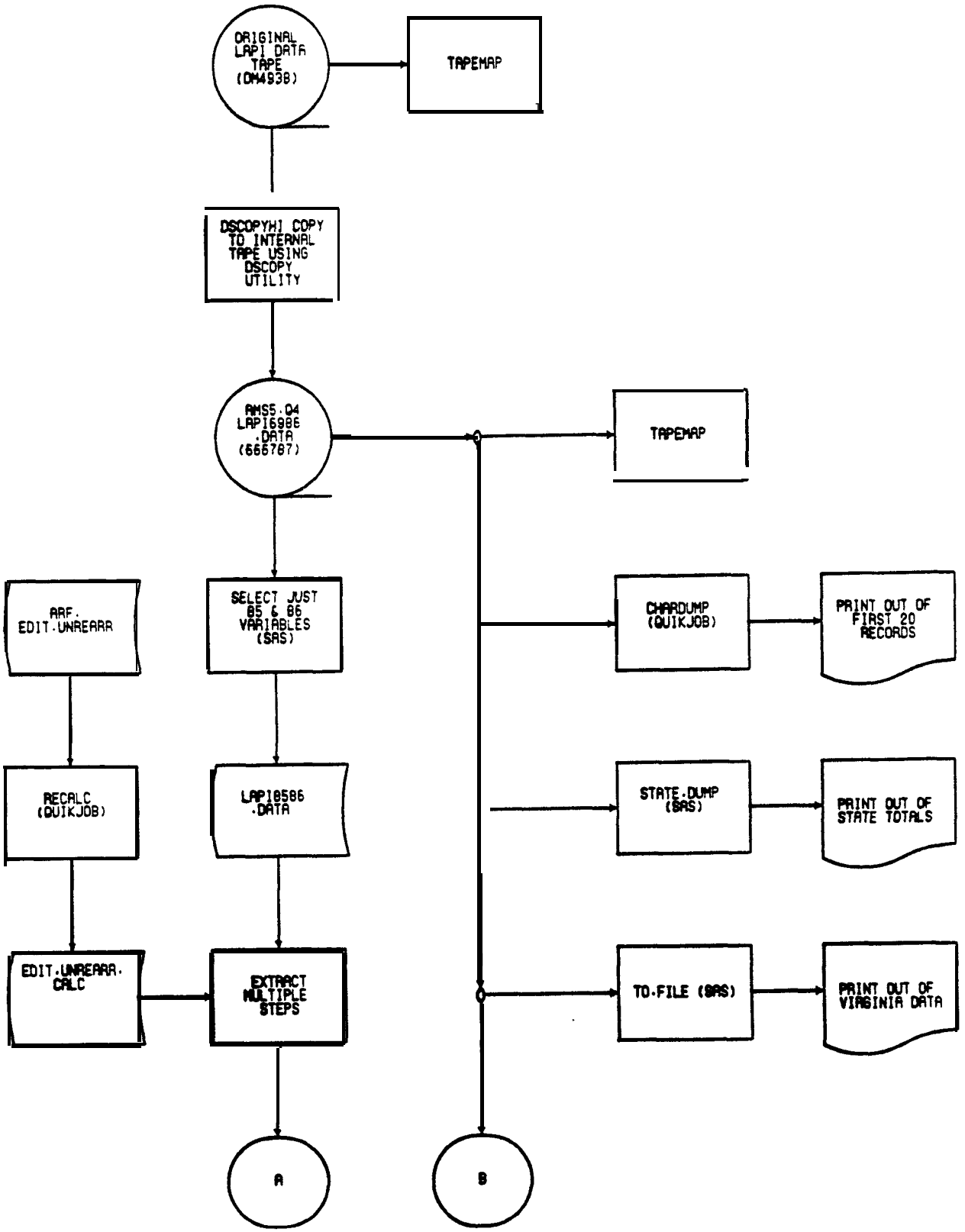


EXHIBIT III.2a - PAGE 2

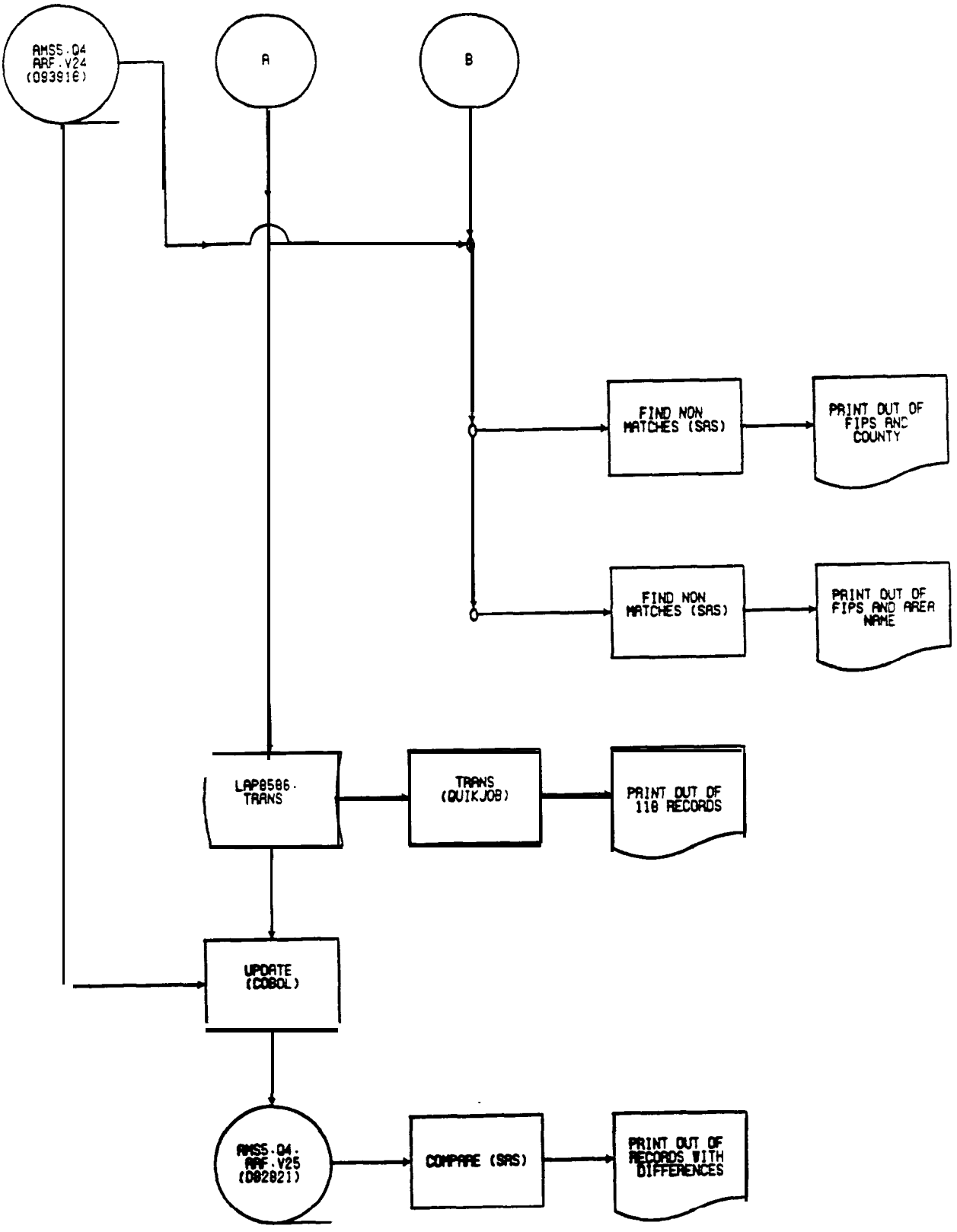


EXHIBIT III.2b

1987 MEDICARE DATA

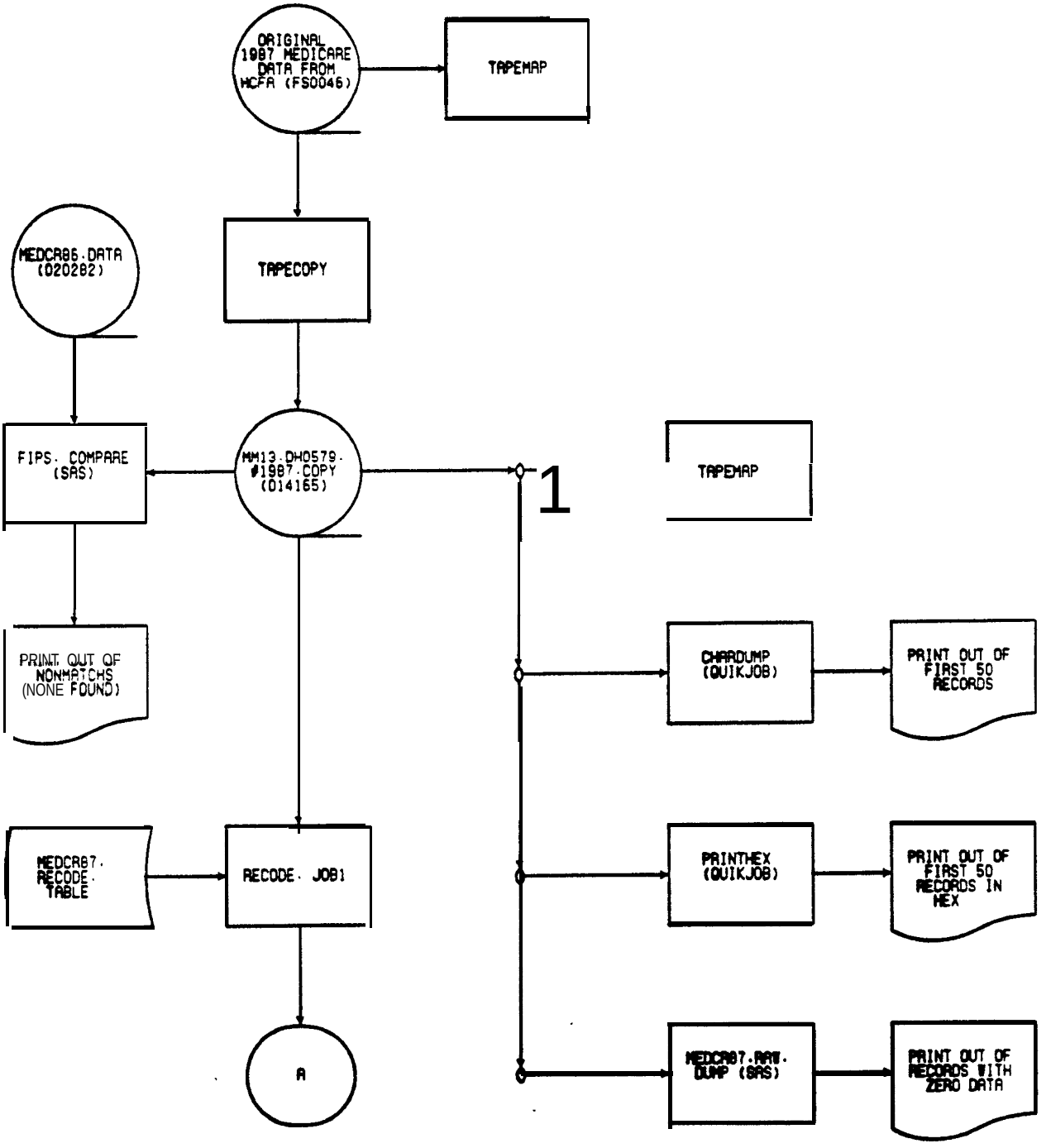
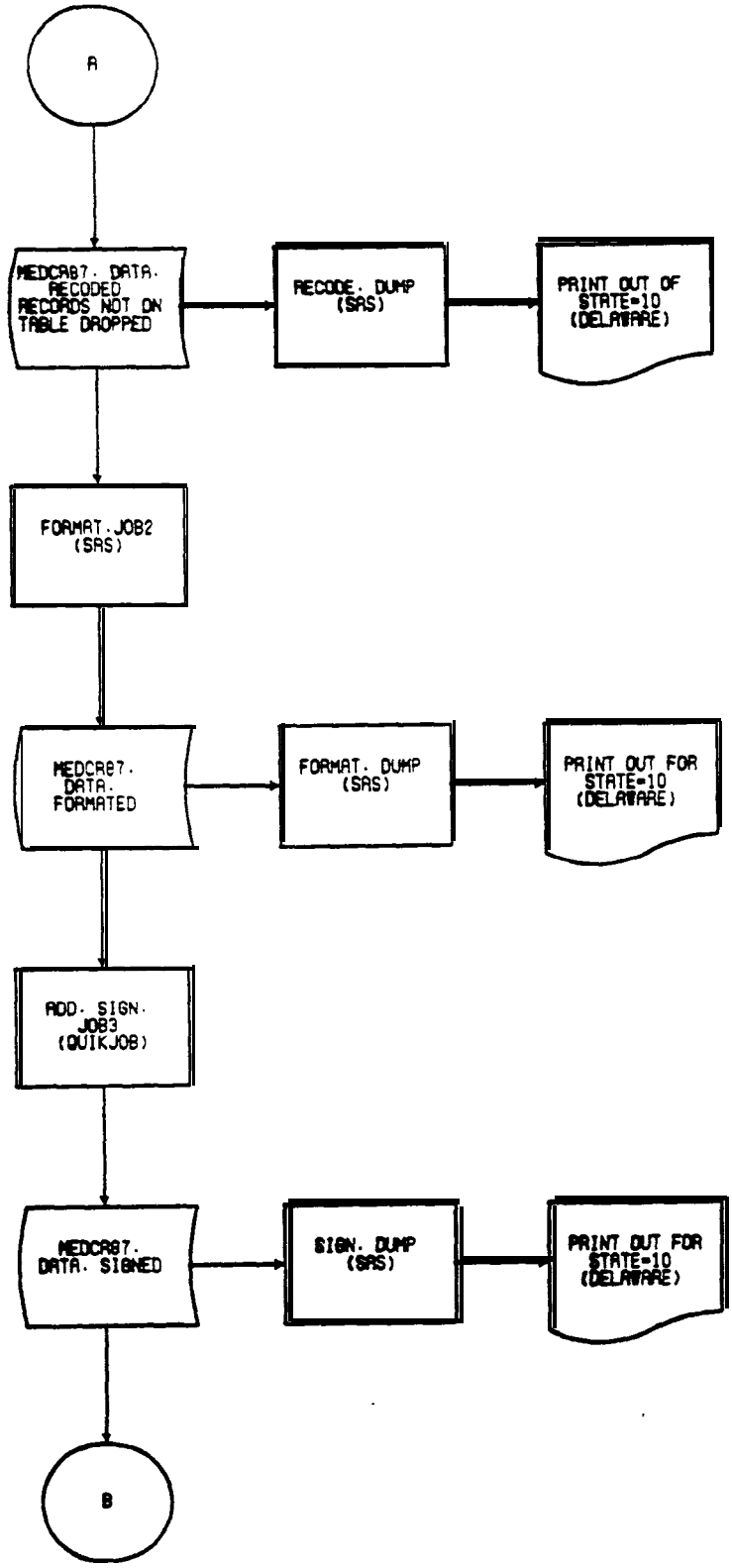


EXHIBIT III.2b - PAGE 2



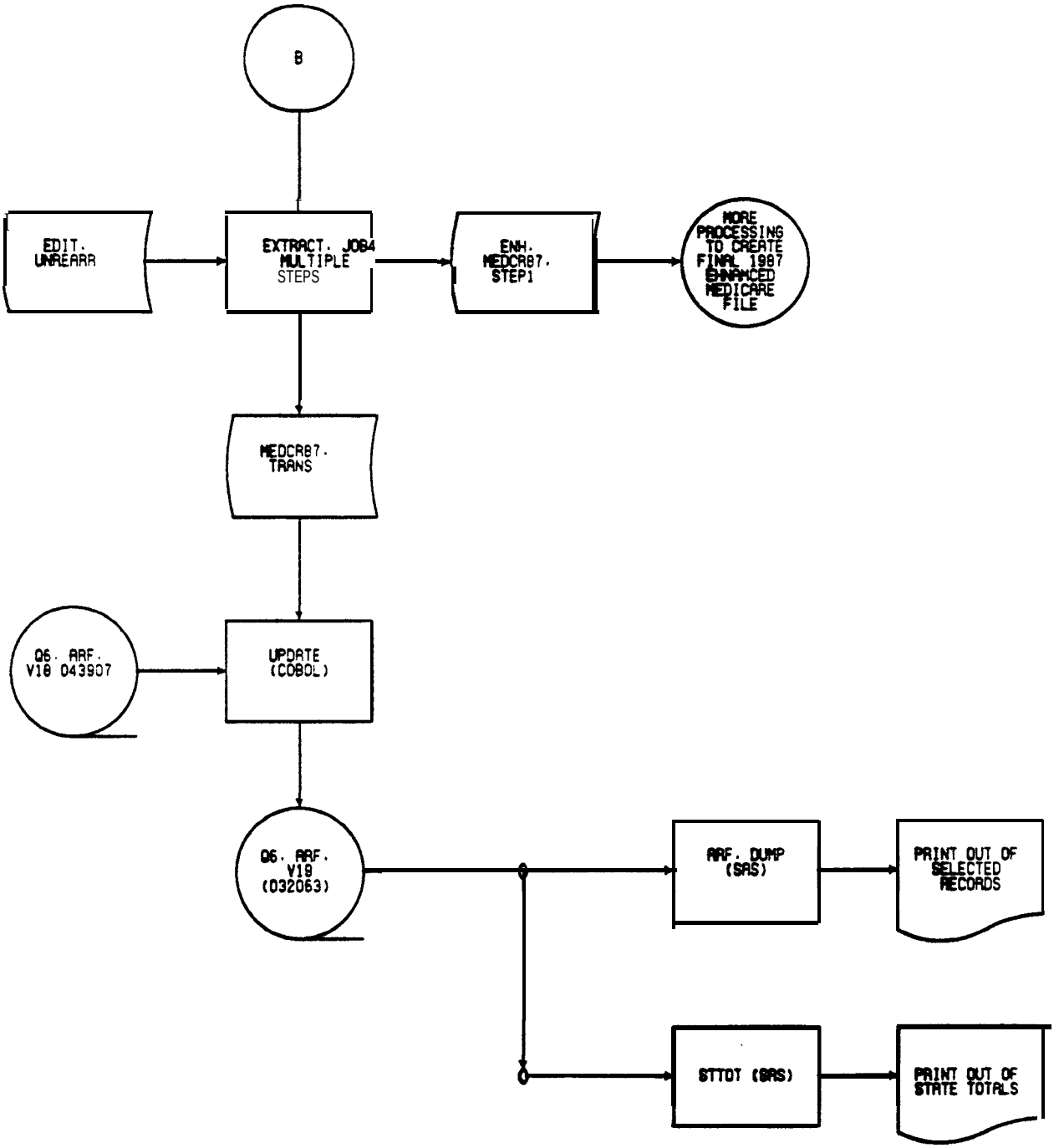

B
EDIT. UNREARR
EXTRACT. JOB4 MULTIPLE STEPS
ENH. MEDCR87. STEP1
MORE PROCESSING TO CREATE FINAL 1987 ENHANCED MEDICARE FILE
MEDCR87. TRANS
Q6. ARF. V18 043907
UPDATE (COBOL)
Q6. ARF. V19 (032063)
ARF. DUMP (SAS)
PRINT OUT OF SELECTED RECORDS
STTOT (SAS)
PRINT OUT OF STATE TOTALS

EXHIBIT III.2c

AHA 1987 HOSPITAL DATA

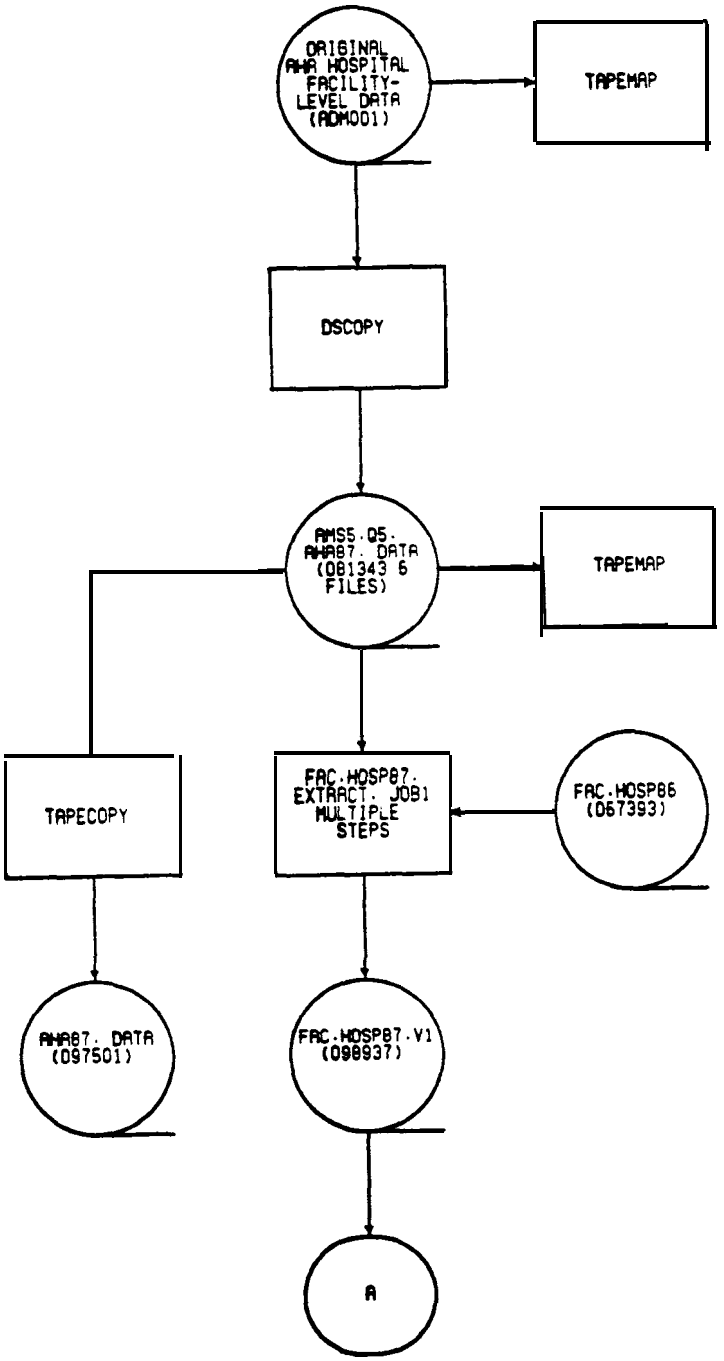
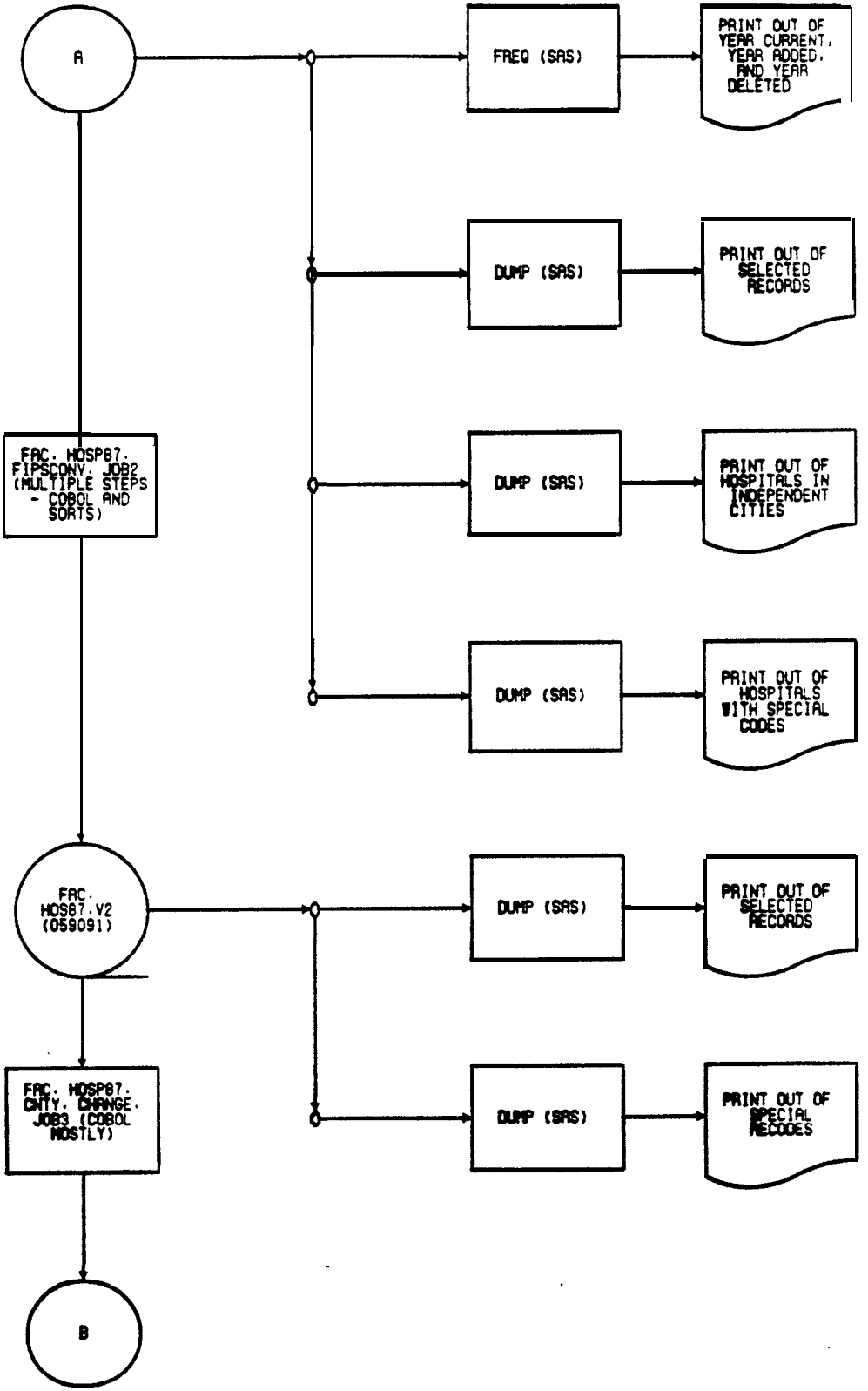
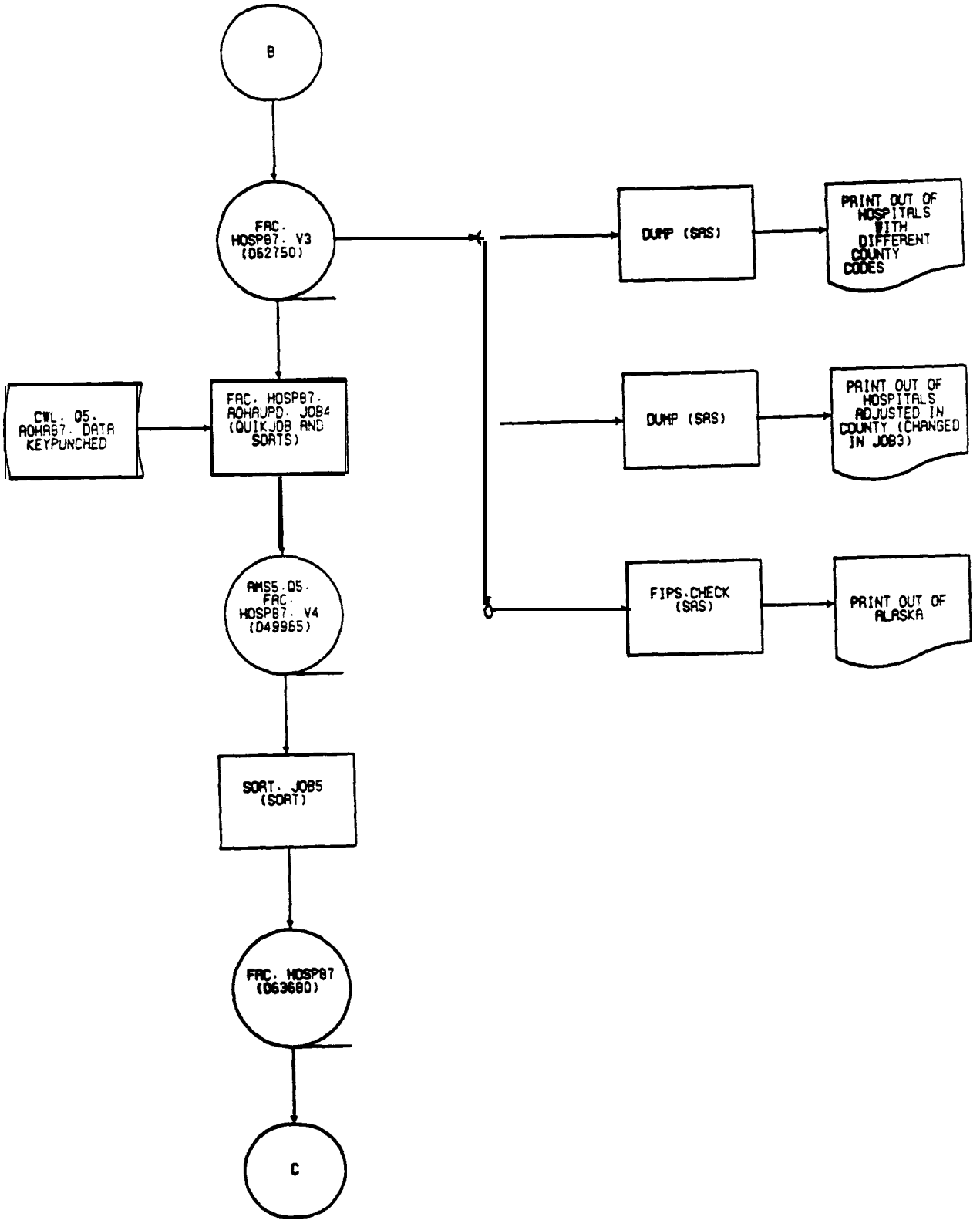


EXHIBIT III.2c - PAGE 2




B
FAC. HOSP87. V3 (D62750)
CWL. 05. AOHA87. DATA KEYPUNCHED
FAC. HOSP87. AOHAUPD. JOB4 (QUIKJOB AND SORTS)
AMS5.05. FAC. HOSP87. V4 (D49965)
SORT. JOB5 (SORT)
FAC. HOSP87 (D63680)
C
DUMP (SAS)
PRINT OUT OF HOSPITALS WITH DIFFERENT COUNTY CODES
DUMP (SAS)
PRINT OUT OF HOSPITALS ADJUSTED IN COUNTY (CHANGED IN JOB3)
FIPS.CHECK (SAS)
PRINT OUT OF ALASKA

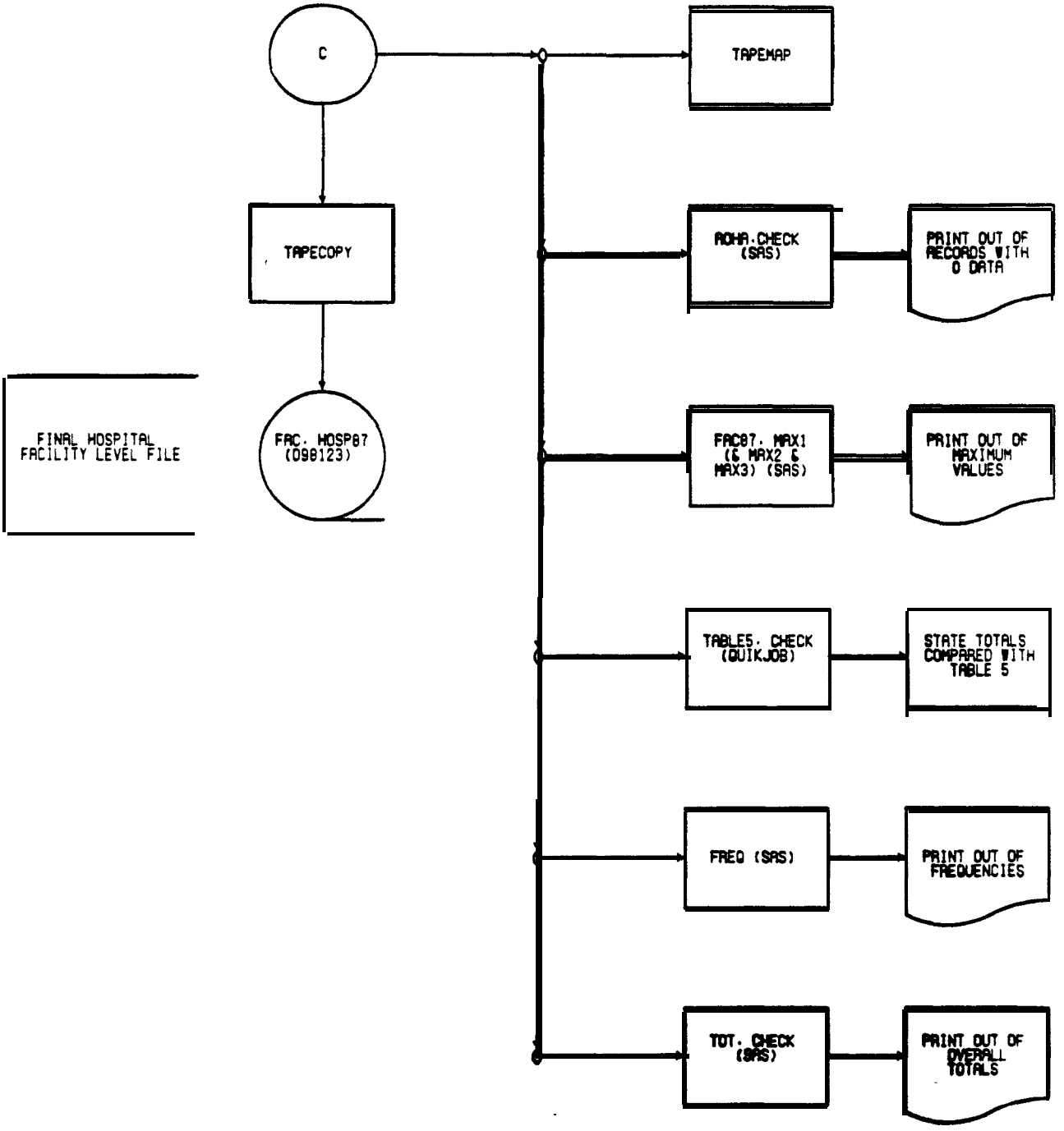
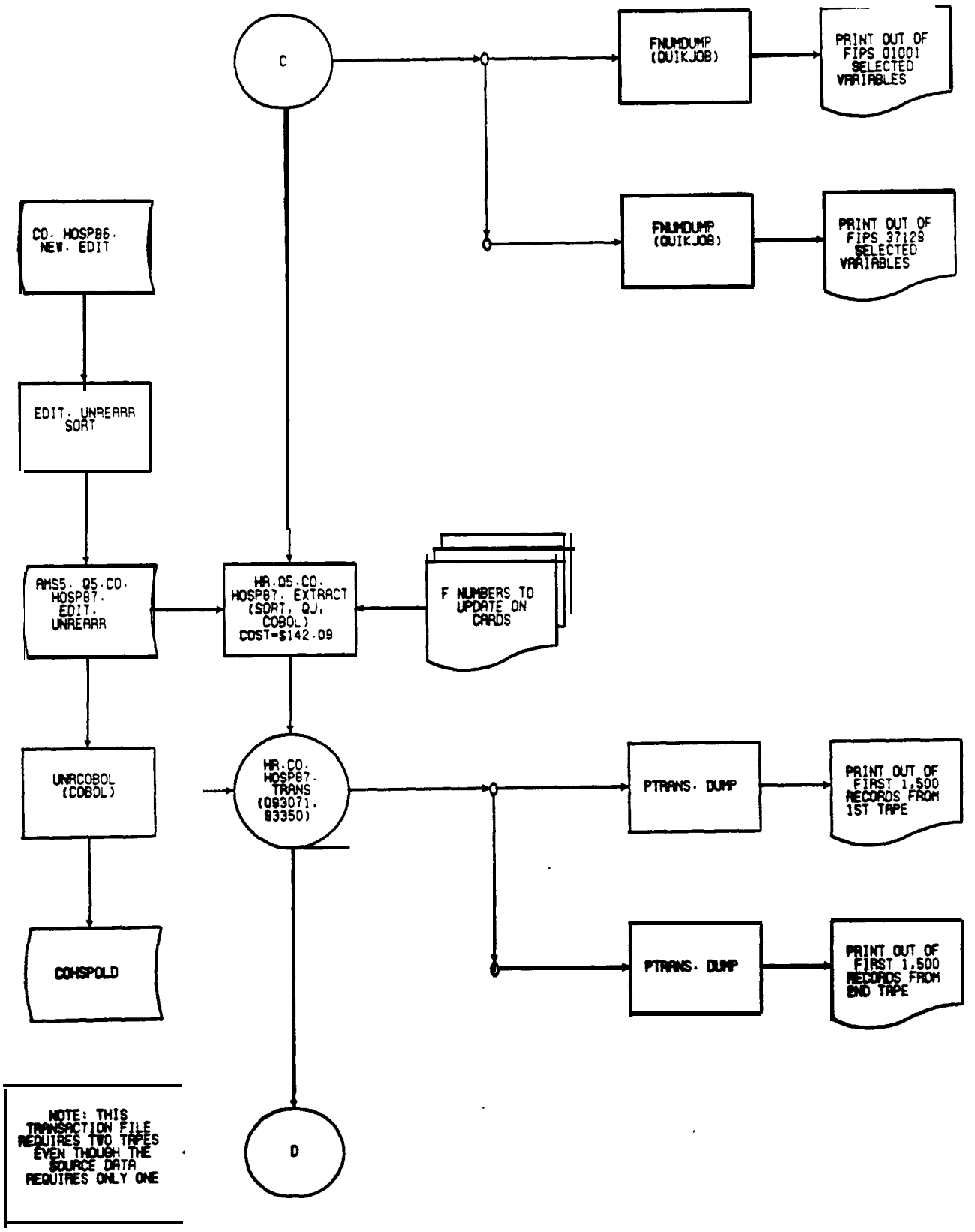
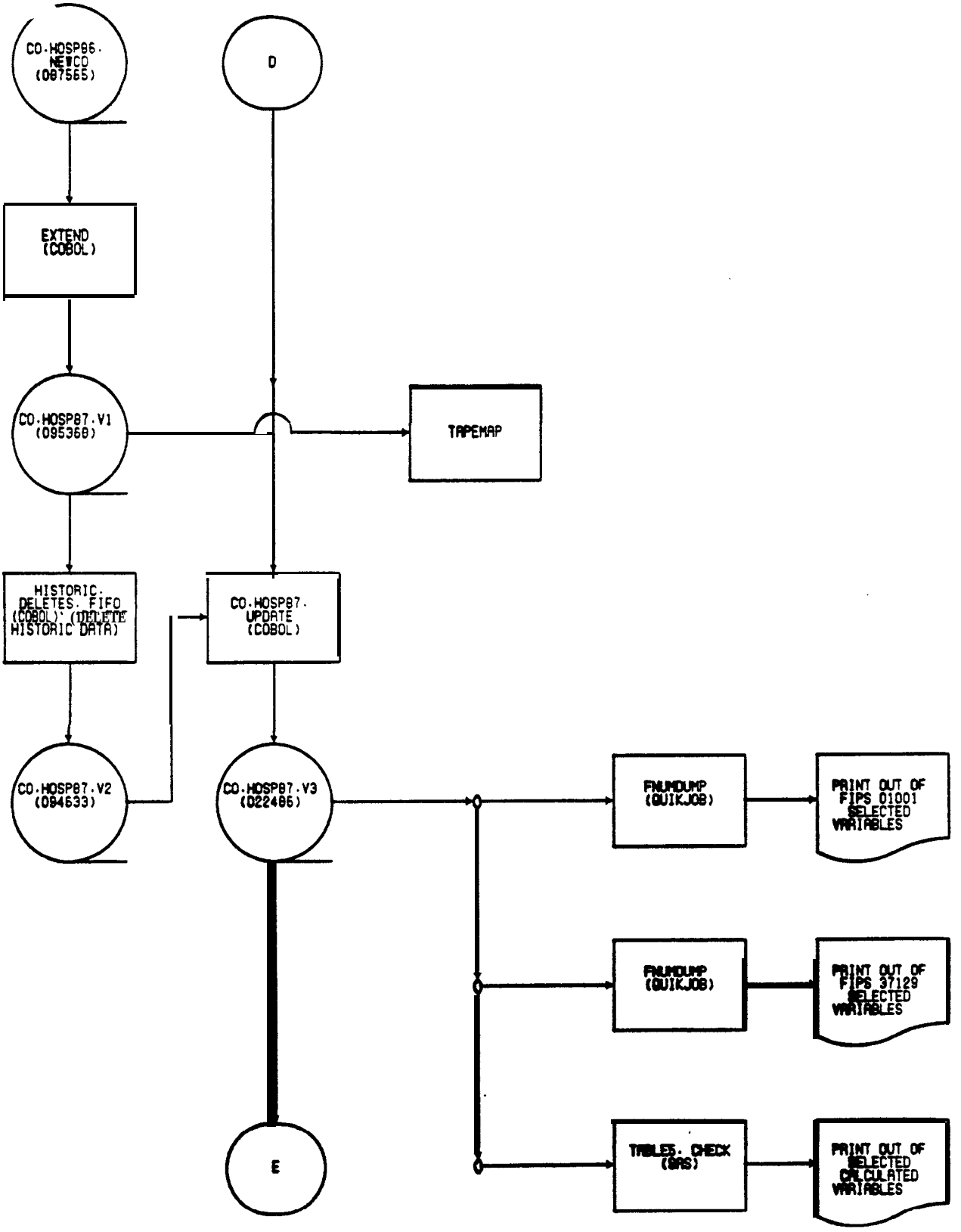
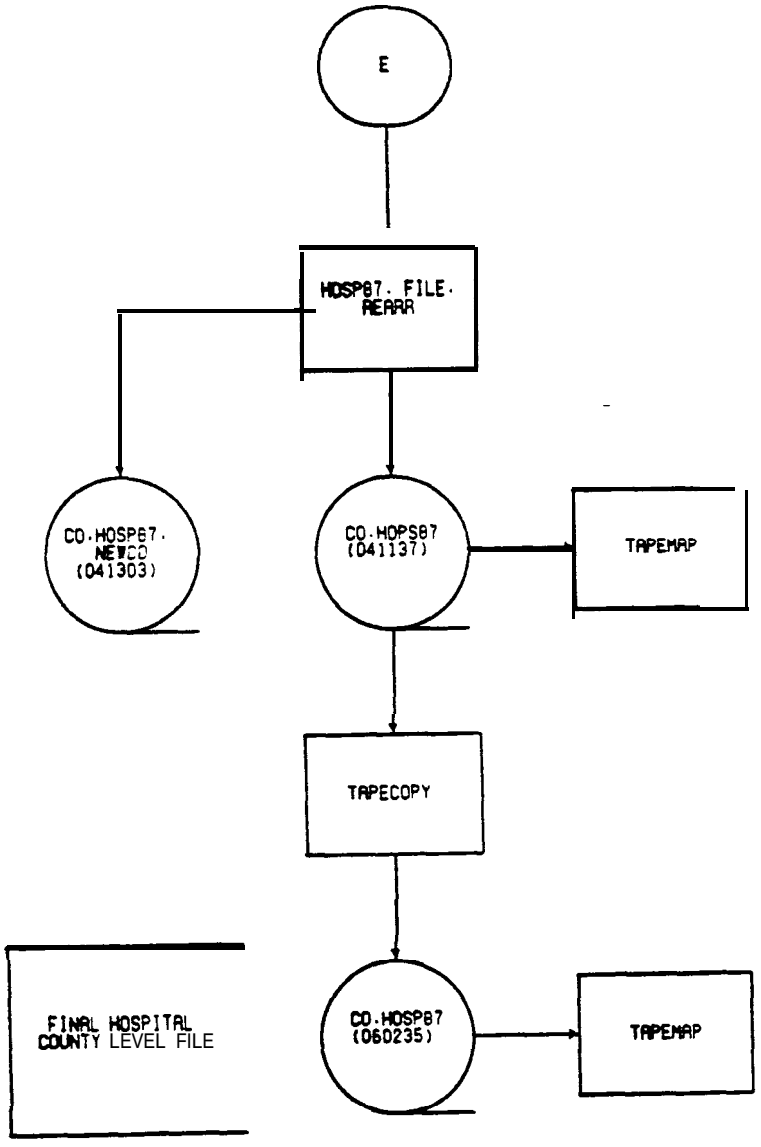

C
TAPEMAP
TAPECOPY
ROHA.CHECK (SAS)
PRINT OUT OF RECORDS WITH 0 DATA
FINAL HOSPITAL FACILITY LEVEL FILE
FAC. HOSP87 (090123)
FAC87. MAX1 (& MAX2 & MAX3) (SAS)
PRINT OUT OF MAXIMUM VALUES
TABLE5. CHECK (QUIKJOB)
STATE TOTALS COMPARED WITH TABLE 5
FREQ (SAS)
PRINT OUT OF FREQUENCIES
TOT. CHECK (SAS)
PRINT OUT OF OVERALL TOTALS

EXHIBIT III.2c - PAGE 5




CO.HOSP86.
NEWCO
(087565)
D
EXTEND
(COBOL)
CO.HOSP87.V1
(095368)
TAPEMAP
HISTORIC.
DELETES. FIFO
(COBOL) (DELETE
HISTORIC DATA)
CO.HOSP87.
UPDATE
(COBOL)
CO.HOSP87.V2
(094633)
CO.HOSP87.V3
(022486)
FNUMDUMP
(QUIKJOB)
PRINT OUT OF
FIPS 01001
SELECTED
VARIABLES
FNUMDUMP
(QUIKJOB)
PRINT OUT OF
FIPS 37129
SELECTED
VARIABLES
TABLES. CHECK
(SAS)
PRINT OUT OF
SELECTED
CALCULATED
VARIABLES
E


E
HOSP87. FILE. REARR
CO.HOSP87. NEWCO (041303)
CO.HOPS87 (041137)
TAPEMAP
TAPECOPY
FINAL HOSPITAL COUNTY LEVEL FILE
CO.HOSP87 (060235)
TAPEMAP

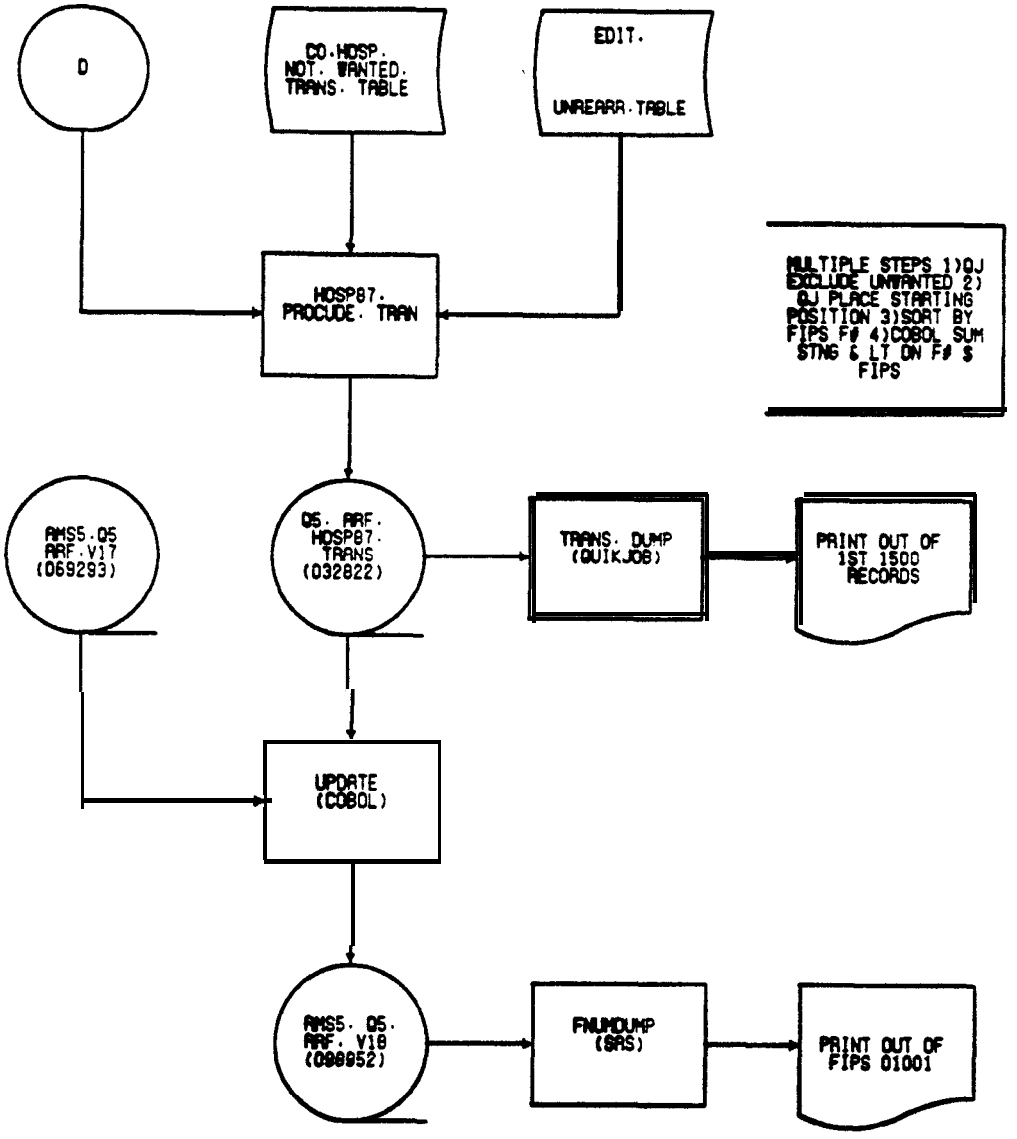

D
CO.HDSP.
NOT. WANTED.
TRANS. TABLE
EDIT.
UNREARR.TABLE
HDSP87.
PRODUCE. TRAN
MULTIPLE STEPS 1)OJ
EXCLUDE UNWANTED 2)
OJ PLACE STARTING
POSITION 3)SORT BY
FIPS F# 4)COBOL SUM
STNG & LT ON F# $
FIPS
AMS5.05
ARF.V17
(069293)
05. ARF.
HDSP87.
TRANS
(032822)
TRANS. DUMP
(QUIKJOB)
PRINT OUT OF
1ST 1500
RECORDS
UPDATE
(COBOL)
AMS5. 05.
ARF. V18
(098952)
FNUMDUMP
(SAS)
PRINT OUT OF
FIPS 01001

TABLE III.1A

MAIN SYSTEM COMPONENTS AND PROCESSES

| Main System Components/Processes | Description |
|---|---|
| For LAPI data: | |
| 1. procure | obtain tape from Bureau of Economic Analysis |
| 2. duplicate | run TAPEMAP and make copy of original tape |
| 3. display | SAS and QUIKJOB programs to dump various records to become more familiar with the data |
| 4. merge | SAS programs to run some test merges to catch possible problems |
| 5. SELECT | SAS program to select desired variables |
| 6. edit | manually edit table to include variables to be added |
| 7. RECALC | QUIKJOB program to recalculate start and end positions on edit table |
| 8. EXTRACT | multi-step program to extract data and create a transaction file |
| 9. TRAN | QUIKJOB program to list first 100 records of transaction file |
| 10. UPDATE | COBOL program to use transaction file to update current version of ARF |
| 11. COMPARE | SAS program which displays records with differences for two population variables |
| 12. review | review all programs and document the process |

TABLE III.1B

MAIN SYSTEM COMPONENTS AND PROCESSES

| Main System Components/Processes | Description |
|---|---|
| For Medicare data: | |
| 1. procure | obtain tape from HCFA |
| 2. duplicate | run TAPEMAP and make copy of original tape |
| 3. display 1 | SAS and QUIKJOB programs to dump various records to become more familiar with the data |
| 4. FIPS.COMPARE | SAS programs to find unexpected FIPS codes |
| 5. edit 1 | make sure MECCA87.RECODE.TABLE is up to date |
| 6. RECODE. JOB 1 | QUIKJOB to recode HCFA state and county codes to FIPS codes based on MEDCA87.RECODE.TABLE |
| 7. RECODE.DUMP | SAS program to print out selected recoded records |
| 8. FORMAT.JOB2 | SAS program to reformat data |
| 9. FORMAT.DUMP | SAS program to print out selected reformatted records |
| 10. SIGN. JOB3 | QUIKJOB program to create numbers with leading signs |
| 11. SIGN.DUMP | SAS program to print out selected signed records |
| 12. edit 2 | modify by hand edit table to include variables to be added |
| 13. RECALC | QUIKJOB program to recalculate start and end positions on edit table |
| 14. EXTRACT | multi-step program to extract data and create a transaction file |
| 15. UPDATE | COBOL program to use transaction file to update current version of ARF |
| 16. display 2 | SAS programs to print selected records and state totals |
| 17. review | review all programs and document the process |

TABLE III.1C

MAIN SYSTEM COMPONENTS AND PROCESSES

| Main System Components/Processes | Description |
|---|---|
| For AHA hospital data: | |
| 1. procure | obtain tape from AHA |
| 2. duplicate 1 | run TAPEMAP and make copy of original tape |
| 3. EXTRACT 1 | multi-step program to extract data at the hospital facility level and merge to the previous version |
| 4. display 1 | numerous SAS jobs to view select records more carefully |
| 5 . FIPSCONV | multi-step program to convert to FIPS codes |
| 6. display 2 | SAS programs to print out various records |
| 7. CHANGE | COBOL program to account for hospitals which "moved" |
| 8. display 3 | SAS programs to print out various records |
| 9. enter | keypunch AOHA data |
| 10. AOHAUPD | QUIKJOB program to merge AOHA data to hospital facility level data |
| 11. SORT | sort the hospital facility level data |
| 12. display 4 | numerous SAS and QUIKJOB programs to insure the integrity of the data |
| 13. duplicate 2 | copy final facility-level file |
| 14. edit | edit county-level edit table |
| 15. UNREARR.SORT | sort county-level edit table |
| 16. EXTRACT 2 | multi-step program to extract data from file and create transaction file |
| 17. display 5 | programs to print out various records |

**TABLE III.1C-** (continued)

| Main System Components/Processes | Description |
|---|---|
| 18. **EXTEND** | **COBOL** program to increase record length of county-level file |
| 19. **DELETES.FIFO** | **COBOL** program to delete data from county-level file |
| 20. **UPDATE** 1 | **COBOL** program to use transaction file to update county-level hospital file |
| 21. display 6 | SAS and QUIKJOB programs to print out various records |
| 22. REARR | program to rearrange variables on county-level file |
| 23. duplicate 3 | copy final county-level file |
| 24. PROCUDE | COBOL program which uses county-level transaction file to create a ARF transaction file |
| 25. display 7 | QUIKJOB program to print out various records |
| 26. UPDATE 2 | COBOL program to use transaction file to update current version of ARF |
| 27. display 8 | SAS program to print out various records |
| 28. review | review all programs and document the process |

components, we list display as a component. This display component is actually composed of various programs to view the data and verify that the previous step worked.

The next column, Description, details both the purpose and the nature of the component. Many of the same components are repeated on all the source data sets. Naturally, the processing of each of the three data sets, the LAPI, the Medicare, and the AHA, begins with the procurement of a tape and the duplication of this tape. Similarly, the final review componentincludes documenting and filing the listings of the computer runs. In addition, a TAPEMAP component, a computer program to list the contents of a tape, is performed on original source tapes and final, deliverable tapes for documentation.

3. Components' Inputs and Outputs

Tables III.2A through III.2C describe all the data sets used in the processing of the three selected source files, listing the components for which they are inputs and outputs. Note that many of the data files serve as input to two or three of the components. For instance, the copy of the original LAP1 data, tape number 666787, serves as input to three components (3. display, 4. merge, and 5. SELECT) as shown in Table III.2A.

4. Example of Component Interrelationships: LAP1 Data

To understand the relationship among the components listed in Table III.2A, the reader will have to refer to both Exhibit III.2A and Table III.1A. To facilitate this process, the following text describes the links among these components for the LAP1 data.

The first component is to obtain the LAPI data tape. Next, this tape is analyzed (depicted by the TAPEMAP box on the flowchart) and duplicated (depicted by the box below the tape).

TABLE III.2A

COMPONENT INPUTS AND OUTPUTS

| Description | VOL/DSN | Output From[1] | Input To[1] |
|---|---|---|---|
| For LAPI data: | | | |
| original LAPI tape | 0M4938 | 1. procure | 2. duplicate |
| LAP1 data | 666787 | 2. duplicate | 3. display |
| | | | 4. merge |
| | | | 5. SELECT |
| ARF old ver. | 093918 | previously2 | 4. merge |
| | | | 9. UPDATE |
| edit table | EDIT.UNREARR | previously2 | 7. RECALC |
| edit table | EDIT.CALC | 7. RECALC | 8. EXTRACT |
| selected LAP1 data | LAP18586 | 5. SELECT | 8. EXTRACT |
| trans. file | LAPI8586.TRAN | 9. EXTRACT | 10. UPDATE |
| | | | 11. COMPARE |
| ARF new ver. | 082821 | 10. UPDATE | 12. review |

'Numbers refer to component numbers on Tables III.1A

*Data and programs used to create most recent working version of ARF.

TABLE III.2B

COMPONENT INPUTS AND OUTPUTS

| Description | VOL/DSN | Output From' | Input To' |
|---|---|---|---|
| For MEDICARE Data: | | | |
| original 87 Medicare | F'S0046 | 1. procure | 2. duplicate |
| Medcr 87 data | 014165 | 2. duplicate | 3. display 1 |
| | | | 4. FIPS.COMPARE |
| | | | 6. RECODE.JOB1 |
| recode table | RECODE.TABLE | 5. edit 1 | 6. RECODE.JOB1 |
| recoded data | DATA.RECODE | 6. RECODE.JOBl | 7. RECODE.DUMP |
| | | | 8. FORMAT.JOB2 |
| formatted data | DATA.FORMATTED | 8. FORMAT.JOB2 | 9. FORMAT.DUMP |
| | | | 10. SIGN.JOB3 |
| signed data | DATASIGNED | 10. SIGN.JOB3 | 11. SIGN.DUMP |
| | | | 14. EXTRACT |
| edit tab. | EDIT.UNREARR | previously[2] | 12. edit 2 |
| edit table | EDIT.UNREARR | 12. edit 2 | 13. RECALC |
| edit table | EDIT.CALC | 13. RECALC | 14. EXTRACT |
| transaction file | MEDCA87.TRAN | 14. EXTRACT | 15. UPDATE |
| enhanced file | EHN.MEDCR87.STP1 | 14. EXTRACT | subsequently |
| ARF old version | 043907 | previously[2] | 15. UPDATE |
| ARF new version | 032063 | 15. UPDATE | 16. display 2 |
| | | | 17. review |

'Numbers refer to component numbers on Tables III.1B.

[2]Data and programs used to create most recent working version of ARF.

TABLE III.2C
COMPONENT INPUTS AND OUTPUTS

| Description | VOL/DSN | Output From' | Input To[1] |
|---|---|---|---|
| For AHA Hospital Data | | | |
| original AHA tape | ADM001 | 1. procurement | 2. duplicate 1 |
| AHA data | 081343 | 2. duplicate 1 | 3. EXTRACT 1 |
| fat.-level prev. ver | 067393 | previously2 | 3. EXTRACT 1 |
| facility-level ver. 1 | 098937 | 3. EXTRACT | 4. display 1<br>5. FIPSCONV |
| facility-level ver. 2 | 059091 | 5. FIPSCONV | 6. display 2<br>7. CHANGE |
| facility-level ver. 3 | 062750 | 7. CHANGE | 8. display 3<br>10. AOHAUPD |
| AOHA data | AOHA87 | 9. enter | 10. AOHAUPD |
| facility-level ver. 4 | 049965 | 10. AOHAUPD | 11. SORT |
| facility-level final | 063680 | 11. SORT | 12. display 4<br>13. duplicate 2 |
| facility-level backup | 098123 | 13. duplicate 2 | |
| prev. ver. edit table | HOSP86.NEW.EDIT | previously2 | 9. UNREARR.SORT |
| new ver. edit table | HOSP87.EDIT.UNRE | 15. UNREARR.SORT | 16. EXTRACT 2 |
| county trans. file | 093071,093350 | 16. EXTRACT 2 | 17. display 5<br>18. UPDATE<br>24. PROCUDE |
| old county-level | 087565 | previously2 | 18. EXTEND |
| county-level ver. 1 | 095368 | 18. EXTEND | 19. DELETES.FIFO |
| county-level ver. 2 | 094633 | 19. DELETESFIFO | 20. UPDATE 1 |
| county-level unrearr | 022486 | 20. UPDATE 1 | 21. display 6<br>22. REARR |
| county-level rearr | 041137 | 22. REARR | 23. duplicate 3 |
| county-level new | 041303 | 22. REARR | subsequent |
| county-level backup | 060235 | 23. duplicate 3 | |
| unwanted records | NOT.WANTED | previously2 | 24. PROCUDE |
| ARF edit table | EDIT.UNREARR | previously[2] | 24. PROCUDE |
| ARF trans. file | 032822 | 24. PROCUDE | 25. display 7<br>26. UPDATE 2 |
| ARF pervious ver. | 069293 | previously2 | 26. UPDATE 2 |
| ARF new version | 098952 | 26. UPDATE 2<br>28. review | 27. display 8 |

'Numbers refer to component numbers on Table III.1C.
[2]Data and programs used to create most recent working version of ARF and/or most recent hospital facility-level file.

This copy of the tape is used as input for the next three components. The display component is illustrated by the four program boxes on the right of the first page of the flowchart (Exhibit III.2A). Computer languages' are shown in parentheses. Similarly, the merge component, which also uses the previous version of the ARF as input, is shown on the right side of the second page. These two components both output printout. The SELECT program outputs the LAPI8586.DATA disk file. This file is in turn used as input to the EXTRACT program. As shown on the flow chart, this EXTRACT program also uses EDIT.UNREARR.CALC (which was first manually edited and then run through the RECALC program). The output from the EXTRACT program is a transaction file which will be used as input by the UPDATE program. After the successful execution of this UPDATE program (and the TRAN program to insure the transaction file was created correctly), the new version of the ARF is nearly complete. The last two components, the program COMPARE and the review of the entire process, are required to guarantee that this new file is complete and error-free.

5. Processing Costs

The ARF includes data from over fifty different sources (or hundreds if yearly files from a particular source are counted separately). Anywhere from twenty to forty versions of the ARF are created each half-year period since every time data from any one source is added to the ARF a new version is created. Generally, less than twenty data fields are added to the ARF from any single data source. The two important exceptions to this are the AHA Hospital data

[1]The principle computer languages used by the ARF contractor include:
- SAS - a commonly used statistical and data manipulation package
- QUIKJOB - a simple procedural language used to manipulate columns of data
- COBOL - a commonly used procedural language especially efficient with data input/output.

(approximately 500 fields are added from this source) and the AMA physicians data (approximately 500 fields are added from this source). The computer costs of processing most of the data sets is under $75 each. Processing the AHA and AMA data is much more expensive, on the order of $1000 to $2000 for the twenty or so major runs required. The Mortality file is also in this expensive range since the original data includes one record per death (around 20,000 records) and must be aggregated to the county level. Single runs on the AMA, AHA, and Mortality file often cost well over $100.

## D. IN-FORMATION MANAGEMENT AND MAINTENANCE PROCEDURES

This chapter describes the quality control, data storage, and data backup procedures employed by the ARF contractor. The quality control procedures are designed to identify errors introduced during data entry and processing. The ARF contractor also utilizes several data storage and backup strategies to guard against accidental loss or corruption of data. The quality control, data storage, and backup procedures and strategies used by the ARF contractor conform to or exceed industry standards. Each of the procedures and strategies are detailed below.

### 1. Quality Control

The updating of the ARF involves extensive quality control. In general, the quality controls are (a) comparing inputs and outputs from each processing step, (b) checking totals (state and national) before and after processing, and (c) inspecting any non-matches. These quality control procedures are generally important with manually entered data because these data sets are more likely to contain errors. These quality controls procedures are discussed below.

a. Manually Comparing Input and Output

Many of the processing steps and programs described in Section 3 Appendix E compare data elements before and after each processing step. These "before and after" comparisons of selected data are manually reviewed to insure that the data were processed correctly. This review process can be seen in the creation of the AHA Hospital facility-level data. After each major step, such as the EXTRACT program, the FIPSCONV program, and the CHANGE program, the ARF contractor displayed the data in order to manually review the results of the previous program.

b. Automated Comparisons of Aggregated Data

The processing of the LAP1 data described above provides a good example of quality control checks utilizing aggregated data. In the case of the LAPI data, the source data from the Bureau of Economic Analysis includes extra records containing totals for each state. By comparing these LAPI state totals to values calculated by aggregating ARF county-level data to state levels, the ARF contractor is able to identify whether previous processing steps corrupted the data. This process also reveals the existence of duplicate county-level records,

c. Identifying and Inspecting Non-Matches

Whenever data files are merged, records that did not match are printed and examined manually. In the event that a nonmatch is the result of a data error, the data is corrected and the merge is attempted again. Sometimes the mismatch is caused by erroneous data which can not be corrected. When this occurs, the ARF contractor documents the error and forwards the explanation to ODAM. Once all the mismatches are explained adequately, the processing is considered to have been completed correctly.

2. Data Storage

The ARF contractor utilizes the various storage media available at the National Institute for Health's computer facilities. This includes permanent and temporary disk storage and magnetic tape.

For many of the data files comprising the ARF system, there is a tradeoff between storage cost and ease of access. In many cases, storage charges could be reduced by storing data in a compressed (binary or packed) format. However, if the data were stored in a more compressed fashion, it would be more difficult to view quickly. In addition, this would most probably require an extra step at the beginning and end of the process to convert it to and from this compressed format.

The ARF contractor has generally opted to store data in the uncompressed, easier-to-access format. Although the data could be stored more efficiently in terms of space, the extra processing required would likely offset any storage cost savings.

The ARF contract's total monthly storage costs for tapes is a little over $1,000 and for online storage about $1,500.

3. Backup

The contractor's data backup procedures seem adequate to prevent the accidental loss of any data. The NIH computer facility automatically creates backups of data and source code stored on the regular public disks. The backups are performed twice a week and are saved for approximately three weeks. Tape copies of important files stored on other NIH direct access devices (TMP disks and MSS) are performed by the ARF contractor. In addition, the ARF contractor copies final versions of all programs and files to tapes. Finally, the ARF contractor

makes multiple copies of all final deliverables (a copy for public use at NTH, a copy for ODAM, a copy for NTIS, and an additional copy stored at the contractor's office).

## E. SYSTEM DOCUMENTATION

In this section, the content and quality of the ARF system documentation are reviewed. Emphasis is placed on evaluating the system documentation intended for the experienced programmer. An evaluation of user and technical documentation and the timeliness, completeness, and accuracy of the ARF was included in Chapter 2.

In general, the ARF contractor documents the system very well. Output from processing specific data files is stored in a library of labeled binders. The labeling convention used allows the output from the processing of any specific data file to be found quickly and painlessly. Each binder includes a few pages summarizing the processing performed. The summary describes the processing of the data set as a whole including the original format of the source data. The remainder of the binder contains source code listings and printouts generated by each program. Each source code listing includes a documentation header containing:

*Program Name
*Program Purpose
*Author of Specifications
*Author of Original Code
*Author, Date and Purpose of Modifications
*Inputs
. outputs
•A Brief Description of the Process

Additional comments are contained in the code at critical points to facilitate the understanding of the program.

A conspicuous omission from the system documentation is system and program flowcharts. Many data processing documentation standards, including the Federal Information Processing Standards Publication Series of the National Bureau of Standards for use by all non-defense government agencies, mandate the use of flowcharts because of their time-proven value in assisting in the maintenance and modification of complex systems.

Each process should be flowcharted to illustrate the flow of data. These flowcharts would prove invaluable in both the maintenance of the current ARF system and in selecting and implementing possible future modifications to the ARF system.

## F. THE ACCESS SYSTEM

### 1. Overview

This ACCESS system was designed to facilitate research utilizing the ARF data by enabling health analysts and other professionals to easily and quickly retrieve data without requiring knowledge of computer programming. To accomplish this goal, the ARF contractor developed a system of QUIKJOB, COBOL, and SAS programs at the NIH computer facilities which allow users to access the data through a series of menus and prompts.

### 2. Capabilities

The ACCESS system provides the user with various options, The user may elect to have the output in one of three different forms: tabular reports, graphics, or a data file (SAS, PC, or EBCDIC "save" files). Similarly, the user may decide to include data from any of the ten, main ARF databases: the Basic ARF, the County-Level Hospital File, the Facility-Level Hospital File, the MD/DO File, the Census File, the Graduate Medical Education File, the Educational Commission for Foreign Medical Graduates File, the GME National Non-Specialty Timeseries

File, the GME National Specialty Timeseries File, and the GME State Non-Specialty Timeseries File. (Note that many of these files are password protected - only for internal use). From each of these files, the user may select any variable. He must, however, refer to the documentation for the variable names (F-numbers). In addition, the user is able to create addition variables with arithmetic and boolean statements. Similarly, the user is able to aggregate records to create the output at a different level. For instance, the user may aggregate the county-level basic ARF data to Small Metropolitan Statistical Areas (SMSA). Using these options, or variations of these options, the user is able to extract the data and manipulate the data as required.

3. How It Works

The software that makes up the ACCESS system consists of three basic parts: 1) preliminary programs, mainly QUIKJOB and COBOL, to prepare the basic ARF data for the next step; 2) a 12,000 line COBOL program which provides the user with the various options and generate the SAS code; and, 3) the SAS code generated on the previous step.

a. Preliminary Programs

In order to run the ACCESS system without having to mount a tape each tune (and therefore greatly reducing turnaround time), the ARF contractor decided to place the basic ARF on mass storage at NIH. However, since the file is too large to fit on mass storage as a single file, the contractor split the file into three. The purpose of these preliminary programs is to split the files into three and create the new edit tables with correct starting positions for each variable on these files. The first file includes codes and classification data and health professions data. The next file contains health facilities, utilization, and expenditure data and the population data

up to 1970 population for black females over 65. The last file contains the rest of the population data (1980 median age), environment data, and health professions.

b. Main COBOL Program

This program accepts and validates all the input from the user. Inputs to this program include an edit table for each file listing information about each variable on a file; an edit table dictionary file listing the actual line number on the edit table of each variable, another file listing the start position and length of all available geographic break levels for each input file; and a help messages file. Using these files, the program is able to verify the user's responses. For instance, if the user specifies an incorrect geographic aggregation variable (which does not appear on the geographic break file), the program flags this problem. Similarly, the program confirms the existence of a variable on a certain file specified by the user. The program, however, does not check for syntax errors on arithmetic and boolean operations entered by the user; these will cause an error on the next step when the SAS job is submitted. After checking each response, the program utilizes the information to generate the proper SAS code.

c. SAS program

This SAS program is generated from the previous COBOL program. It may be submitted in that form. The user, however, may also decide to save the code for later editing and submission.

4. Limitations

The ACCESS system does not really help the user selecting variables; basically, the user must be familiar with the variable names (F-numbers) before using the ACCESS system. The ACCESS system will provide the user with a description of the variable after the user has entered

the F-number. In addition, the default for time series data is the most recent year available (if a user enters an F-number without a year, the ACCESS system will provide the most recent data for that variable). However, if the user does not know the correct F-number, he is unable to select a variable. If the user could view the variables, along with descriptions, and highlight selected variables, the ACCESS system would be much more user-friendly.

In addition, when generating reports, the SAS program will only list five variables on a page despite the fact that the variables may contain quite small values.

## G. OTHER SUBSIDIARY FILES AND SYSTEM

This final section in the documentation chapter provides a short description of the files and systems not covered in the previous sections. Although the basic Area Resource File is the best known and central part of the ARF system, there are numerous other files which generally include more detailed data. Due to this additional detail, many of these tiles are only available for use within the Bureau. This section will briefly describe some of these various files and two other subsystems not previously described (the creation of the documentation and the creation of the profile reports).

### 1. AHA Hospital Data Files

Earlier in this chapter, the process for adding the AHA Hospital Data to the basic ARF was described in detail. The first set of programs resulted in the production of a facility-level file. The next set of programs produced a county-level files. And finally, the aggregated variables were added to the ARF. These two files created in the process are actually available to the Bureau for internal use. By using one of these data sets, an analyst of the Bureau has access to a wider selection of variables and, for the facility-level file, observations at a more detailed level.

2. Enhanced Files

Similarly, many of the other subsidiary files are useful by-products created during the process of adding data to the basic ARF. Similar to the AHA data files, these files offer the analyst a wider selection of variables to user. Often, the ARF contractor refers to these files as enhanced files (basically, source data files with the county identifiers including the modified FIPS codes added to the beginning of each record). Almost thirty files fall in this category including the Enhanced Medicare files (documented earlier), the MD/DO file, and the Census file.

3. Graduate Medical Education Files

The Graduate Medical Education files entail a somewhat different process. The original data is available to the ARF contractor at the GME program-level. The ARF contractor creates four files from this data (in addition to the basic ARF data): the GME file (program-level), the GME National Non-Specialty Timeseries File, the GME National Specialty Timeseries File, and the GME State Non-Specialty Timeseries File. The first file is created with an EXTRACT/ UPDATE system similar to the processed used for the basic ARF. The latter three are created by different aggregations of the first file.

4. Creation of Documentation

The system for creating the documentation involves a series of programs to reformat the edit table. As mentioned earlier, the edit table includes one record for each variable on the basic ARF. Each of these records includes the variable's start position, length, and description. Creating the Technical Documentation from this file is a rather straightforward process. By adding headers and page breaks and reformatting the records, this file is converted into the

Technical Documentation. These documentation programs, if desired, also create the file layouts for COBOL and PL/1.

## 5. Profile Reports

The Profile Report system is a series of SAS (and some QUIKJOB) programs which select, aggregate, and display in table format various information. The first set of programs are used to create the SAS input statements for the files from which data is extracted, the basic ARF, the MD/DO County-level file, and the AHA Hospital County-level file. Then, Virginia independent cities are aggregated back to the original county. Finally, a series of SAS programs calculate various values and output the data in table format.

# IV. FRAMEWORK FOR EVALUATING CHANGES TO THE ARF SYSTEM

In the course of collecting user comments and documenting the system, a wide variety of suggested changes to the ARF system were identified. These proposed modifications, offered by BHPr staff, outside users, and MPR, ranged from general comments to very specific, detailed suggestions.

With the goal of providing BHPr with a set of specific and detailed recommendations concerning future development of the ARF, MPR first prepared a list of all the suggested modifications. The list was organized based on the problems the suggested modification was designed to address.

Summary evaluations were performed on each of the suggested modifications. These summary evaluations were designed to identify those suggestions that were feasible, cost effective, and likely to be successfully implemented. Suggestions that satisfied these criteria were then evaluated more thoroughly and detailed cost estimates were developed.

## A. PROBLEM AREAS AND INITIAL SUGGESTED MODIFICATIONS

### 1. Problem Areas

Based on user comments and MPR's documentation of the ARF system, MPR identified the following as problem areas:

1) Underutilization
2) Content and Structure of the Data
3) Content and Structure of the Documentation
4) Difficulty in Accessing the Existing Data
5) Cost of maintenance and operation of the Existing System

These are problem areas in the sense that users identified errors or perceived that there were errors or room for improvement in these areas. The suggested modifications addressed these problem areas.

Within each problem area there were separate issues raised. Exhiiit IV-1 provides a brief description of each issue raised categorized by the appropriate problem area.

2. Suggested Modifications

Some of the problems described above can be resolved by multiple approaches, others have only one potential resolution. In many cases, a single resolution may address a number of problems simultaneously. An annotated listing of the suggested modifications is presented in Exhibit IV-2.

## B. SUMMARY EVALUATIONS: IDENTIFYING CANDIDATES FOR MORE THOROUGH EVALUATION

BHPr requested that MPR concentrate on that recommendations that could be accomplished with limited funding. Recommendations requiring substantial expenditures by HRSA or end users did not receive detailed evaluation. MPR was asked to limit their detailed analysis to those recommendations that could be appended to the current ARF support/development contract, and to recommendations required by MPR's contract for the evaluation of the ARF.

In addition, some of the suggested recommendations were the result of erroneous perceptions and therefore were not appropriate. Other suggestions had, in MPR's view, a limited likelihood of successful implementation or were unlikely to solve the intended problem.

EXHIBIT **IV-l**

ANNOTATED LISTING OF PERCEIVED PROBLEM AREAS

1. Underutilization

- At present, there is a small number of users.
- The ARF is not being utilized extensively by the health research community.

2. Content and Structure of the **Data**

- The cycle for incorporating new data was considered by users to be too long (this is related to timeliness of the data).
- Data are coded inconsistently. **Two** problem areas were defined: 1) population variables currently appear in whole numbers and in hundreds (**100s**) and 2) missing values are coded as zeroes, which confuses some users.
- Some users saw data elements, such as prevailing charges, that were incorrect.
- Important sources of data are presently not included on the ARF, such as Allied Health Professions, HMO industry data, and Census data.
- The years included in the time-series are inconsistent across variables.
- Explicit criteria for including or excluding data sets is needed for a clear understanding of the source data.
- The present ARF is missing selected geographic identifiers, such as the Social Security Administration (SSA) state and county codes and New England Consolidated Metropolitan Area (NECMA) codes.
- The current treatment of Virginia independent cities makes the ARF incompatible with external sources.
- The coding of categorical variables is too coarse. Categories are sometimes structured so that almost all observations fall into only one category.

- The geographic coding of the present ARF is limited to the county-level. The ARF data is desired on smaller geographic areas (Zip Code) or larger geographic areas (state or national). These levels are not easily obtained within the current structure of the ARF.

- The current longitudinal structure of ARF makes use of the file cumbersome for users interested in either cross-sectional analyses.

3. Content and Structure of the Documentation

- The current documentation is unclear and confusing, and make it difficult for a user to find information on data elements. The user's guide and technical documentation are organized poorly and are inconsistent with one another.

- The current structure of the documentation lacks an index and subject cross references.

- There are numerous typos, and the text is not being updated consistently.

- Some of the variables on the file are not described in the user's guide.

- The merging/filtering procedures used for adding data to the ARF are not explained adequately/accurately.

- The present version of the documentation does not include complete references of the source data.

- The present version of the documentation has an unclear use of abbreviations within the text.

4. Difficulty Accessing Existing Data

- To use the ARF requires technical (programming) skills.

- The current ACCESS system at NIH is limited (only five variables can be printed in a tabulation at a time).

- There has been difficulty in working with NTIS[49].
- The most recent version of ARF is available only on magnetic tape. The diskette versions are outdated, incomplete, or expensive.
- The user must convert the ARF from character format in order to use with other software (i.e. SAS, SPSS, and dBASE).

5. Cost of the ARF System in terms of Maintenance/Operation

- The data is stored in an uncompressed format.
- There are not enough flowcharts in the present system documentation.
- The merge processing as currently performed is inefficient.
- At present, there is unnecessary use of non-procedural languages.

[49]As mentioned previously in this report, ODAM has terminated their contract with NTIS.

EXHIBIT IV-2

ANNOTATED LISTING OF SUGGESTED MODIFICATIONS

1. Outreach

Three types of outreach are possible, all of which are intended to increase utilization.

a. Newsletter.

A brief newsletter could be produced (especially for DHHS and other federal agency users) that is distributed each time the file is updated. It could list new data added to the file (which can be adapted from the existing documentation appendix), as well as changes to the ACCESS system. The newsletter would also include other user information and information on projects that have used the file.

b. User Group.

This could be internal to HRSA or external (e.g. through the Association of Public Data Users). It would provide information sharing on how and why the file is used.

c. Product Information Dissemination.

More widespread marketing of the ARF is possible through more aggressive publicity at professional meetings or through a different sales organization. Currently, the potential uses of the ARF are not widely understood.

2. Immediate Documentation Improvements

The current documentation is awkward to use and could be improved considerably by some relatively simple actions:

a. Crosswalk between user and technical documentation.

This change would reference, in the technical documentation, the variables that are described more fully in the user documentation. It would provide variable names (F-numbers) in the user documentation so that users will know which variables are being discussed. Page headings could also be improved.

b. Dump of records.

Including a dump of one or two records in the technical documentation would show users how the data are coded.

c. Correct existing errors.

See Chapter II of this report for a list of errors and out-of-date information.

### 3. Long Range Documentation Improvements

Some options for improving the documentation repair the awkwardness of the current arrangement, others radically change the way in which data would be documented.

a. Codebook.

A codebook, in the true meaning of the term, would be developed. It would have values for each variable and summary statistics (minimum, maximum, mean) for each.

b. Complete Descrintions of all Variables.

The current user documentation would be expanded to describe coding, source, merge problems, nonresponse, and editing of each variable or set of variables.

c. Page References in the Technical Documentation.

The technical documentation could be expanded to show the page number in the user documentation where the variable is more fully described.

d. Remove Variable Descrintions from Technical Documentation.

The user documentation would include all the variable descriptions.

e. Revise Order of Presentation in User Documentation.

The user documentation would describe variables in the same order as the technical documentation.

f. Imnrove Source Data References.

A complete reference list would be provided.

g. Contact Person.

The documentation would list a contact to handle queries for each data source and for the file in general.

h. Glossary of Abbreviations.

Provide a glossary of abbreviations, and ensure that abbreviations are used consistently.

i. Suppressed Data.

There could be a description of the manner in which applicable data are suppressed, including a description of all suppression flags in a separate section of the user documentation, and clear references to the associations between flags and data elements.

j. Describe Time Series Data.

Provide tables showing which variables in the file may be used for time-series analysis.

k. Electronic Codebook.

This is a codebook in an PC-readable format, bundled with keyword/subject search software.

4. Data Content

The file contents were criticized for being incomplete, inaccurate, and out of date. Among the options for resolving some of these problems are the following:

a. Speed up Acquisition of Key Data.

BHPr should improve the speed of acquisition where possible and should produce updates as soon as possible after acquisition of key data on physicians, hospitals, morbidity, and mortality.

b. Consistent Data Coding.

We identified data that are unnecessarily presented in different units.

c. Explicit Criteria for Inclusion or Exclusion.

We would like forma1 criteria for inclusion or exclusion. For example: exclude sources with a response rate below 60 percent.

d. Introduce Consistent Missing Values Policy.

At present, two sorts of missing values are used: flags and zero-filling. With certain variables it is difficult to distinguish between a zero representing a suppression field, a zero representing the genuine value for a data element, or a zero meaning a missing value for the data element.

e. Correct Specific Data Errors.

We can provide a list of data errors found by users and by MPR.

f. Add New Sources.

Several new sources of data have been recommended (see Chapter II of this report).

g. Trv to Obtain Complete Time Series.

For example, try to get complete data for every fifth year for all types of data.

h. Add New Geographic Codes to the County File.

For example, SSA county codes would help users of HCFA and SSA data. NECMA codes were recommended for New England counties.

i. Re-evaluate Virginia Cities Policy.

Some users found the current policy hard to use, especially when merging ARF data with external data sets.

j. Improve Population, Morbidity and Mortality.

In places, these data series are inconsistent and incorrect.

k. Improve Grouping of Data.

Some variables (e.g., hospital occupancy rates) have been summarized in ways that do not provide the best information. We recommend reviewing the distribution of grouped data and making new decisions on how to summarize data in light of the current environment.

1. Implement Better Procedures to Avoid Duplicates Being Represented in the File.

This may be a documentation problem rather than a data problem. The example cited is for nursing facilities.

5. Organization of Data Elements/Products

It has been suggested that ARF be structured by geographic levels, or that cross-sectional and time-series data be separated. These resolutions would resolve objections that ARF is not local enough, that ARF is unwieldy, and that ARF removes historical data. In addition, BHPr should consider making the county profile reports accessible through a menu-driven request system that could answer many of the queries handled by HRSA staff from outsiders.

a. Data at Different Geographic Levels

- Zip Code: The suggestion is to replace the county file with a zip-code-level file.

- State File: The suggestion is to augment the county-based file with a state-level file containing key data available at the county level and other data available only at the state level.

b. Have Two Data Files.

The time-series (or complete historical) file would be maintained with all data that had ever been in the file. The most current data would be presented in a current, cross-sectional file, with a cutoff that excluded data earlier than, for example, five years previous.

c. Profile Reports Available on Demand.

The county profile reports would be available from a menu-driven system.

6. Delivery Systems/Access Techniques

The following resolutions are proposed as options for dealing with difficulty in accessing the current ARF system.

a. Automated Mainframe/PC downloading.

Develop a software system to automate transfer of data from a mainframe to a microcomputer.

b. Bulletin Board Providine Direct Access to Data Elements.

Install ARF data on a microcomputer system. The system would contain software allowing PC users to download specific data elements.

c. Bulletin Board Download of Complete Files.

A microcomputer-based system allowing remote PCs to download complete ARF data files.

d. Direct, On-Line Access.

Allow remote users to access a mainframe containing all data.

e. Use of Non-Procedural Languages.

Restructure and install data for use with SQL (Structured Query Language) or QBE (Query By Example) software.

f. Optical Discs.

Distribute optical discs containing ARF data.

g. Modify the ACCESS System.

Modify ACCESS system to make it more flexible and easier to use.

h. Provide Training.

- Improve technical skills
- ACCESS System

EXHIBIT IV-3

## SUMMARY EVALUATION CRITERIA

1. What problem will the proposed modification solve? MPR has identified the following problem areas: underutilization, content and structure of the data, content and structure of the documentation, difficulty accessing data, and cost of maintenance and operation of the ARF system.

2. How badly is the proposed modification needed? This section will address the importance of the problem in relation to the costs of the modification.

3. What is the effect of the proposed modification on the current ARF system programs and procedures? MPR will investigate how the proposed modification will affect the ARF contractor in the creation of the ARF.

4. How would the modification affect users? MPR will describe how the changes will affect users of the ARF system.

5. What are the development costs associated with the proposed modification? These costs will incorporate all costs, from the development of the design specifications to the testing of the programming.

6. What are the implementation costs? These costs will incorporate the costs of adding the new procedure to the production of the ARF the first time.

7. What are the operational costs? These costs will incorporate the average cost of processing in future years (perhaps in terms of differences in operational costs compared to the cost system).

8. Is additional user training required?

9. Will the ARF contractor require additional hardware or software?

10. Will users require additional hardware or software? Is it readily available at a reasonable cost?

11. Are alternatives available? MPR will describe alternative options. This will enable HRSA to compare the cost and benefits of recommendations which address similar problems.

12. What is the possibility of success? Is the modification likely to solve the problem?

TABLE IV-1

Overview of Proposals Addressing Problem Areas

| PROBLEM AREA | PROPOSED RESPDNSE | CODE | SPECIFIC MODIFICATION TO BE CONSIDERED | REFERENCE NUMBER |
|---|---|---|---|---|
| Underutilization | Expand Outreach Activities | C | Expand Product Information Dissemination (Newsletter, Professional Meetings, etc.) | Al |
| | | | Develop Newsletter/Disseminate Infonaation | |
| | | | Develop/Revise Brochures, Materials | |
| | | U | Organize User Group | A2 |
| | | I | Provide an Electronic Codebook | A3 |
| Data Problem | Correct Errors/Improve Timeliness | C | Correct Specific Data Errors on Existing File | B1 |
| | | I | Expanded Data | B2 |
| | | | Add New Data Sources | |
| | | | Focus on Obtaining Complete Time-Series | |
| | | | Add New Geographic Codes | |
| | | | Expand Population, Morbidity. and Mortality Data | |
| | | C | Data Editing Procedures (Front-End) | B3 |
| | | | Standardize Data Coding Procedures | |
| | | | Provide Explicit Criteria for Inclusion/Exclusion of Data | |
| | | | Standardize Missing Data Codes | |
| | | | Reevaluate Policy for Virginia Independent Cities | |
| | | | Improve Categorization of Data | |
| | | | Implement Better Procedures to Avoid Duplicates | |
| | | U | Speed Data Acquisition/Alter Distribution Cycle | B4 |
| Organization /structure | Modify Structure/Organization | U | Provide Data at Other Geographical Levels | Cl |
| | | U | Create Current Cross-sectional; Historical/Time-Series Files | C2 |
| | | U | Expand Availability of Profile Reports | C3 |
| Documentation | Modify Docmentation | C | improve Existing Documentation | D1 |
| | | | Include a Dump of Data with the Documentation | |
| | | | Expand Table of Contents | |
| | | | Correct Existing Identified Errors | |
| | | | Provide Complete Description of Source Data and References | |
| | | | Provide a Contact Person for Source Data | |
| | | | Document and Utilize a Consistent Set of Abbreviations | |
| | | | Explain Suppression Fields and Flags | |
| | | | Provide Tables of Available Tim-Series Data | |
| | | I | Develop Codebook | D2 |
| | | | Provide Cross-Walk Between User and Tech. Documentation | |
| | | | Coordinate Variable Order in User and Tech. Documentation | |
| | | | Page References linking Tech. to User Documentation | |
| | | | Move Descriptions to User Documentation/Add Complete Variable Descriptions | |
| | | I | Add Summary Statistics to Codebook | D3 |

TABLE IV-1

Overview of Proposals Addressing Problem Areas
(continued)

| PROBLEM AREA | PROPOSED RESPONSE | CODE | SPECIFIC MODIFICATION TO BE CONSIDERED | REFERENCE NUMBER |
|---|---|---|---|---|
| Access/Delivery | Alter Access/Delivery Methods | | Manual: | |
| | | I | Magnetic Tape (Alternative Formats) | E1 |
| | | I | Distribute on Alternative Media (Optical Disk) | E2 |
| | | | Computerized: | |
| | | U | Automated Access to Individual Data Elements/Observations | E3 |
| | | | Manual Access to Individual Data Elements/Observations | |
| | | | Mainframe: | |
| | | I | Flat File/Non-menu Driven Software | E4 |
| | | I | Flat File/Menu Driven Software | E5 |
| | | I | DBMS/Non-menu drive Software | E6 |
| | | I | DBMS/Menu-driven Software | E7 |
| | | | Microcanputer: | |
| | | I | Flat File/Non-menu Driven Software | E8 |
| | | I | Flat File/Menu Driven Software | E9 |
| | | I | DBMS/Non-menu drive Software | E10 |
| | | I | DBMS/Menu-driven Software | E11 |
| | | | Manual Access to ARF Data Files | |
| | | I | Microcomputer Bulletin Board | E12 |
| Efficiency/Cost | Modify system to reduce costs | U | Store Data in Canpressed Format | F1 |
| | | C | Expand Flowcharts in System Documentation | F2 |
| | | I | Reexamine/streamline Merge Procedures | F3 |
| | | I | Utilize Procedural Languages | F4 |
| | | I | Computerized Data Entry | F5 |
| | | U | Artificial intelligence-based algorithms | F6 |
| | | I | DBMS | F7 |
| | | I | Microcomputer | F8 |

CODE: C - Critical; I - Important: U - Useful

# V. EVALUATION OF SELECTED POTENTIAL CHANGES

This chapter describes, costs out, and discusses those potential changes to the ARF that were identified after a preliminary review to have the most promise (see Chapter IV). The discussion in this chapter focusses on changes intended to: increase utilization, improve the quality of the data in the ARF, improve the documentation, improve access to the file or delivery methods, increase efficiency of the system, or reduce costs. Recommendations are given in Chapter VI in the context of decisions that only BHPr can make about the purpose and future of the ARF.

## A APPROACH

For each of the five areas of potential improvement, we describe the content of the proposed change and the purpose it would serve. The costs of introducing the change are presented, with the underlying assumptions. We recognize that alternative assumptions could be used. Thus the assumptions employed need not be regarded as describing the only way in which the change might be implemented, but rather should be viewed as one way of implementing the change, chosen for illustration. Finally, we discuss the advantages and disadvantages of the changes to each area of potential improvement, in the context of the alternative actions that could be taken to achieve the same goal.

## B. CHANGES INTENDED TO INCREASE UTILIZATION

### 1. Problem

BHPr sees the ARF as an important resource whose use should be increased, both because the expense of maintaining the ARF needs to be justified and because it is a valuable, but under-

used tool for analysis. At present it is not clear how many users of the ARF there are and it is certain that there are hidden uses of the file (based on the literature review). Nevertheless, there are fewer users than there might be in BHPr, HRSA in general, and other health agencies. In addition, non-health agencies could use the file more than they do. Based on the interviews we conducted, we determined that reasons for limited use include lack of knowledge about the contents and potential uses of the file, a sense of the documentation being cumbersome, and insufficient technical programming skills.

## 2. Explanation and Purpose of Change

There are numerous ways in which the accessibility of the ARF could be increased. We believe that improved information and the availability of an electronic codebook are two of the most direct ways of achieving this, particularly among federal government users.

### a. Expand Dissemination of Information about Product

Increased information about the ARF could take many forms; we have focussed on two approaches:

1. Change: Develop and distribute a newsletter

   Purpose: Provide information to users and potential users of the ARF about changes in file contents (additions and deletions), uses of the file and names of project directors who have used the file, and changes (for example to ACCESS) or improvements to the overall system. The newsletter would include comments from users as well as information from BHPr

2. Change: Increase availability and distriiute descriptive materials about the ARF.

   Purpose: Provide information about the ARF to potential new markets through professional and other interest groups and publications. Would include distribution of brochures at professional meetings and marketing of the ARF through a sales organization.

The newsletter perhaps has the most potential for increasing use within the federal government while the other information dissemination may have more potential for increasing outside use.

b. Electronic Codebook

Some users referred to the size and format of the ARF file documentation as a deterrent to their use of the ARF data file. Later in this chapter, we discuss various suggested changes involving the content and structure of the printed ARF documentation. These changes are intended to make the documentation more comprehensive and easier to use. Another method for simplifying use of the documentation and thus the data file is to provide a codebook in electronic form:

| | |
|---|---|
| Change: | Develop an electronic codebook for the ARF documentation. The codebook would consist of a user-friendly microcomputer software package which would enable users to easily obtain brief descriptions of the names, location, and content of ARF variables. |
| Purpose: | The electronic codebook would utilize computerized search algorithms to allow convenient access to information about data on the ARF file. In addition to being easier to use, a diskette containing the codebook would be less expensive to distriiute than either a magnetic tape or the printed documentation. |

3. Costs and Assumptions

The costs of increasing the dissemination of information about the file and making an electronic codebook available are presented here.

a. Increase Dissemination of Information about Product

The costs of this activity are largely labor costs, printing costs, and the cost of hiring a marketing group to produce and disseminate materials. Table V.l shows the labor hours and principal other direct costs for each of the two activities.

Develon and Distribute a Newsletter. We assume that BHPr would take responsibility internally for the newsletter but that the ARF contractor may supply some of the material. The activities involved would include:

- Prepare a prototype newsletter after internal discussions about what should be included, send out to BHPr users for comment and update accordingly
- Develop a distribution plan: decide how extensive the mailing would be and develop a mailing list
- Distribute the newsletter: assuming a semi-annual newsletter to coincide with the new ARF, prepare, print and mail the newsletter (we assumed a 4-sided newsletter, with 200 copies distributed using internal PHS mail or federal frank)

We estimate that this activity would require 86 hours of planning and 88 hours a year to produce and send out the newsletter.

Develon and Distribute Other Materials. We assume that BHPr would hire a marketing consultant to assist in the revamping of the currently available materials and in developing and implementing a distribution plan. The costs of this activity could be almost any amount BHPr chose. The activities would include:

- Select a marketing organization and decide which materials were to be developed or revised, and working with the marketing organization to develop or revise the materials (assumes a brochure and a booklet)
- Develop a distribution plan, with the assistance of the marketing organization
- Implement a marketing plan which would include attendance at 3 professional association meetings per year and printing of 500 brochures and booklets per year

TABLE V.1

COST ESTIMATES FOR CHANGES INTENDED TO INCREASE ARF USE:

1. INFORMATION DISSEMINATION

| Activity | Labor Hours for ARF Contractor | | | BHPr Hours | | Other Direct Costs | |
|---|---|---|---|---|---|---|---|
| | System Analyst | Programmer/Analyst | Clerical | Staff | Clerical | Printing | Marketing |
| 1. Newsletter Plan & Implement - (1 Year) | 24 | 40 | 4 | 66 | 40 | 40 | |
| 2. Marketing | 0 | 0 | | go | | | 13,000 |
| Total | 24 | 40 | 4 | 146 | 40 | | 13,000 |

Our estimate of the total costs for this activity are 80 hours for BHPr, marketing organization at $10,000 for development of materials and $3,000 a year for dissemination, plus annual printing costs of under $500.

b. Electronic Codebook

Estimated costs associated with developing and maintaining the electronic codebook are presented in Table V.2. During the development and implementation phase, the codebook software must be developed and the database converted for use with the codebook. The following activities must be performed:

Specify the functional requirements of the software
Design, code and test the software
Prepare documentation and a user's guide
Prepare final package and installation procedures
Convert the existing documentation to electronic codebook format
Specify and flag search keywords

As new versions of the ARF become available, the electronic codebook must be updated. The update process involves reloading the revised documentation database and specifying search keywords for new variables. Non-labor costs are estimated at $1,000 for software for the development/implementation tasks.

4. Advantages and Disadvantages

Both of these approaches are likely to improve public relations for the Bureau since they would show that BHPr is interested not only in maintaining the file for internal use but also in providing a research tool for researchers outside of BHPr. The less expensive of the two approaches to increasing use of the file is increased information dissemination, and within this

TABLE V.2

COST ESTIMATES FOR ELECTRONIC CODEBOOK

| Activity | Labor Hours for ARF Contractor | | | | BHPr Hours | | Other Direct Costs |
|---|---|---|---|---|---|---|---|
| | Sys. Analyst | Programmer | Analyst | Clerical | Staff | Clerical | |
| **FIRST YEAR DEVELOPMENT/IMPLEMENTATION** | | | | | | | |
| a. Develop Codebook Software | 128 | 304 | 104 | 80 | 0 | 0 | $1,000 |
| b. Create Initial Codebook Database | 48 | 56 | 104 | 16 | 0 | 0 | Under $500 |
| **ON-GOING ANNUAL OPERATIONS** | | | | | | | |
| a. Update Codebook Database | 24 | 24 | 40 | 8 | 0 | 0 | Under $500 |

approach the most fruitful of the activities described would, in the short run, be the development of the newsletter. The newsletter can be accomplished fairly cheaply, and would fill an information gap that exists within HRSA.

As indicated at the start of the chapter, many of the changes described in subsequent sections may make it easier to use the ARF and hence may increase users indirectly, but none are as likely to increase the number of users as information dissemination.

An electronic codebook as an adjunct to the ACCESS system could prove very useful, but would have to be evaluated in terms of the number of users who would benefit, relative to the cost. At present, the user must sift through two one-hundred page documents to discover what is contained in ARF. As mentioned in the codebook recommendation below, many important items, such as cross-references and the means, maximums, and minimum values for variables, are missing. The electronic codebook would allow the user to have access to all this information through user-friendly retrieval software. Upon entering the program, the user would be prompted through a series of menus and queries which would reveal exactly what information is available. The user would be able to compare similar variables, and view means, maximums, and minimums for any variable. The system could even include a program generator which would create the code (SAS, COBOL, or parameter file) to extract data.

## C. CHANGES INTENDED TO ADDRESS DATA PROBLEMS

Bad impressions of data problems in the ARF are hard to change, even long after the errors have been corrected. While increased information about the file may reduce misperceptions about the file, increased avoidance of problems might be more effective.

1. Problem

The ARF is not perfect, as it contains a few instances of misleading, unusable, or incorrect data. These problems originate both in the source data and in the procedures for editing source data prior to merging with the ARF.

2. Explanation and Purpose of Change

The potential solutions, some proposed by users and some based on a review of the file and editing procedures, include correcting specific errors found in the file, expanding the data base to include additional data sources and fields, and modified front-end procedures for editing and merging data with the data base. The-se activities would improve the value of the file for some users, though a careful evaluation would be needed to avoid introducing new problems for some users in the process of helping others. The changes described here do not all have equal merit; choosing among them is the subject of Chapter VI.

a. Correct Specific Errors

BHPr corrects errors in the file whenever this is required. The change that we describe includes designating an outreach person to analyze and respond to potential errors and to provide feedback to the person providing notice of an error. The purpose of this change is to ensure both that potential errors are promptly reviewed and corrected when necessary and that users are provided with explanations for why the file has or has not been corrected.

b. Expand the Data Base

The expansion of the data base would include the addition of new data sources, an alternative approach to maintaining timeseries data on the file, the addition of geographic codes requested by users, and the modification of the population fields in the file:

1. Change: Add new data sources
   Purpose: Make new data sources available by a continual search for new sources (none are currently proposed)

2. Change: Try to obtain complete time-series for key variables regularly (rather than acquiring annual data for a few types of data)
   Purpose: Make time-series analyses more straightforward through a more complete set of data for specific years

3. Change: Add new geographic codes that users have requested (and possibly make trailer records for some types of area)
   Purpose: Improve the ease with which other major data sources not included in the ARF can be merged with the ARF

4. Change: Expand or revise the existing population, morbidity, and mortality data
   Purpose: Improve the value and use of the ARF, particularly to other HRSA agencies

c. Modify Editing (Front-End) Procedures

The ARF front-end procedures would be modified to improve the consistency of presentation of data from different sources and to avoid adding incorrect data to the ARF in the future:

1. Change: Standardize the units or codes for data from different sources
   Purpose: Make it easier for the user to use time-series data from different sources

2. Change: Provide explicit criteria for when a data source will be included or excluded from the file
   Purpose: Eliminate inconsistencies in the quality of the data in the file and provide the user with a clearer understanding of how much review the source data receive before addition to ARF so that they can decide whether or not data meet their own standards of acceptability

3. Change: Re-evaluate and standardize missing value codes
   Purpose: Provide uniform and unambiguous missing values

4. Change: Reevaluate the policy of maintaining separate observations for five of the Virginia cities
   Purpose: Simplify use of the ARF across years in which data for Virginia cities are handled differently and simplify merging with other county level data

5. Change: Improve Categorization of Data
   Purpose: Provide more useful data by changing the ranges on aggregated data

6. Change: Implement better procedures for avoiding inclusion of duplicates in the file
   Purpose: Improve the accuracy of the data included in the ARF

## 3. Costs and Assumptions

The costs of introducing the changes are described with underlying assumptions in this section. We have assumed that most of the changes would be made by the ARF contractor with supervision from BHPr. Table V.3 shows the total costs in terms of labor hours by type.

### a. Correct Specific Errors

The process we have assumed would have an ODAM staff member handle queries about data errors, review them with the ARF contractor, report back to the caller, and prepare an article for the next newsletter describing the problem. and its resolution. We have estimated BHPr staff time of 48 hours annually for this activity. We have not included ARF contractor hours because the time spent reviewing data would occur regardless of the change.

### b. Expand Data

Four activities are included here which require policy decisions followed by additional programming.

TABLE V.3

COST ESTIMATES FOR CHANGES INTENDED TO IMPROVE DATA CONTENT IN THE ARF

| Activity | Labor Hours for ARP Contractor | | | BHPr Hours | | Other Direct Costs |
|---|---|---|---|---|---|---|
| | System Analyst | Programmer/Analyst | Clerical | Staff | Clerical | Data |
| 1. Correct Specific Errors (Annually) | 0 | 0 | 0 | 36 | 12 | |
| 2. Expand Data | | | | | | |
| a. Add new Data Sources | 28 | 136 | 0 | 24 | 4 | 500/year |
| b. Focus on Obtaining Computer Time Series | 56 | 112 | 0 | 64 | | |
| c. Add New Geographic Codes (Annually) | 4 | 44 | 0 | 4 | | |
| d. Expand Population Data (Once) | 13 | 308 | 0 | 16 | | |
| Total | 101 | 600 | 0 | 108 | 16 | |
| 3. Modify Front-End Procedures | | | | | | |
| a. Standardize Data Coding (Once) | 10 | 128 | 0 | 16 | | |
| b. Provide Explicit Criteria for Including Data Sources (Once) | 10 | 0 | 0 | 44 | | |
| c. Standardize Missing Data Codes (Once) | 8 | 436 | 0 | 8 | | |
| d. Reevaluate VA cities Policy (Once) | 12 | 220 | 0 | 4 | | |
| e. Improve Categorization (Once) | 8 | 532/40 | 0 | 4 | | |
| f. Implement Procedures to Avoid Duplicates (Once) | 16 | 168 | 0 | 4 | | |
| Implementation (Annual) | | 80 | 0 | | | |
| Total | 64 | 1,564/1,072 | 0 | 80 | | |
| GRAND TOTAL | | | | | | |

DRAFT FINAL

Add New Data Sources. Under this task we include the following activities:

- Develop a plan for identifying new data, disseminate for comment, and revise
- Add new data as the sources are identified (we have assumed an additional two sources a year under this more aggressive approach); this requires evaluation of the data, front-end review and editing, merging, and documenting in the user materials.

We estimate that the ARF contractor would require 164 hours for this additional work and BHPr staff would require 28 hours.

Focus on Obtaining Complete Time Series. To make this change, two activities would be required:

- Review current availability of data across years and prepare detailed tables. Request input from users with respect to uses and review their suggestions
- Select specific years for which data acquisition would be concentrated, revise file, and focus future efforts on the selected cycle

We estimate that the ARF contractor would require 168 hours to implement the change in the first year and that BHPr staff would spend 64 hours evaluating and selecting a cycle.

Add New Geographic Codes. The activities would be a once a year acquisition and merging of one additional geographic code which would require:

- Decision on which code to merge and whether to add trailers to the file (assume SSA and NECMA codes are the first to be added)
- Writing and testing front-end and merge programs and documenting the data in the file

Assuming no trailer file development, we estimate that the total hours for the ARF contractor would be 52 hours a year.

Expand Population Data. Under this task, the current population (including mortality and natality data) would be reviewed with other HRSA agencies to establish which variables are most important to those agencies and to decide whether they can be included on the file. The assumptions are:

- Prepare a memo and meet with other agencies
- Prepare a plan for modifying ARF as needed
- Modify data in the file for selected years (five years assumed) and revise documentation

We have estimated 321 hours for the ARF contractor and 16 hours for BHPr staff in order to modify the first ARF data file. We have assumed no additional costs in later years.

c. Modifv Editing (Front-End) Procedures

All of the changes in this section require policy decisions concerning how and to what extent the ARF system should modify source data. We have shown the budgets for these activities separately, but the required reviews of past practice and the meetings to decide if these practices should change could be coordinated. Our cost estimates assume that this coordination occurs.

Standardize Data Coding Procedures. This change would include a planning and policy decision and a recreating the ARF file:

- Review the current policy of passing data through without change, including a meeting with HRSA users, and preparation of a policy memo

- Identify affected data (two sources assumed of which one is population), revise front end programs, merge, and document

We have assumed a one-time cost of 138 hours of ARF contractor and 16 hours of BHPr time.

Provide Explicit Criteria for Inclusion or Exclusion of Data. This activity includes:

- A review of past practice and preparation of a memo that describes that policy
- Meet with HRSA users to decide if the past policy is acceptable, formulate future policy, and document the policy in the user documentation

We have assumed a fairly brief review and revision of policy requiring 10 hours of ARF contractor and 44 hours of BHPr time.

Standardize Missing Data Codes. Activities are:

- Review past policy and determine future policy, through preparation of a memo and meeting with HRSA users
- Revise the missing values by identifying all sources known to have missing values, revising the programs to change missing data, rerunning the merge, and documenting the change

Although reviewing the policy would not take extensive time (16 hours of ARF contractor and 8 hours of BHPr staff), we estimate that rerunning the ARF could be most labor intensive, requiring at least 428 ARF contractor hours.

Reevaluate Policy for Virginia Cities. The activities include review and decision, as well as reprogramming:

- Meet with major users and establish needs with respect to Virginia cities and prepare a memo summarizing future policy (ARF contractor 12 hours, BHPr 4 hours)

- Identify which merges would need to be rerun in light of a new policy and rewrite the merge programs (10 sources per year assumed; 176 hours of ARF contractor)

- Documentation of the change (44 hours of ARF contractor)

Imnrove Categorization of Data. This activity would include:

- Review all aggregated data on the file to establish whether the aggregations are providing useful information and establish a system for review in the future

- Select the variables to be revised and produce documentation of decisions

- Regroup and revise the selected variables (five sources across five years assumed); requires rewriting front-ends and re-merging the data sources

Under these assumptions we have estimated 540 ARF contractor hours. However, the grouping policy could be changed only for future merges for the selected variables at the relatively low cost estimate of 48 ARF contractor hours.

Imnlement Better Procedures to Avoid Duplicates. The activities include policy review and re-merging:

- Review past policy, including conducting a meeting with key users, and prepare a decision memo which would define when and how the search for duplicates would be conducted. The memo would also review past merged sources for which there are concerns. This would require contacting the agency who provided the source. file

- Write a model duplicate search program and test

- Run a duplicate search on previously merged sources (5 assumed) including rerunning the merges if duplicates are found (184 ARF contractor hours for this and the preceding two tasks)

- Running duplicate searches on 10 sources/year in future (80 hours per year)

4. Advantages and Disadvantages

The three approaches to improving data usability may be considered optional since BHPr already has policies for deciding whether data should be added and removes bad data when it is identified. However, it is now a useful juncture for reviewing past data handling policies with the goal of improving them whenever the cost of doing so is reasonable.

The results of the set of activities described in this section can not be achieved directly by other means although they may be achieved indirectly to a limited degree by increased information about data sources and coding in the documentation and through increased training in programming and analysis techniques.

Formalizing procedures for finding and handling data errors (approach (a) above) is a relatively low cost activity and would result in BHPr being and appearing to be helpful to other agencies.

In general, BHPr has done a good job of identifying new data sources and adding them to the file, so the benefits of adding additional resources here are less clear. Of the four data expansions suggested, the modification of population data to meet other HRSA bureau needs and the addition of new geographic codes seem to have the most merit. The adoption of limited cycles of time series data could save BHPr a great deal, at the cost of reduced timeliness in some key fields.

As we discuss further in Chapter VI, many users suggested changes that counter the current philosophy of the ARF. This is seen most acutely in the suggestions for changing the front-end editing policies that have developed over the years. Without some of these changes, the ARF will continue to include bad data from time to time, and this will affect users attitudes towards it. The costs of modifying the front-end policies must be measured against the costs of

bad press. Modifying the missing values procedures has the highest cost and does not seem critical as it has not impaired past use of the file for most users. Improving categorization of data has a high cost but may have a higher payoff and could be accomplished quite cheaply if old fields were not modified. The other changes are inexpensive and likely to result in improved communication with users (especially providing explicit criteria for inclusion of data).

## D. CHANGES INTENDED TO IMPROVE DOCUMENTATION

The documentation provided to users consists of printed "technical" and "user" sections that correspond to a file layout and background information on the source data in the file. There are many ways to make the file easier to use that could promote an increase in new ARF users and more frequent and appropriate use by the current users. The changes discussed here were suggested by users or came out of a detailed review of the documentation.

### 1. Problem

The documentation could be improved in two ways: (1) the explanation of the content and meaning of the variables in the file and (2) the organization of the documentation. It is very important that users have a clear idea of the content of variables on the file and that they are aware of any problems with particular variables. Documentation should therefore be as complete, clear, and consistent as possible. A number of users complained that it was hard to find individual variables in both the technical documentation and the user documentation. With a file as large as ARF, this may always be somewhat of a problem, but there are a number of ways to help the user find more information about the variables.

## 2. Explanation and Pumose of Changes

There are three ways in which the documentation could be enhanced, which can be thought of as staged improvements to the existing documentation with the goals of improving both content and organization. First, there is a set of improvements to the existing documentation that could be made without a major change in the way the documentation is currently maintained and presented. Beyond such basic changes there are more substantial modifications that cumulatively culminate in the production of a codebook. Finally, the codebook could be augmented with detailed statistics on each variable on the file. Each of these changes would improve the usability of the documentation.

### a. Improve Existing Documentation.

Some improvements to the existing documentation are critically important. We have identified eight improvements, none of which would require any substantial changes in the existing documentation, but all of which, to varying degrees, would help the users find and understand the data in the file. Furthermore, any more elaborate improvement, such as the development of a codebook, would benefit if these changes had already been made. These improvements and the reasons for them are as follows:

1. Change: Include a dump of one observation in the file
   Purpose: Help the user to understand the way a particular variable is coded in the file, especially when this is not clear from the technical or user documentation

2. Change: Provide a more detailed contents to the technical documentation
   Purpose: With a few more sub-headings in the contents, users will find the section of the file they are looking for more quickly

3. Change: Correct existing errors
   Purpose: There will always be a continuing need to correct errors made in previous additions to the documentation

4. Change: Provide complete descriptions of the source data and references
Purpose: It is not always possible to determine exactly which data are on the file from the current list of sources. Users should be able to check source data for additional information if they wish to do so.

5. Change: Provide contact persons for source data whenever possible
Purpose: Assist users in learning about a data source before using it

6. Change: Use and document a consistent set of names and abbreviations
Purpose: Clarify the technical documentation, given the very limited space available for each variable

7. Change: List and explain suppression fields and flags
Purpose: Encourage the informed use of the fields flagged as having data suppressed

8. Change: Provide tables of time-series data
Purpose: Assist the user to identify the time-series available and the latest data available for frequently used types of data

b. Develon a Codebook.

A codebook for the ARF that provided a more complete description for every variable on the tile would be very valuable to users. It would save them having to read the file to understand how some variables are coded, and in some cases having to return to the source data to understand exactly what the variables mean. A staged approach is possible in which steps could be taken that fall short of a complete codebook but that have some of its features. The development of a complete codebook could thus be gradual. Four specific changes are described below with the purpose each would serve:

1. Change: Provide a cross-walk between technical and user documentation
Purpose: A flag would alert the user that the variable is discussed more fully in the user documentation and all discussions of variables in the user documentation would include the variable name from the technical documentation

2. Change: Coordinate variable order in the technical and user documentation
   Purpose: Currently, it is not easy to be certain whether a specific variable is described in the user documentation. By observing the same general order in the two sections, the user would find descriptions more rapidly

3. Change: Provide page references linking the technical and user documentation
   Purpose: Rather than simply flagging the fact that a variable is discussed in the user documentation, this change would provide the actual page number where the discussion is found

4. Change: Develop complete descriptions of each variable and move descriptions to the user documentation
   Purpose: The user would be provided with more than the one line description currently available in the technical documentation, and would thus have a better understanding of the variable. The technical documentation would be simplified in layout and the variabie descriptions would be combined with the other information on source data provided in the user documentation.

c. Provide Summary Statistics.

As an adjunct to the codebook, a set of summary statistics would be very helpful to users. Summary data show how the data in the file are distributed, help the user to understand the data content, and provide evidence on whether the file has been read correctly. The summary statistics would consist of frequency distributions for variables taking on integer or alphabetic codes (such as the 1985 AHEC code) and means and ranges for continuous variables (such as total population).

3. Costs and Assumptions

The costs of improving the documentation and the assumptions underlying the estimates are presented for each of the three phases of improvement.

a. Improve Existing Documentation.

The costs for this activity are largely labor costs for the ARF contractor to revise the existing documentation or prepare new parts of the documentation. Moreover the costs are generally not very great. Table V.4 shows the labor hours by type of labor for each of the eight activities.

Include a dump of one record in the documentation. This activity will require that the ARF contractor:

- Write a program to dump a record of the file
- Modify the documentation to incorporate the dump and prepare an explanation for the user documentation on how to read the dump
- Run the dump program each time the public version of the ARF is prepared

We estimate that 28 person hours of the ARF contractor would be required to implement this activity, and an additional 4 hours would be required every time the ARF is updated. BHPr time for review of the proposed change is estimated at 4 hours.

Provide More Detailed Contents to the Technical Documentation. The current contents would be expanded to would allow the user to find the data they were looking for more quickly. For example, under the heading of "POPULATION" the following subheadings could appear:

- Total population and population by age, sex, and race
- Households and Families
- Births

TABLE? V.4

COST ESTIMATES FOR IMPROVING EXISTING DOCUMENTATION

| Activity | Labor Hours for ARF Contractor | | | BHPr Hours | | Other Direct Costs ($) |
|---|---|---|---|---|---|---|
| | Systems Analyst | Programmer/Analyst | Clerical | Staff | Clerical | |
| 1. Include a Dump of One Record (Implement) | 4 | 32 | | 4 | | |
| (Annual) | | 4 | | | | |
| 2. Provide More Detailed Contents (Once) | 4 | 80 | | | | |
| 3. Correct Errors in Existing Documentation (Annual) | | 20 | | | | |
| 4. Provide Complete Descriptions of Source Data (Once) | | 56 | | | | |
| 5. Provide list of Contact Persons (Annually) | | 48 | | | | |
| 6. Revise Naming and Abbreviations (Once) | 4 | 40 | | 4 | | |
| 7. Provide a List and Explanation of the Suppression Fields (Once) | | 8 | | | | |
| 8. Prepare Tabks of Time Series Data (Once) | 4 | 56 | | 4 | | |
| Total | 16 | 320 | | 12 | | |

- Mortality rates and deaths--infant deaths, deaths by cause, and by age
- Morbidity
- Income
- Employment and Unemployment
- Housing

All that would be required would be the preparation of a more detailed list and the modification of the documentation to provide the more detailed contents. We estimate a total of 84 ARF contractor labor hours.

Correct Errors in the Existing Documentation. This continuing activity is assumed to require 5 errors to be corrected per year, at 4 hours per error.

Provide Complete Descriptions of Source Data with References. To prepare complete descriptions of source data will require that the ARF contractor (or BHPr) staff perform the following:

- Review all sources of data on the file, and prepare a complete list by year
- Check whether the data were ever published, and prepare references for all published data
- Update the source list at the end of the technical documentation with this new material
- Update the technical documentation to ensure that the names given in the source list are given for each variable

This would require an estimated 56 hours of ARF contractor labor. Once this update has been completed, subsequent updates will require only minimal additional time. Therefore, subsequent updates have not been budgeted.

Provide List of Contact Persons with Source Information. For each major source that is willing to provide it, a contact name would be included in the documentation, enabling users to discuss the data with the designated contact person. This activity would require:

- Contact all the major data sources for the name of the contact person each time the documentation is produced for a new public version of the ARF
- Update the source list with the names of the contact persons

This activity would require an estimated combined 24 person hours of ARF contractor and BHPr time for each version of the public file, or 48 hours per year.

Revise the Naming and Abbreviations in the Documentation. The inconsistent names and abbreviations used in the documentation would be revised once by taking the following steps:

- Prepare a sorted list of abbreviations used in the documentation with their meanings
- Review the variable descriptions within a specific subject (for example physicians) and identify ambiguity
- At a meeting with BHPr agree on a revised list of abbreviations and descriptions that avoids duplicates and is consistent within a source
- Revise the technical documentation

This would require an estimated 44 hours of ARF contractor time plus 4 hours BHPr review time. Subsequent updates would require little time and are not budgeted.

Provide a List and an Explanation of Suppression Fields and Flags. This change would provide a list of all suppression fields with an explanation in an identified section of the user documentation (such as with the missing values). We estimate that 8 hours of ARF contractor

time will be sufficient to make the list, write the explanation(s) and modify the user documentation. Subsequent updates should take only minimal time and are not budgeted.

Prepare Tables of Time-Series Data. This will produce a series of tables showing which years particular data items are available. To prepare these tables will require a careful review of the documentation for comparable data items in different years, preparation of the tables for key items, and modification of the user documentation to include the tables. This will require and estimated 60 ARF contractor hours and BHPr input of 4 hours. Subsequent updates should take only minimal time and are not budgeted.

b. Develou a Codebook.

**The** costs of this activity are largely the labor costs of the ARF contractor to revise the documentation and the way it is produced. Table V.5 shows the labor hours by type for each of the four activities.

Develou Cross-Walk between User and Technical Documentation. Developing a cross-walk would require three activities:

- Identifying every variable in the technical documentation for which additional information is provided in the user documentation (in the final development of a codebook, this step would be unnecessary because all variables listed in the file layout would be documented elsewhere)
- The technical documentation would be modified to include a flag
- The user documentation would be modified to show variable names

For these activities, we have estimated 124 hours for the ARF contractor. Subsequent updates should take only minimal time and are not budgeted.

TABLE V.5

COST ESTIMATES FOR DEVELOPING A CODEBOOK

| Activity | Labor Hours for ARF Contractor | | | BHPr Hours | | Other Direct Costs ($) |
|---|---|---|---|---|---|---|
| | System Analyst | Programmer/Analyst | Clerical | Staff | Clerical | |
| 1. Develop a Crosswalk Between User and Technical Documentation (Once) | 4 | 120 | | | | |
| 2. Coordinate Variable Order in User and Technical Documentation (Once) | 4 | 36 | | | | |
| 3. Provide Page References in the Technical Documentation (Annual) | 2 | 48 | | | | |
| 4. Develop Complete Descriptions of Variables and Move Into User Documentation (Once) | 28 | 820 | | 24 | | |
| Total | 36 | 1,024 | | 24 | | |

Coordinate Variable Order in the Technical and User Documentation. Under this activity, the user documentation would be reorganized to match the order in the technical documentation. This activity is estimated to take 40 ARF contractor hours, for deciding on the appropriate order and updating the user documentation (a one-time activity).

Provide Page References in the Technical Documentation to the User Documentation. This activity requires that the user documentation be indexed so that page references can be given, and that the appropriate indexes are then added to the technical documentation. We estimated 50 hours per year for the ARF contractor if the activity is done manually. Alternatively, this cross-referencing could be automated so that the index in the user documentation would be picked up automatically every time the technical documentation was prepared. Because there would be no need to undertake the automated version if the full codebook were to be the ultimate goal, we have not budgeted the automated version of the page referencing.

Develop Complete Descriptions of Variables and Move Them to the User Documentation. The three activities that complete the development of a codebook are:

- Writing variable descriptions, which requires first that the format and content of the descriptions be specified and that the descriptions be written. We assumed 6,200 variables, with an average of ten variables sharing almost identical descriptions (except for such variations as year), and that each group of ten would take one hour to write (624 ARF contractor hours)
- The technical documentation would be reformatted, with the removal of the description (52 ARF contractor hours)
- The user documentation would be respecified and new documentation incorporating the variable descriptions and the existing information would be prepared (172 ARF contractor hours)

These are one-time costs of implementing the system. Subsequent costs would be for writing descriptions for new variables. These would be budgeted at 10 per hour; this is not shown in Table VS.

c. Provide Summary Statistics.

The three activities that were assumed for this task were: to determine which variables would be described by frequency distributions and which by means, minimums, and maximums; to write a program to produce the descriptive data and arrange it into the order required for inclusion in a codebook; and to run summary statistics every time the public version of the ARF is produced. There is a high one-time cost of doing this (294 hours ARF contractor time); based on an estimate of two minutes per variable in the ARF to classify and flag it by type (mean or frequency) and 84 hours to write a program that runs the statistics and adds them to the documentation. After the initial work is completed, running and checking the statistics and adding them to the documentation each time the public ARF is produced is estimated to be about 80 hours per year. Table V.6 shows the costs for this activity as well as the total costs for the other two documentation activities.

4. Advantages and Disadvantages

Making the three changes to the documentation would result in the ARF coming to resemble other public use files in the quality of the supporting documentation. Given that ARF is a secondary rather than a primary data source, the philosophy of adding so much documentation may well be questioned. One change that could affect the need for additional documentation is frequency of use: frequent users will need less additional documentation than occasional users, and will be less intimidated by the current format. Given' the mix of users,

TABLE V.6

COST ESTIMATES FOR MODIFYING THE DOCUMENTATION

| Activity | Labor Hours for ARF Contractor | | | BHPr Hours | | Other Direct Costs(S) |
|---|---|---|---|---|---|---|
| | Systems Analyst | Programmer/Analyst | Clerical | staff | Clerical | |
| 1. Improve Existing Documentation | 16 | 320 | | 12 | | |
| 2. Create Codebook | 36 | 1,024 | | 24 | | |
| 3. Add Summary Statistics to Codebook (Once) | 8 | 286 | | | | |
| (Annually) | 8 | 72 | | | | |

however, it is clear from the user comments that some changes to the documentation would be helpful, and of the three groups of changes discussed here, making modifications to the existing documentation is clearly the cheapest, while development of a full codebook is the most expensive.

## E. CHANGES INTENDED TO IMPROVE SYSTEM EFFICIENCY

BHPr has expressed a continuing concern over the costs associated with the operation, maintenance and enhancement of the ARF system. As a result, one of the mandates given to MPR was to analyze the ARF system and make recommendations which would increase system efficiency, where efficiency would be measured by the overall cost of operation, maintenance, and development. The lower the cost, the more efficient the system.

### 1. Problem

Several factors contribute to the overall operating costs of the ARF system. These include:

#### a. Ease of system maintenance

The ARF system is a large and complex system containing both manual and computerized operations. The extent to which these processes are clearly documented affects the costs of future problem resolution and system modifications. Such modifications may include the use of new technologies or new tools in the preparation of the ARF files. The more thoroughly the ARF system is documented, the less expensive future troubleshooting and modification will be. A thoroughly documented system will also ease the transfer of the ARF processing to another contractor if BHPr desires to do so in the future.

b. Processing (CPU) efficiency

A substantial amount of data manipulation is required in order to transform, aggregate, and load source data into the ARF system. The data processing procedures utilize CPU, input/output, and storage resources. More efficient use of these resources will reduce costs.

c. Manual onerations

The current ARF utilizes data obtained in both machine-readable and non-machine-readable formats. The non-machine-readable data require data entry and additional error checking, both of which are labor-intensive activities. Streamlining or reducing the need for these manual activities may result in cost reductions.

d. Use of expensive computing resources

Until recently, the ARF system relied almost exclusively on expensive mainframe processors for the data processing activities. With the availability of more powerful microcomputers, transfer of processing from the mainframe to the less expensive microcomputer environment is now viable[50].

e. Storage and access methods

The current ARF system relies heavily on batch processing of flat files. The emergence of alternative storage and access techniques offered by data base management software offers the possibility of reduced storage requirements and simplified access. The reduced storage requirement promises cost reductions, although the savings are offset in part by the required

[50]The current ARF contractor has already begun to utilize microcomputers more extensively in the ARF system processing.

DBMS overhead. The use of non-procedural access languages, such as SQL, requires fewer programming resources than traditional methods, which in turn translates into reduced costs.

2. Explanation and Purpose of Changes

Modifications addressing each of these issues were suggested by BHPr, users, and MPR. The specific modifications are described in detail below.

a. Changes Affecting System Maintenance

Many data processing documentation standards, including the Federal Information Processing Standards Publication Series of the National Bureau of Standards, for use by all non-defense government agencies, require the use of flowcharts because of their value in maintenance and modification of complex systems, including ARF.

As discussed in Chapter III, the current ARF system documentation contains flowcharts describing processing flows. However, MPR found that these flowcharts sometimes omitted detailed file information and, in some cases, provided only summaries of the processing.

In addition, the flowcharts do not appear to be stored in a centralized document available for general reference. The flowcharts were drawn freehand rather than with a computerized flowchart system. This discourages active maintenance of up-to-date flowcharts.

1. Change: Expand the number and detail of the flowcharts contained in the ARF system documentation. The flowcharts should be maintained in machine-readable format using a microcomputer-based software package such as EasyFlow 2. The flowcharts, both the machine-readable files and hard copy, should be stored in a centralized location in addition to being stored with the printed output.

   Purpose: Expanding the number and detail of the system flowcharts will assist programmers in maintenance of the existing system and will prove

invaluable to future developers who may not be familiar with the details of the current system.

b. Changes Affecting Processing (CPU) Efficiency

The ARF contractor uses both COBOL and the SAS programming language throughout the ARF system. An important feature offered by SAS is ease of programming. However, SAS programs generally require more computer resources than COBOL programs performing the same functions. Therefore, the ARF contractor should review their use of SAS within the ARF system and replace SAS-based programs with COBOL versions where programs are likely to be used repeatedly without modification.

1. Change: Investigate the Expanded Use of COBOL instead of SAS. Examples of good candidates for conversion are:

*SELECT program for LAP1 data
- STATE.DUMl? program for LAPI data
- TO.FILE program for LAPI data
- MEDCR87.RAW.DUMP program for Medicare data
- DUMP program used numerous times on the hospital data

The ARF contractor should provide ODAM with a report justifying the continued use of SAS rather than COBOL for each process currently utilizing SAS. The report would list existing SAS processes and the number of times the process is repeated each year. Estimates of processing costs, including time-related costs, for both the mainframe and microcomputer environments would also be provided. The ARF contractor should also estimate the cost of developing COBOL equivalents and provide information on processing costs for both the mainframe and microcomputer COBOL versions.

Purpose: Over time, cost savings will be realized by making use of more efficient procedural languages.

2. Change: Modify the EXTRACT and UPDATE processes for selected data files by allowing multiple variables per record.

Purpose: Customizing the EXTRACT and UPDATE process for selected sources will result in significant cost savings.

c. Changes Affecting Manual Onerations

Until recently, data entry performed by the ARF contractor utilized simple text editor software. Quality controls such as range checks, inspection for spurious characters, and other editing were performed manually. The ARF contractor has embarked on utilizing custom data entry screens and checking routines to reduce the keying errors and minimize the reliance on manual checking operations. Large data entry operations have traditionally used such software. With the increased availability of low-cost data entry software for microcomputers, the ARF contractor would realize both a reduction in cost and improvements in data quality by automating the data entry with existing data entry software packages.

1. Change: The ARF contractor should provide ODAM with a report describing the manual data entry which is performed as part of the current system. For each data entry process, the ARF contractor should justify the use of customized SAS screens and menus rather than microcomputer data entry software packages such as QPL (Questionnaire Programming Language) or SPSS Data Entry II, a module of the SPSS/PC+ software system. Estimates of development and processing costs, including time-related costs, for each approach should be provided.

   Purpose: Low cost packages such as QPL contain useful features which would be expensive to duplicate in a customized system developed using SAS or some other procedural language. Using QPL may lower development costs and result in a more flexible data entry system.

d. The Expanded Use of Microcomputers

Suggested changes described elsewhere discuss the use of microcomputers in distributing documentation and in providing access to ARF data. The change described below addresses the use of microcomputers in the development, operation, and maintenance of the ARF data files.

1. Change: The ARF contractor is currently transferring a substantial portion of the ARF processing to microcomputers. The suggestion is that this transfer should continue.

   Purpose: This change should reduce costs and improve efficiency.

e. The Use of DBMS Software

The suggested change described below addresses the use of data base management system software as a tool in developing, operating, and maintaining the ARF system.

1. Change: The ARF contractor should provide ODAM with a system design providing the details of an ARF system utilizing DBMS software such as ORACLE, INGRESS, or Model 204. The report would describe the file structures, storage requirements, processing efficiency, and costs of using a DBMS as a tool in developing the ARF files.

   Purpose: The report is a necessary step in the transformation of the ARF system to take advantage of the efficiencies offered by DBMS software.

3. Costs and Assumptions

Estimated costs for the modifications related to system efficiency are shown in Table V.7. The assumptions used to derive the estimates are described below.

a. Changes Affecting System Maintenance

Expansion/Computerization of System Flowcharts. The ARF contractor would perform the following activities:

identify processes to flowchart
create actual flowcharts

Non-labor costs would include one-time software purchases, which will be under $500.

TABLE V.7

COST ESTIMATES FOR CHANGES INTENDED TO IMPROVE SYSTEM EFFICIENCY

| Activity | Labor Hours for ARF Contractor | | | | BHPr Hours | | Other Direct Costs |
|---|---|---|---|---|---|---|---|
| | Sys. Analyst | Programmer | Analyst | Clerical | Staff | Clerical | |
| a. Changes affecting system maintenance | | | | | | | |
| (i) Expansion/computerization of flowcharts | 20 | 60 | 0 | 0 | 0 | 0 | Under $500 |
| b. Changes affecting processing (CPU) efficiency | | | | | | | |
| (i) Report on expanded use of COBOL rather than SAS | 20 | 20 | 0 | 0 | 0 | 0 | Under $500 |
| (ii) Modify EXTRACT and UPDATE programs | 60 | 260 | 0 | 0 | 0 | 0 | $500 |
| c. Changes affecting manual operations | | | | | | | |
| (i) Report on use of data entry software package | 104 | 136 | 192 | 2.4 | 0 | 0 | $1000 |
| d. Expanded use of microcomputers | | | | | | | |
| (i) Continued transfer to microcomputer | (1) | (1) | (1) | (1) | (1) | (1) | (1) |
| e. DBMS software | | | | | | | |
| (i) System Design report on use of DBMS software | 110 | 140 | 80 | 0 | 0 | 0 | $1000 |

(1) Because the ARF contractor is performing this task under their current contract, costs were not estimated here.

b. Changes Affecting Processing (CPU) Efficiency

Expanded Use of COBOL Rather Than SAS. The ARF contractor would perform the following activities:

identify programs to recode

recode programs

document changes

Mainframe processing costs would not exceed $500.

Modify EXTRACT/UPDATE Programs for Selected Files. The ARF contractor would perform the following activities:

execute test runs to check if it is worthwhile to rewrite code
determine which merges to rewrite
determine best way to merge data for each source
write and execute merge
document the revised merge procedure

Mainframe processing costs would not exceed $500 per source. Merge procedures for two sources would be modified.c.Changes Affecting Manual Onerations

Use of Data Entry Software Packages. The ARF contractor would perform the following activities:

identify the specific data entry/cleaning operations which are currently being performed manually
develop criteria for converting process
apply criteria for converting process
specify the software requirements
obtain packages for test purposes
select package
prepare a final task report

Approximately $1,000 would be required for purchasing software for the evaluation. Other costs would be under $500.

d. The Expanded Use of Microcomputers

Continued Transfer of Processing to Microcomputers. No costs are estimated for this proposed change. This activity is currently being performed by the ARF contractor. Presumably, costs for this activity are included in the current ARF contract and, as a result, we anticipate no change in expected costs.

e. The Use of DBMS Software

Report on Using DBMS Software. The ARF contractor would perform the following activities:

access available software
select and purchase DBMS software
decide optimal number of files and file levels
determine which variables and sources are placed in each file
report on findings

Approximately $1,000 would be required for purchasing software for the evaluation. Other costs would be under $500.

4. Advantages and Disadvantages

a. Changes Affecting System Maintenance

Using EasyFlow 2 or a similar microcomputer package offers a cost-effective way for the ARF contractor to expand the number and detail of the system flowcharts and will foster updating the flowcharts as future processing proceeds. Maintaining the flowcharts in a central location will promote their use as a maintenance and development tool.

b. Changes Affecting Processing (CPU) Efficiency

Expanded Use of COBOL Rather Than SAS. MPR believes that the cost/time significant savings can be realized by utilizing procedural languages like COBOL rather than SAS. This is especially evident when working in a microcomputer environment. For example, MPR developed a simple, yet flexible, data dump program using a test extract of ARF data. Exhibit V-l provides a listing of the COBOL source code. Development costs on a microcomputer using Realia COBOL were minimal (less than one hour of programmer time). The compiled COBOL program generated data dumps in less time than it took for PC SAS to load into the microcomputer's memory. The most time-consuming aspect of using COBOL rather than SAS is the preparation of the FD (File Description). In many cases, the FD is required by later processing steps, such as the EXTRACT program.

Modify EXTRACT and UPDATE Programs for Selected Files. As microcomputers are used more extensively in ARF processing, CPU and Input/Output efficiency becomes more important. Modifications to the EXTRACT and UPDATE programs will result in meaningful time savings. For example, the transaction file for the county-level hospital file requires two tapes while the source data and the output from the UPDATE program each require only one tape. By converting the transaction file to include data for a county in the same record and eliminating unnecessary descriptive variables, the size of the transaction file would be reduced by approximately 50 percent and the number of records processed would drop substantially.

Exhibits V-2a and V-2b provide a comparison of the structure and size of the transaction files using the current and proposed format, The proposed structure results in a transaction file significantly smaller than the file using the current structure.

EXHIBIT V.1a

```
 IDENTIFICATION  DMSION.
 PROGRAM-ID. DUMP.
 AUTHOR. MATHEMATICA.
 ENVIRONMENT  DIVISION.
 CONFIGURATION  SECTION.
 SOURCE-COMPUTER. IBM-PC.
 OBJECT-COMPUTER. IBM-PC.
*
* PROGRAM TO DUMP FILE WITH 100 CHARACTERS PER LINE
*
 INPUT-OUTPUT  SECTION.
 FILE-CONTROL.
      SELECT TEST1 ASSIGN TO 'TEST1.DAT[N]'
          FILE STATUS IS TEST1-STATUS.
 DATA  DIVISION.
 FILE  SECTION.

FD TEST1
   LABEL  RECORDS  STANDARD
   RECORD CONTAINS 32000 CHARACTERS
   BLOCK CONTAINS 0 RECORDS.

01  TEST1-REC.
   05 TEST1-ARRAY        PIC X(100) OCCURS 320 TIMES.

 WORKING-STORAGE  SECTION.
 01 I                    PIC S9(9) COMP  VALUE  +o.
 01 J                    PIC S9(9) COMP VALUE +0.
 01 TEST1-COUNT          PIC S9(9) COMP  VALUE  +o.
 01 DISP-COUNT           PIC ZZZ,ZZ9.
 01 TEST1-STATUS         PIC X(2).
 01 NUMBER-TO-DUMP        PIC S9(9) COMP  VALUE  +10.
 01 LENGTH-RECORD         PIC S9(9) COMP VALUE +172.
 01 NUMBER-LINES          PIC S9(9) COMP  VALUE  +o.
 01 FROM-CC               PIC S9(9) COMP  VALUE  +o.
 01 TO-CC                 PIC S9(9) COMP  VALUE  +0.
 01 EOF-TEST1             PIC X(1) VALUE 'N'.
 01  RULER-LINE.

   05  RULER-LINE-l       PIC X(10) VALUE ' STARTING'.
   05  RULER-LINE-2       PIC X(50) VALUE
       '----+----1---- +----2 ____ + ____ 3----+----4----+----5'.
   05  RULER-LINE-3       PIC X(50) VALUE
       '----+----6----+ ____ 7 ____ + ---- 8 ____ +----9---- +----0'.
   05  RULER-LINE-4       PIC X(10) VALUE ' ENDING'.
```

EXHIBIT V.la (continued)

```
PROCEDURE DIVISION.
START-UP.

    OPEN INPUT TEST1.
    DISPLAY 'Enter number of records to dump?'
    ACCEPT NUMBER-TO-DUMP.
    DISPLAY 'Enter record length?'
    ACCEPT LENGTH-RECORD.
    COMPUTE NUMBER-LINES = ((LENGTH-RECORD - 1) / 100) + 1.
    PERFORM MAIN-LOOP THRU MAIN-LOOP-EXIT
          VARYING I FROM +1 BY +1 UNTIL
          I > NUMBER-TO-DUMP OR EOF-TEST1 = Y'.
    PERFORM EOJ.

MAIN-LOOP.
    READ TEST1 AT END
       MOVE 'Y' TO EOF-TEST1.
    IF EOF-TEST1 = 'N'
      ADD +1 TO TEST1-COUNT
      MOVE TEST1-COUNT TO DISP-COUNT
      DISPLAY ' RECORD #' DISP-COUNT
      DISPLAY RULER-LINE
      PERFORM PRINT-DUMP THRU PRINT-DUMP-EXIT
          VARYING J FROM +1 BY +1 UNTIL
          J > NUMBER-LINES.

MAIN-LOOP-EXIT.
   EXIT.

PRINT-DUMP.
   COMPUTE FROM-CC = (J - 1) * 100 + 1.
   COMPUTE TO-CC = J * 100.
   IF J = NUMBER-LINES
         MOVE LENGTH-RECORD TO TO-CC.
   DISPLAY FROM-CC '' TEST1-ARRAY(J)'' TO-CC.

PRINT-DUMP-EXIT.
   EXIT.

EOJ.
   EXHIBIT NAMED TEST1-COUNT.
   CLOSE TEST1.
   STOP RUN.
```

EXHIBIT V.lb

```
RECORD #        I
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0 ENDING
000000001 12001FLA000090111838018050004800881101491600911700660000971000632500061000430008100088 12001FLA000090  000000100
000000101 111838018050004800881101491600911700660000971000632500061000430008100088                                 000000172

 RECORD #        2
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0-0 ENDING
000000001 12003FLA000058509909001820000000076400045200076100082800128000079500007000020000000000012003FLA000058 000000100
000000101 509909001820000000076400045200076100082800128000079500007000020000000000                                 000000172

 RECORD #        3
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0-0 ENDING
000000001 12005FLA000075811223012230000900445700532800493600420800785100619800014000060001300008 12005FLA000075 000000100
000000101 811223012230000900445700532800493600420800785100619800014000060001300008                                 000000172

 RECORD #        4
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0-0 ENDING
000000001 12007FLA000029308443002450000000083500089300086900094500149300122400005000000000300000 12007FLA000029 000000109
000000101 308443002450000000083500089300086900094500149300122400005000000000300000                                 000000172

 RECORD #        5
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0-0 ENDING
000000001 12009FLA000099513827037490003201212601254001198301115402125801941700076000510003800026 12009FLA000099 000000100
000000101 513827037490003201212601254001198301115402125801941700076000510003800026                                 000000172

 RECORD #        6
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0---0 ENDING
000000001 12011FLA000121118128116260012503160603897404264004019106594505178400228001410026000124 12011FLA000121 000000100
000000101 118128116260012503160603897404264004019106594505178400228001410026000124                                 000000172

 RECORD #        7
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0---0 ENDING
000000001 12013FLA000056807290000980000000036600043600028600026900060200065300004000020000000000012013FLA000056 000000100
000000101 807290000980000000036600043600028600026900060200065300004000020000000000                                 000000172

 RECORD #        8
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0 ENDING
000000001 12015FLA000069013732008910000600148500160600172100168900272600370400016000050001500004 12015FLA000069 000000100
000000101 013732008910000600148500160600172100168900272600370400016000050001500004                                 000000172

 RECORD #        9
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0---0 ENDING
000000001 12017FLA000062910919008750000300190700164800170100168600333600398300018000110000800001 12017FLA000062 000000100
000000101 910919008750000300190700164800170100168600333600398300018000110000800001                                 000000172

 RECORD #       10
 STARTING ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0-0 ENDING
000000001 12019FLA000059213526009700001200339000237400361500484300734300377100023000160001000008 12019FLA000059 000000100
000000101 213526009700001200339000237400361500484300734300377100023000160001000008                                 000000172
```

EXHIBIT V.2

Format of Current Transaction File

| | Transid | Field* | Start* | Length* | Description* | Data |
|---|---|---|---|---|---|---|
| Size | 5 | 7 | 5 | 2 | 37 | 35 |
| | 1 | VAR1 | 100 | 5 | descip. 1 | For 1,1 |
| | 1 | VAR2 | 105 | 5 | descip. 2 | For 1,2 |
| | 1 | VAR3 | 110 | 7 | descip. 3 | For 1,3 |
| | 1 | vAR4 | 117 | 3 | descip. 4 | For 1,4 |
| | 1 | VAR5 | 120 | 8 | descip. 5 | For 1,5 |
| | 2 | VAR1 | 100 | 5 | descip. 1 | For 2,1 |
| | 2 | VAR2 | 105 | 5 | descip. 2 | For 2,2 |
| | 2 | VAR5 | 120 | 8 | descip. 5 | For 2,5 |
| | | Length of record | | 91 | File size 728 | |

Format of Proposed Transaction File

| | TRANSID | VAR1 | VAR2 | VAR3 | vAR4 | VAR5 |
|---|---|---|---|---|---|---|
| Size | 5 | 5 | 5 | 7 | 3 | 8 |
| | 1 | For 1,1 | For 1,2 | For 1,3 | For 1,4 | For 1,5 |
| | 2 | For 2,1 | For 2,2 | missing | missing | For 2,5 |
| | | Length of record | | 33 | File size 66 | |

*These fields are no longer necessary, and will be incorporated into the customized program.

c. Changes Affecting Manual Operations.

QPL, a public domain software system developed by the General Accounting Office, provides a set of powerful tools for entering, manipulating, and printing data. The system contains routines which are relevant to data entry within the ARF system:

KEYPUNCH - This routine simplifies entering new data or editing previously entered records. Multiple data fields appear on the screen. The program will check for many types of errors as the data are entered. After adding or editing a record, the user presses an ADD key to update the main data file. Data is stored as a flat file with fixed length fields.

The routine also makes it easy to review records. Records may be retrieved sequentially or through the use of search criteria. KEYPUNCH also contains a print routine for use in manual data verification. Data is printed with variable names and in a descriptive format (e.g. "Yes" answers would appear as 'Yes" rather than a numeric code).

Options allowing the user to control the format and characteristics of the screen, such as color and highlighting, are built into the routine.

COLLECT - This routine is functionally similar to KEYPUNCH except that data fields appear on separate screens. The user may page forward or backwards to review and correct entered data. The data file is automatically updated after information is entered into the last data field. The routine also incorporates built-in editing features. Data may be stored in a flat file with fixed fields, a flat file with comma separated fields, or as two files -- one with fixed and one with comma separated fields. The comma separated files enable convenient translation into other software packages such as dBase III or Lotus 123.

TOSAS - This routine takes a flat file in fixed format and generates the SAS program statements needed by PC SAS to create a SAS dataset. Similar programs which generate SPSS and C language code are also available.

Using QPL may lower development costs and result in a more flexible data entry system.

d. The Expanded Use of Microcomputers.

There are numerous advantages to doing processing on the PC: no cost associated with each run; immediate completion for small jobs; more user-friendly software to facilitate most any task; and simpler processes for viewing samples of data. On the other hand, microcomputers

have certain costs associated with their use not common to mainframes: only one task may be executed at a time; sorting can require a great deal of time and storage; and Input/Output-intensive jobs can take hours instead of minutes. These two points emphasize the importance of tailoring the application to the microcomputer rather than merely duplicating the mainframe processing on the PC.

The ARF contractor should be aware of the relative efficiency considerations of microcomputers and mainframes. Other recommendations (converting SAS jobs to COBOL and creation of a DBMS) address this difference. The contractor should utilize this conversion as an opportunity to optimize the procedures for the microcomputer.

e. The Use of DBMS Software.

As more ARF processing is transferred to microcomputers, processing efficiency and storage requirements become more important. Based on our review of the current ARF system (described in Chapter III) and our experience with developing and manipulating similar, large data files, MPR believes the DBMS approach offers the possibility of worthwhile storage and time savings.

Restructuring the ARF system to incorporate DBMS software will not be a simple task. ODAM and the ARF contractor, because of their intimate knowledge with the ARF, are in the best position to produce the first step in the process, a report outlining a detailed system design.

In investigating the possibility of using database management system software, the system design report should examine both the possible DBMS software and the file structures. Many software packages such as ORACLE, INGRESS, and Model 204 have both mainframe and PC versions. The structure of the files (called tables) and the variables (called columns) they include determine the efficiency of the database. In determining which variables to include on a table,

the report should address the often conflicting efficiency considerations. The report should detail the tradeoff between space, processing time, and retrieval time.

Currently, each time the ARF contractor adds or modifies a variable on the basic ARF, they must read and write the entire file (about 90 MB). The report may find it is more efficient to store the data in separate files, organized by source and subject. To add a new variable, the contractor would add the variable to one of the much smaller subfiles. In addition, the subfiles could dramatically reduce storage requirements (especially important for microcomputer processing) by eliminating records for counties where data may be deficient.

## F. CHANGES AFFECTING ACCESS TO AND DELIVERY OF ARF DATA

A recurring issue in our discussions with BHPr and other users was ease of use of the ARF data file. Some users felt that changes to the ARF documentation would enable them to utilize the ARF file more easily and more frequently. Suggestions related to ARF documentation were discussed above.

Other users felt that modifications to format and media would simplify use of the file. Providing a magnetic tape with data pre-formatted for use with analysis packages, such as SAS, would minimize a researcher's reliance on scarce technical programming resources. Other users expressed interest in using optical technologies, which enable users to bypass mainframe processing and the need for programmer resources.

Users and BHPr also felt that the use of alternative data access techniques would facilitate use of the data. Some users felt that DBMS software, operating on either microcomputers and mainframes, would provide an easy-to-use vehicle for non-programmers to access ARF data. Others were interested in using microcomputer-based bulletin boards for ARF data dissemination.

1. Problem

The suggestions offered addressed the following main issues:

Was there a better way to distribute copies of the ARF data file?
Was there a way to provide for more convenient computerized access to individual data elements and observations?
Was there a way to provide for more convenient computerized access to ARF data for use on microcomputers?

2. Explanation and Purpose of Changes

Each of the main issues listed above are addressed by the following specific proposed changes:

a. Manual (Non-computerized) Delivery Systems

1. Change: Provide ARF data on magnetic tape in SAS format

Purpose: In addition to providing data in character format, the basic ARF file would be made available for direct use by SAS. The ARF system would be modified by adding a step at the end of the processing stream in which the character file would be transformed into SAS format. The SAS format file would contain SAS labels for each variable. By providing the file in SAS format, non-programmers would be able to access the file more easily. SAS procedures could be used without requiring DATA steps to transform the character data.

2. Change: Distribute ARF data on Optical Disk

Purpose: Providing the ARF on optical disks will allow the microcomputer user to access the complete file without having to resort to downloading data or handling magnetic tapes.

b. Computerized Access to Individual Data Elements and Observations

**As discussed in Chapter III, the mainframe ARF** ACCESS **system provides an user-**interface to the existing basic ARF file. The ACCESS system contains a set of simple menus which prompt users for the variables and counties for which they would like data. Once

extracted from the full ARF file, the data is converted to SAS format for additional analysis and presentation.

The ACCESS system has not been used extensively outside of ODAM. To encourage direct computerized access of individual data elements and observations, several suggested changes were identified:

1. Change: Provide additional and regular training for the existing ACCESS system (Mainframe - Flat File/ACCESS Software). Up-to-date documentation and training manuals would be maintained.

   Purpose: ODAM and the ARF contractor would conduct regularly scheduled training sessions in the use of the existing ACCESS system. This training would refresh the skills of users who received ACCESS training several years ago and would provide an introduction to the system for new staff.

2. Change: Enhance the current ACCESS System (Mainframe - Flat File/Enhanced Menu-Driven Software). Additional menus would be added to the ACCESS system. The user would be able to point to the desired area and variables they would like to extract. The system would incorporate help screens describing the definitions and other characteristics of the variables. The system would also be expanded to generate both structure extended and comma separated extracts to facilitate transfer into dBase III and other microcomputer software packages. The expanded menus would be developed using the same procedural languages as the existing ACCESS System.

   Purpose: This would provide a more intuitive and easier-to-use interface. The ability to quickly obtain extracts of the ARF file without having to invest substantial amounts of time in learning the intricacies of the ARF file will facilitate use of ARF by casual users; i.e., those users who require ARF data on an irregular or infrequent basis.

3. Change: Replace the existing mainframe ACCESS System with a system utilizing Data Base Management System software (Mainframe - DBMS/Non-Menu-Driven Software). The new system would provide interactive access to on-line data, in contrast to the existing batch-oriented, flat-file ACCESS System. Users would be presented with relatively simple menus developed using the DBMS language.

   Purpose: This change would accomplish many of the same goals as the existing ACCESS System. However, replacing the current system with a

DBMS system offers reduced storage and simplified maintenance and expansion.

4. Change: Replace the existing mainframe ACCESS System with a system utilizing Data Base Management System software and an enhanced menuing system (Mainframe - DBMS/Menu-Driven Software). Similar to Change 3 listed above, the new system would provide interactive access to on-line data, in contrast to the existing batch-oriented, flat-file ACCESS System. However, users would be presented with an enhanced menu system allowing pointing and access to complete help screens.

   Purpose: This change would accomplish many of the same goals as the existing ACCESS System. However, replacing the current system with a DBMS system offers reduced storage and simplified maintenance. As in Change #2, the enhanced menu system -would facilitate use by infrequent users.

5. Change: Transfer the existing ACCESS System to a microcomputer and provide additional training.

   Purpose: This change is identical to Change #1 except that the ACCESS system will reside on a microcomputer rather than the mainframe. Mainframe operating costs would be eliminated. Savings would be offset in part by microcomputer and software acquisition costs.

6. Change: Transfer the existing ACCESS System to a microcomputer and incorporate an enhanced menu system.

   Purpose: This would provide the same advantages as Change #2 but in a microcomputer environment.

7. Change: Transfer the existing ACCESS System to a microcomputer-based DBMS.

   Purpose: This would provide the same advantages as Change #3 but in a microcomputer environment.

8. Change: Transfer the existing ACCESS System to a microcomputer-based DBMS and implement an enhance menu system.

   Purpose: This would provide the same advantages as Change #4 but in a microcomputer environment.

c. Computerized Access to ARF Data Files

Some of the changes described above address the format and media used to provide users with physical copies of the ARF data. Other changes dealt with techniques which would provide users with direct, computerized access to individual records or data elements. Another approach is to distribute extracts of the ARF data through telecommunications.

1. Change: Install a microcomputer-based bulletin board containing ARF data extracts. The ARF contractor would prepare a set of data extracts containing information from the ARF. The data extracts would be stored in condensed, character format. Microcomputer users would telephone the bulletin board, select the file containing the desired data, and download the file to their own microcomputer. Users would also be supplied with a program which would convert the character file to dBase format. The bulletin board software would control access to the data files. The bulletin board would also be used to transmit messages between ODAM, the ARF contractor, and users.

   Purpose: Using a bulletin board, users would copy complete data extracts and, once the data is on their own PC, users would be able to discard unneeded records and variables. File content would be determined by ODAM and the ARF contractor.

3. Costs and Assumptions

Estimated costs for the modifications related to system efficiency are shown in Table V.8. The assumptions used to derive the estimates are described below.

a. Manual (Non-computerized) Delivery Systems

Provide ARF data on magnetic tape in SAS format. would perform the following activities:

write and execute a SAS conversion program
prepare documentation (including executing a PROC CONTENTS)

The expected computer costs of creating the SAS file is under $1,000 per year.

TABLE V.6

COST ESTIMATES FOR CHANGES AFFECTING ACCESS TO AND DELIVERY OF ARF DATA

| Activity | Labor Hours for ARF Contractor | | | | BHPr Hours | | Other Direct Costs |
|---|---|---|---|---|---|---|---|
| | Sys. Analyst | Programmer | Analyst | Clerical | Staff | Clerical | |
| **FIRST YEAR DEVELOPMENT/IMPLEMENTATION** | | | | | | | |
| a. Manual (Non-computerized delivery systems) | | | | | | | |
| (i) Provide data in SAS format | 0 | 16 | 16 | 0 | 16 | 0 | Under $500 |
| (ii) Distribute ARF data on optical disk | 30 | 200 | 30 | 0 | 20 | 0 | $15,000 |
| b. Computerized access to individual data elements and observations | | | | | | | |
| (i) Provide training/maintain existing ACCESS System | 88 | 96 | 104 | 24 | 0 | 0 | Under $500 |
| (ii) Provide training/expand existing mainframe ACCESS system | 368 | 568 | 336 | 24 | 0 | 0 | $7,500 |
| (iii) Provide training/replace Access System with mainframe DBMS | 400 | 464 | 368 | 24 | 0 | 0 | $8,600 |
| (iv) Provide training/mainframe DBMS with enhanced menu system | 568 | 688 | 528 | 24 | 0 | 0 | $12,200 |
| (v) Provide training/transfer current access system to microcomputer | 136 | 296 | 160 | 24 | 0 | 0 | $7,000 |
| (vi) Provide training/microcomputer ACCESS with enhanced menu system | 416 | 768 | 392 | 24 | 0 | 0 | $7,000 |
| (vii) Provide training/microcomputer DBMS | 408 | 504 | 384 | 24 | 0 | 0 | $8,000 |
| (viii) Provide training/microcomputer DBMS with enhanced menu system | 576 | 728 | 344 | 24 | 0 | 0 | $8,000 |
| c. Computerized access to ARF data fiks | | | | | | | |
| (i) Bulletin board | 232 | 608 | 192 | 16 | 0 | 0 | $12,000 |

TABLE V.8

COST ESTIMATES FOR CHANGES AFFECTING ACCESS TO AND DELIVERY OF ARF DATA
(continued)

| Activity | Labor Hours for ARF Contractor | | | | BHPr Hours | | Other Direct Costs |
|---|---|---|---|---|---|---|---|
| | Sys. Analyst | Programmer | Analyst | Clerical | Staff | Clerical | |
| ON-GOING ANNUAL OPERATION | | | | | | | |
| a. Manual (Non-computerized delivery systems) | | | | | | | |
| (i) Provide data in SAS format | 0 | 16 | 16 | 0 | 16 | 0 | Under $500 |
| (ii) Distribute ARF data on optical dish | 200 | 350 | 160 | 120 | 150 | 0 | $30,000 |
| b. Computerized access to individual data elements and observations | | | | | | | |
| (i) Provide training/maintain existing ACCESS System | 128 | 304 | 192 | 16 | 0 | 0 | $6,000 |
| (ii) Provide training/expand existing mainframe ACCESS system | 156 | 432 | 176 | 16 | 0 | 0 | $10,200 |
| (iii) Provide training/replace Access System with mainframe DBMS | 120 | 200 | 152 | 16 | 0 | 0 | $15,000 |
| (iv) Provide training/mainframe DBMS with enhanced menu system | 120 | 248 | 152 | 16 | 0 | 0 | $15,000 |
| (v) Provide training/transfer current access system to microcomputer | 136 | 344 | 208 | 16 | 0 | 0 | $500 |
| (vii Provide training/microcomputer ACCESS with enhanced menu system | 164 | 472 | 192 | 16 | 0 | 0 | $500 |
| (vii) Provide training/microcomputer DBMS | 128 | 240 | 168 | 16 | 0 | 0 | $500 |
| (viii) Provide training/microcomputer DBMS with enhanced menu system | 128 | 280 | 168 | 16 | 0 | 0 | $500 |
| c. Computerized access to ARF data files | | | | | | | |
| (i) Bulletin board | 48 | 132 | 40 | 8 | 0 | 0 | $500 |

Distribute ARF data on Optical Disk. The ARF contractor would perform the following activities:

develop a marketing strategy (including production promotion package and distribution system)
establish training and technical support
deliver CD-ROM application to end-users
define user interface requirements
define delivery system
analyze database characteristics and select index/retrieval software
create custom user interface and documentation
index data
format software and data
simulate CD-ROM access/retrieval - premaster formatted data
arrange for master/replicate CD-ROM discs

The cost would include $10,000 for access/retrieval software, $2,000 for hardware, and $15,000 for the creation of the master.b. Computerized Access to Individual Data Elements and Observations

Training activities will be required regardless of the actual access vehicle selected. Training materials will need to be prepared and training classes conducted.

In addition to providing training, the ARF contractor will be responsible for the development, implementation, and on-going maintenance (correcting errors, making minor changes, and performing backups) for each access option:

Maintain existing ACCESS System. The following activities are assumed during the development/implementation phase:

Review, update, and expand database loading documentation
Review and test current database load procedures
Review, update, and expand ACCESS System documentation
Review and test current ACCESS System

Each year, the following activities are assumed:

Load revised database
Perform periodic testing
Make minor modifications to the ACCESS System

Non-labor costs associated with development and implementation are assumed not to exceed $500. Non-labor costs experienced each year are assumed to be:

$3,000 for database storage (MSS storage at the NIH Computer Center)
$1,000 for CPU/Storage for periodic testing
$2,000 for CPU/Storage for incorporating minor modifications/changes
Under $500 for other direct costs

Existing ACCESS System with enhanced menu system. The following activities are assumed during the development/implementation phase:

Review, update, and expand database loading documentation
Review and test current database load procedures
Specify requirements for enhanced menu system
Specify design for enhanced menu system
Code, test, install, and document the enhanced menu system

Each year, the following activities are assumed:

Load revised database
Perform periodic testing
Modify/change the enhanced menu system

Non-labor costs associated with development and implementation are estimated at:

$1,500 for CPU/Storage during coding
$5,000 for CPU/Storage during testing
$1,000 for CPU/Storage during installation

Annual non-labor costs are estimated at:

$3,000 for database storage (MSS storage at the NIH computer center)
$1,200 for CPU/Storage for periodic testing
$6,000 for CPU/Storage for implementing minor modifications/changes

MPR believes that developing and maintaining an enhanced menu system written in SAS or a procedural language will require substantial resources.

Replace with mainframe DBMS. The following activities are assumed during the development/implementation phase:

Specify database design
Develop and test database cleaning/load procedures
Clean/load database
Prepare database documentation
Specify access system requirements
Specify access system design
Code, test, and install access system
Document access system

Each year, the following activities are assumed:

Load revised database
Perform periodic testing
Implement minor modifications/changes

Non-labor costs associated with development/implementation are:

$4,000 for CPU/Storage during development/testing of cleaning/load procedures
$1,000 for cleaning/loading the database
$3,600 for coding/testing/installing access system

Annual non-labor costs are:

$12,000 for database storage (on-line storage at the NIH computer center)
$1,000 for periodic testing
$2,000 for implementing minor modifications/changes

Cleaning refers to preparing the basic ARF tile for loading into the DBMS. The menu system is assumed to be roughly equivalent to the current ACCESS system.

Provide training/replace with mainframe DBMS and enhanced menus. The activities to be performed are the same as above, except that an enhanced menu system will be developed. This is reflected in an increase in the professional hours required. Non-labor access system coding/testing/installation costs are estimated at $7,200 rather than $3,600. Other non-labor costs are assumed to be the same as above.

Transfer existing ACCESS System to microcomputer. The assumptions are similar to those described above, except for the additional tasks of transferring the software to the microcomputer and, if necessary, obtaining a microcomputer for use as a dedicated ACCESS machine.

The following activities are assumed during the development/implementation phase:

Modify the existing system for use on the microcomputer
Review, update, and expand database loading documentation
Review and test database load procedures
Review, update, and expand ACCESS System documentation

Each year, the following activities are assumed:

Load revised database
Perform periodic testing
Make minor modifications to the ACCESS System

Non-labor costs associated with development and implementation are associated with obtaining a microcomputer for use as an ACCESS machine:

$6,000 for a microcomputer with 300MB hard disk
$1,000 for language compilers and miscellaneous software

Non-labor costs experienced each year are assumed to be:

$500 for hardware maintenance

Microcomputer ACCESS System with enhanced menus. This is similar to the task described above except that the ACCESS System resides on a microcomputer. During development/implementation, the following activities are assumed:

Modify the existing system for use on the microcomputer
Review, update, and expand database loading documentation
Review and test database load procedures
Review, update, and expand ACCESS System documentation
Design, code, test, and install the enhanced menu system

Each year, the following activities are assumed:

Load revised database
Perform periodic testing
Modify/change the enhanced menu system

Non-labor costs associated with development and implementation are associated with obtaining a microcomputer for use as an ACCESS machine:

$6,000 for a microcomputer with 300MB hard disk
$1,000 for language compilers and miscellaneous software

Non-labor costs experienced each year are assumed to be:

$500 for hardware maintenance

Microcomputer DBMS with simple access system. The following activities are assumed during the development/implementation phase:

Specify database design
Transfer database to microcomputer
Develop and test database cleaning/load procedures
Clean/load database
Prepare database documentation
Specify access system requirements
Specify access system design
Code, test, and install access system
Document access system

Each year, the following activities are assumed:

Transfer/Load revised database
Perform periodic testing
Implement minor modifications/change

Non-labor costs associated with development/implementation are:

$6,000 for a microcomputer with 300MB hard disk
$1,000 for language compilers and miscellaneous software

Non-labor costs experienced each year are assumed to be:

$500 for hardware maintenance

Microcomputer DBMS with enhanced menu. The activities to be performed are the same as above, except that an enhanced menu system will be developed. This is reflected in an increase in the professional hours required. Non-labor costs are the same.

c. Computerized Access to ARF Data Files

Microcomputer Bulletin Board. The following activities are assumed during the development/implementation phase:

Specify the functional requirements of the bulletin board system (BBS)
Select BBS software
Specify/obtain hardware (including telephone lines)
Install hardware and software
Test hardware and software
Document the BBS and specify maintenance procedures
Design the BBS database
Create the BBS database of the mainframe
Transfer the BBS database to the microcomputer
Convert the BBS database to dBase format
Install the BBS database in the BBS software
Document the BBS database and specify update procedures

Each year, the following activities are assumed:

Create the revised BBS database on the mainframe
Transfer the revised BBS database to the microcomputer
Convert the revised BBS database to dBase format
Install the revised BBS database on the BBS
Document the revised BBS database

Non-labor costs during the development/implementation phase are hardware-related:

$6,000 for microcomputer with 300MB hard disk
$4,000 for modems (4 modems at $1000/each)
$1,000 bulletin board software
$1,000 Telephone installation (4 lines at $250/each)

Annual non-labor costs are assumed to be:

$500 hardware maintenance
$2,400 telephone charges (4 lines at $50/month)

4. Advantages and Disadvantages

a. Manual (Non-computerized) Delivery Systems

Provide Data in SAS Format. Transforming the data into SAS format prior to distribution allows individual users to immediately use SAS PROCs. A file containing the instructions necessary to transform ARF into SAS format is currently available as part of the documentation package. As a result, it is relatively easy for a programmer to modify this file to create a program to generate a customized SAS data extract.

SAS files often require more space than their character equivalents. The blocksize limit may be reached, requiring the file to be split. More than one physical tape volume may also be required.

Distribute the ARF Data on Optical Disk. The large storage capacity offered by optical disks would enable researchers to directly access the entire ARF without using any mainframe facilities. Available optical disk software also includes extensive search and display capabilities. Many packages allow users to generate data extracts for use with analysis software such as dBase or PC-SAS.

Optical data storage is an emerging technology with some incompatibilities between different systems. The technology is also relatively expensive and has not been met with widespread acceptance in the user community.

b. Computerized Access to Individual Data Elements and Observations

Expanded training would refresh the skills of users who received ACCESS training several years ago and would introduce new staff to the ARF. Training activities would provide these benefits regardless of the vehicle used to provide computerized access.

The current ACCESS system provides NIH computer users with a relatively simple method of obtaining ARF data in SAS format. Compared to other access vehicles suggested, it has the advantage that no additional development costs are needed. Its disadvantages are that the user must identify the desired variables using the Fnumber and that users must have NIH computer accounts.

Enhancing the menus makes it easier to access the ARF data. Creating and maintaining these enhanced menus can be costly, particularly when using procedural languages like SAS or COBOL.

DBMS software offers reduced storage requirements and lower maintenance and development costs. These advantages are offset by the cost of on-line storage. The use of microcomputers rather than mainframes will require acquisition of special hardware. Hardware costs would be roughly equal to the cost of one year of mainframe on-line storage. The use of microcomputer offers the future opportunity of providing direct access to data elements and individual records for users outside of the federal government.

c. Computerized Access of ARF Data Files

The use of a bulletin board offers a compromise between the distribution of physical copies of the ARF data and approaches which allow direct access to individual records and data elements. Transfers from the mainframe would no longer be necessary. Users would be able to use the dBase files immediately or translate them into Lotus or PC-SAS format. No special programming would be required by researchers.

A dedicated bulletin board will require hardware expenditure and some continuing maintenance. An alternative would be to provide the files on a commercial bulletin board system.

# CHAPTER VI.

## ARF IN THE 1990s: SUMMARY AND RECOMMENDATIONS

The preceding sections of this report de-scribed MPR's evaluation of the ARF System. MPR addressed many aspects of the ARF system, ranging from system efficiency to user perceptions to possible uses of alternative technologies. The scope of MPR's evaluation reflected MPR's contractual requirements as well as suggestions offered by BHPr, end-users outside of BHPr, and MPR's own staff. Perceived problem areas were identified and specific activities designed to address these problem areas were discussed.

There are certain areas of the ARF system where few, if any, problems were identified. Other aspects are viewed by BHPr, end-users, or MPR as more problematic. A review of the identified problem areas and proposed responses shows a wide variety of activities which should be considered when preparing future ARF riles. Some of the activities are straight forward and could be implemented quickly and inexpensively. Others will require major modifications to existing ARF System procedures at considerable expense. Many of the modifications were contradictory; implementing one modification may have the effect of making another irrelevant. Other modifications complement each other and should be implemented in tandem.

The wide range of suggested activities reflects not only the complexity of the ARF system but also the variety of opinions within BHPr and the user-community about the purpose and quality of the current ARF system and data files. Prior to committing additional resources to the ARF system, MPR feels that BHPr, as sponsoring agency for ARF, must make some fundamental decisions concerning the goals of the ARF in the 1990's. Specifically, BHPr must first decide:

1)Who is the ARF being prepared for?

The answer to this question helps determine the answer to a second fundamental question:

2)What should the ARF be?

Once these decisions are made, it is reasonable to ask:

3)Is ARF being prepared in the most cost-effective/efficient manner?

BHPR goals for ARF in the 1990s, plus MPR's specific recommendations for achieving these goals, are discussed below.

## A. BHPr GOALS FOR ARF

The specific recommendations described below are appropriate only within the context of the overall goals of the ARF System. Therefore, prior to implementing any of these modifications, BHPr must make some decisions about the basic goals of the ARF system.

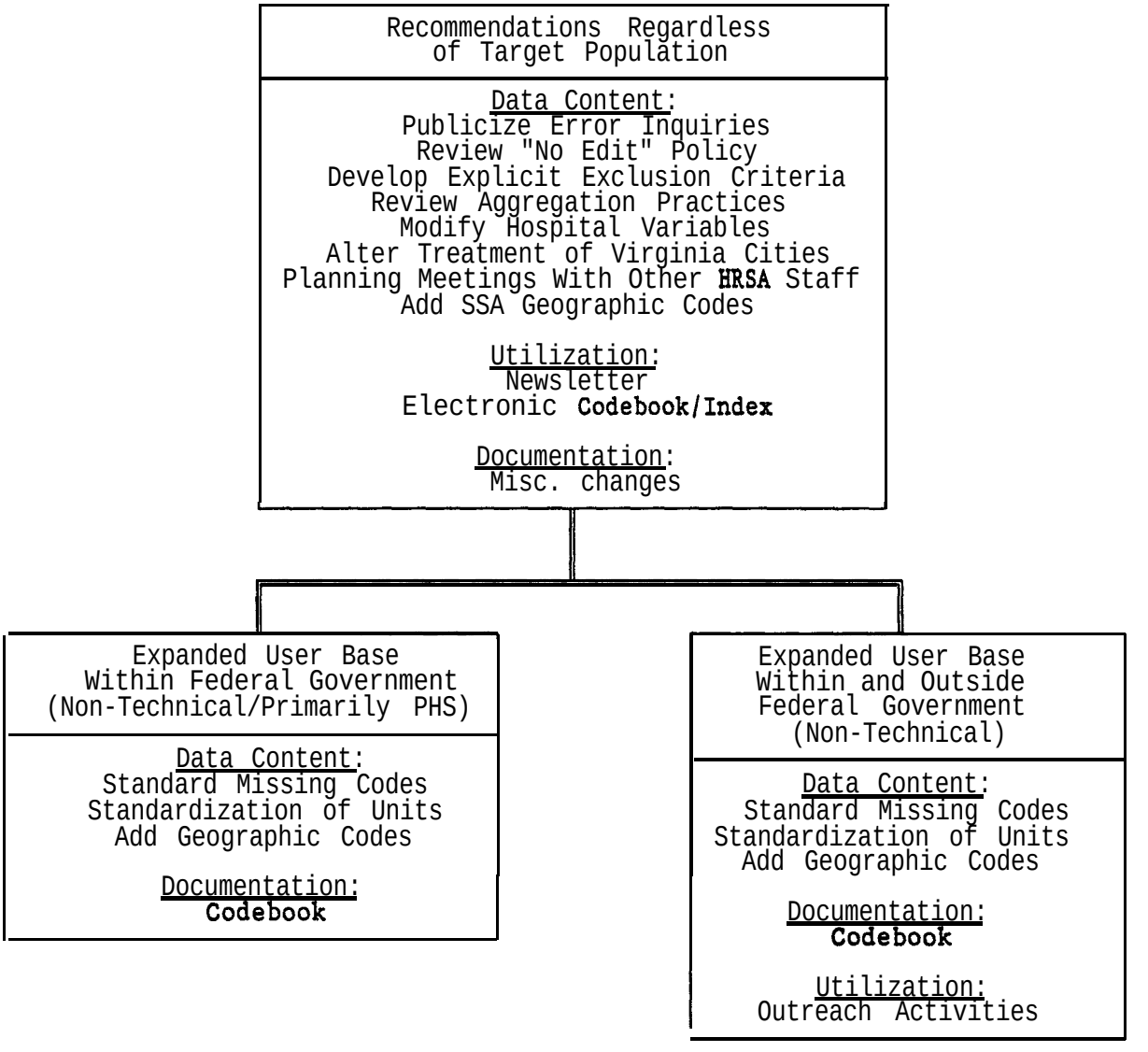
Who is the ARF file being prepared for?

The basic issue BHPr must address concerns the target population of users for the main ARF file. Is the ARF intended for ODAM, other users within BHPr, users outside of BHPr, or some combination of these groups? The answer to this question has a direct bearing on whether some of the recommendations listed below should be implemented. The relationships between the recommendations and BHPr's target population decision are summarized in Exhibit VI.1.

ODAM prepared the ARF originally for its own use, but subsequently made the file available to others both within and outside of BHPr. It is not possible to be sure how many of these outside users there are. Some researchers have used the file once for a specific project; others use the file repeatedly. For some researchers, the ARF was used as a primary data

EXHIBIT VI.1

RECOMMENDATIONS RELATED TO ARF TARGET POPULATIONS



source; for others, the file was a secondary source used to validate information gathered elsewhere. Other researchers, after obtaining the file, decided that the ARF was inappropriate for their needs.

We do know that about 40 copies of the ARF are distributed outside of the Public Health Service annually. Within the Public Health Service, but outside of ODAM, there are approximately ten users per year.

It may be possible to expand the ARF user base, but to do this, the difficulties mentioned by some users outside of ODAM would have to be addressed. In addition to issues related to data quality, content, and timeliness, some users felt that the ARF was difficult to use. Technical programming resources were required to access the file. Confusion related to documentation content and documentation organization were also cited as impediments to using the ARF. Users with the technical programming capabilities and experience using large data files were most likely to use the ARF successfully. These technically proficient users were most likely to access the source data directly and utilize ARF as a secondary data source.

If BHPr desires that the ARF's intended target population should be ODAM and selected data processing-proficient non-ODAM users, only a few changes to the ARF system are required.

If, on the other hand, BHPr feels that the non-programmer research community, both within and outside of BHPr, should utilize ARF more extensively, more substantial changes to the ARF system are required. To promote utilization by these users, BHPr should instruct ODAM and the ARF contractor to engage in activities related to information dissemination and outreach, documentation content and organization, and data access and delivery mechanisms. Specific recommendations for each of these subject areas are described below in Section C.

What should ARF be?

Based on MPR's communications with ODAM and MPR's discussions with other BHPr staff and users outside of BHPr, there appear to be disagreements or misunderstandings about what the Area Resource File is or should be.

The current main Area Resource File could be considered to be a consolidated data bank containing data from several sources. The value added by ODAM and the ARF contractor is primarily in consolidating sub-county data to the county level and in merging data from several sources into a single physical data file. Generally, source data is accepted "as is".

Some BHPr and outside users desire a more integrated research database. Such a database would require a shift from the "as is" philosophy to a more active role in manipulating the data prior to inclusion in the main ARF file. An integrated research database would strive to implement consistent treatment of data elements regardless of the source, but need not regard source data as "absolutely clean". Whether an integrated research database is necessary depends upon the target user population.

If BHPr is satisfied with serving the current user base, little or no change to the ARF system and procedures is required. If BHPr decides to expand the ARF user base, BHPr must also shift the emphasis of the ARF more towards the integrated research approach. BHPr would need to instruct ODAM and the ARF contractor to modify the existing ARF system and procedures. Many of the specific recommendations related to data content described below are designed to transform the main ARF into a more integrated research database.

Regardless of BHPr's decision concerning the future orientation of the ARF system, ODAM and the ARF contractor should seriously consider implementing several recommendations

dealing with ARE data content. MPR feels that these changes would facilitate ODAM's use of the file (see Section B).

Is the work being performed efficiently?

Irrespective of BHPr's decisions concerning the direction of ARE in the 1990s, MPR has identified several areas where processing efficiencies or cost-reductions may be realized. These recommended changes are relatively minor in nature and are unlikely to result in substantial savings. However, they would streamline some of the processing procedures or simplify future system maintenance. These specific efficiency recommendations are described in Section F.

## B. RECOMMENDATIONS: CURRENT USER BASE

### 1. Data Content

The ARF appears to be more or less what its current users need. It was clear from the interviews that most users were reasonably satisfied with the ARF's content. Naturally, some users could think of improvements, and some had specific suggestions for new data that should be added. These suggestions may be considered as marginal improvements, inasmuch as there were no major changes in content requested.

Although the rate of problem data in the file appears to be relatively low (considering the large number of sources included) it might be possible to reduce the error rate further if a more aggressive editing policy for source data were introduced. This change would constitute a revision to the current philosophy of the ARE (adding data "as-is"), though the procedures required to implement it would not be very difficult.

BHPr already makes corrections to the ARF when errors are determined to be present, a process which we assume will be ongoing. However, we recommend that:

*When BHPr receives a query about data errors, a marked effort should be made to respond directly to the query and to let other users know (via the newsletter recommended in Section C) whether or not a change is needed.*

This will have the benefit of reducing rumors about problems in the ARF (which tend to have long lives), at relatively low cost.

Our further recommendations are first that:

*BHPr review its current "no editing policy (the policy is not, in any case, applied absolutely consistently, since data at the sub-county level have to be aggregated and implicit decisions get made about how to enter the data in the county file).*

In the context of this review, we also recommend that:

*BHPr should set out explicit criteria for whether or not data should be included in the file.*

BHPr implicitly makes such decisions both when it decides against adding a data source and when it decides to delete a data source.

There are three other editing changes we would recommend BHPr to explore, again regardless of what the file should be or who it is for, all of which would add value to the file. First is the issue of coarse categorization in the ARF that arises when data are aggregated from the sub-county level to the county level, and which can reduce the usefulness of the data provided. We recommend that:

*BHPr should conduct an ongoing review of current aggregation practices be undertaken, and that specific changes should be ma& immediately to the coding of the hospital occupancy rate and hospital bed variables.*

Cleaning up past data in the file would be quite expensive, so that it may make better use of BHPr resources to make the changes for future additions to the file only.

A second editing issue that we believe BHPr should revisit is the coding of five of the Virginia cities. We recognize that BHPr has already changed its practices to accommodate previous complaints, but the result is that other users are complaining. We recommend that:

*BHPr should consider accommodating both types of users through a trailer file appended to the main file and tables showing for which years and variables Virginia cities are combined with the county and for which they are separate.*

Finally, we strongly recommend that:

*BHPr should consider looking for duplicates in source data before aggregating sub-county level data and adding it to the ARF.*

This is normal practice for researchers merging data to an existing data base.

There appears to be less feedback about data needs between BHPr and other HRSA users than we would have expected. Even if this impression is erroneous, it was an issue to some HRSA users. Therefore:

*We recommend more discussions between the ARF staff and other HRSA staff about which additional data might be included on the file. In particular, we recommend that at the start of every new cycle of ARF production, ODAM and other HRSA staff meet to discuss potential additions and changes for the next public version of the ARF.*

Such a change may also increase the use of ARF among HRSA users outside of ODAM.

Another change we would recommend, even with no change in users is that:

*BHPr should look at one area of data content: the addition of new geographic codes to the ARF. If it is administratively feasible to add the SSA state and county codes, we believe it would be useful to HCFA's contractors.*

2. Utilization

The most fruitful and least expensive way of increasing both uses and users among the current target user population is through an informal newsletter, produced twice a year, to precede the distribution of a new ARF file. The HRSA users we talked to claimed to be unaware when new versions of the ARF were available, and they were also unaware of changes (including corrections) to older versions of the ARF and some were unfamiliar with the ACCESS system to the ARF. A four-page newsletter covering access to and uses of the file, as well as

adjustments, additions, and deletions to the file, would keep the ARF and changes to it in mind. We recommend that:

*BHPr should produce and distribe a newsletter to all the HRSA bureaus and known ARF users.*

Regardless of the ARF target population BHPr selects, MPR believes an electronic codebook or index would be a useful addition to the ARF system. The large number of variables and the physical size of the printed documentation increases the difficulty in using the ARF. An electronic codebook or index would simplify use of the documentation, particularly for those who do not use the ARF on a frequent basis. Therefore:

*We recommend that BHPr and the ARF contractor develop an electronic codebook as a supplement to the printed ARF documentation.*

3. Documentation

The current documentation is a basic, no-frills description of the ARF contents in two sections: a file layout with brief descriptions and a description of the data sources. Three quarters of the interview respondents said they understood the documentation and two thirds said that data problems were explained fully enough for them. Nevertheless, a series of suggestions was made by users for improving the documentation. These suggestions were augmented from MPR's detailed review of the documentation.

If the ARF should be more or less what it is now, we recommend that:

BHPr should make *improvements to the existing documentation. Of the seven possible improvements that we reviewed, we rank the order of importance as follows:*

*1. Correct existing errors (an ongoing activity that should be maintained)*

*2. Include a dump of several records in the documentation*

*3. Provide a more &tailed contents to the technical documentation*

*4. Prepare tables of time-series data available in ARF*

*5. Provide* **complete** *descriptions of* **source** *data with references*

*6. Provide a list and explanation of the suppression fields and flags*

*7. Provide the names of contact persons for source data whenever possible*

Without changing the nature of the ARF radically, there are intermediate changes that would make it more user-friendly. One such change, which we would recommend in a modified approach to improving the ARF is that:

*BHPr should include summary statistics as part of the ARF documentation*

However, we do not regard this as essential if the ARF is to stay more or less as it is.

## C. ADDITIONAL RECOMMENDATIONS: EXPANDED USER BASE WITHIN THE FEDERAL GOVERNMENT

If the decision is made to expand the user base outside of BHPr, particularly if the target population is not technically trained in programming and research, we believe that a fully-integrated data-base would be advisable for which two additional areas of editing policy would then need to be revised:

*Missing values should receive a standardized and unambiguous code.*

Furthermore, we recommend that:

*Standardization of units across comparable variables, such as population, should be* introduced.

With a wider base of users, additional data might also be required. In this case, we would recommend that:

*BHPr should in&de other types of geogmphic codes like the New England Consolidated Metropolitan Area code in the ARF.*

*If* BHPr would like to increase uses among current users elsewhere in the federal government and among those outside:

*BHPr may also wish to send the newsletter to other health agencies as well as other regular users in the federal government (such as CBO and GAO) and possibly to regular users outside the federal government.*

If the ARF is to become a fully integrated database, a codebook should be added to the documentation. The decision to move to such a database and full codebook should be tied to the response to the question of who the users are. The current users can apparently manage with the existing documentation. If the user base is expanded, the amount of technical assistance requested would presumably also increase. An increase in technical assistance could be offset by making a codebook available to users. Thus our conditional recommendation is that:

*If BHPr decides to move to a fully integrated database, with the goal of increasing the number and type of users, a full codebook should be developed*

## D. ADDITIONAL RECOMMENDATIONS: EXPANDED USER BASE OUTSIDE THE FEDERAL GOVERNMENT

To increase the user base outside the Public Health Service would require more expensive approaches such as using a marketing organization to develop brochures and disseminate them to new potential markets. We believe that BHPr must make a very careful assessment of the costs and advantages of expanding the user base outside of the Public Health Service. An increased user base will generate increased numbers of technical assistance questions from a larger base of nontechnical users. As we make clear elsewhere in this chapter, wide expansion of the ARF would require changes in documentation and access and delivery methods. These changes would be costly. If there were a way of collecting a fee for distribution of the file, there would be some financial advantage to BHPr of increasing use among outsiders that might offset the

increased costs. We are not aware of any methods that BHPr could now employ to collect and benefit from such fees.' Thus:

> *We do not recommend that BHPr undertake a major expansion of the ARF user base outside the federal government with the intent of generating a significant revenue stream. It is not clear to us that there is a large market for a research file like ARF. A substantial expansion of users would likely increase costs without commensurate financial benefits. However, with some modifications and outreach activities, the ARF would provide a useful service to the research community and generate positive publicity for BHPr.*

## E. RECOMMENDATIONS CONCERNING DELIVERY/ACCESS SYSTEMS

The delivery and access methods selected for the basic ARF are tied closely to the fundamental question: Who is the ARE for?

### 1. A Hvbrid Delivery/Access System

There are basic differences in the needs of researchers wishing to use the basic ARF file. Some analyses require processing large amounts of data while others require the use of a relatively small number of data elements. There is also a wide range in technical programing skills of researchers.

In developing the ARF system, ODAM has attempted to address the conflicting needs of researchers by implementing a hybrid delivery/access system. For those researchers with the need to process relatively large amounts of data, a magnetic tape containing the basic ARF is available. Use of the tape version of the basic ARF requires some technical programming skills.

For those researchers within the federal government wishing to access only a few variables, ODAM developed the ACCESS System The ACCESS System is an on-line system designed to

---

'Fees collected in this way would have to be returned to the U.S. Treasury.

simplify the creation of data extracts and the use of the extracts with statistical, tabulation, and graphics software. The system is intended for both technical and non-technical users.

The ACCESS System has not received strong user acceptance. The ACCESS System apparently is considered intimidating by novice computer users. Users must have NIH computer center accounts, therefore limiting users to federal agencies and their contractors. Many users are also interested in using their microcomputers for analyzing and displaying ARF data. Obtaining data from the ACCESS System for use on microcomputers is too cumbersome for many.

The two groups of users are both important. In order to satisfy both groups, BHPr should continue the hybrid approach to delivery of ARF data:

> *We recommend that BHPr continue to distribute the basic ARF in character format on magnetic tape.*

Furthermore:

> *We recommend that BHPr discontinue support for the ACCESS System and that BHPr replace it with a microcomputer-based Bulletin Board System.*

This approach will enable technically-proficient groups to use the full ARF file while simultaneously providing users interested in relatively few data items with a convenient mechanism for downloading ARF data extracts to their microcomputer.

*2.* Other Recommendations

Other recommendations involving access/delivery methods are:

> *We recommend that BHPr &lay using optical disk technologies for distributing ARF data. We suggest that BHPr address the use of optical disks technologies at the time of the re-competition of the next ARF support contract.*

If BHPr is interested in providing users with direct access to specific records or data elements (as compared to the data extracts available on the Bulletin Board), there are a number of options available:

> *We conditionally recommend that the ACCESS System be discontinued and replaced with a DBMS-based system The recommended DBMS is ORACLE.*

MPR feels that maintaining the current ACCESS system, with its low usage rates, is not justified because of its costs. Higher usage rates will require enhancements to the user interface. Enhancing the current ACCESS System offers lower immediate expenditures but higher maintenance costs and restricted flexibility. Despite the higher storage costs, the use of a DBMS offers lower maintenance costs and more flexibility. The ORACLE DBMS offers an easy-to-use and flexible interface, is available at the Parklawn Computer Center, and also operates on a microcomputer. Operation on a microcomputer offers the prospect of lower development costs.

## F. RECOMMENDATIONS CONCERNING SYSTEM EFFICIENCY

The ARF system is generally efficient and significant reductions in processing costs are not likely. There are, however, several changes or issues which could be addressed in order to reduce processing costs and facilitate future maintenance and development:

> *We recommend that the ARF contractor expand the number and detail of flowcharts contained in the ARF system documentation The ARF contractor should use microcomputer flowcharting software in preparing the flowcharts. The expanded set of system flowcharts will be valuable for future maintenance and development activities.*

> *The ARF contractor should provide BHPr with a report justifying the continued use of SAS rather than more efficient procedural languages such as COBOL. MPR believes that selected processes within the ARF system could be made more efficient through the use ofprocedural languages, especially when using microcomputers.*

> *The ARF contractor should customize the EXTRACT and UPDATE procedures for selected source files in order to allow multiple variables per record This will result in substantial reductions in the size of selected transaction files.*

*The ARF contractor should justify the use of SAS screens and menus rather than data entry software such as QPL for data entry. MPR feels that development costs would be lower and that more flexible data entry routines would be produced using QPL.*

*The ARF contractor should continue the transfer of processing from the mainframe to a microcomputer environment.*

*The ARF contractor should provide BHPr with a system design providing details of an ARF system based on a DBMS such as ORACLE. The design report should describe the file structures, storage requirements, processing efficiency, and other costs of using the DBMS as a tool in the development of the ARF files. Information in the report will enable BHPr and the ARF contractor to effectively analyze the implications of using new software technologies in the ARF system*

**MATHEMATICA**
Policy Research, Inc.

P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535

600 Maryland Avenue SW, Suite 550
Washington, DC 20024-2512
(202) 484-9220

HRSA 89-179
3606

MATHEMATICA
Policy Research, Inc.

HRSA 89-179
3606

Contract No.: HRSA 240-89-0011
MPR Reference No.: 7866

# EVALUATION OF THE AREA RESOURCE FILE (ARF) DATA SYSTEM

[redacted] FINAL REPORT:
VOLUME II-
APPENDIX F

March 9, 1990


Authors:
Judith Wooldridge
David Edson
Joshua Rozen
Lisa Schopler


Prepared for:

Department of Health and Human Services
Health Resources and Services Administration
5600 Fishers Lane, Rm. SC-02
Rockville, Maryland 20857

Project Officer:

Stuart Bernstein


Prepared by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ. 085432393

Principal Investigators:

Judith Wooldridge
David Edson

HRSA 89-179

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed **Modification:** Al
**Name:** Product Information Dissemination

---

SUMMARY OF PROPOSED MODIFICATION: More widespread information dissemination of the ARF through:

- marketing of the ARF, possibly through the use of a sales organization
- through distribution of brochures, for example, at professional meetings
- a brief newsletter produced especially for DHHS and other federal government users each time the public version of the ARF is updated.

The newsletter would describe:

- new data added to the file (adapted from the documentation)
- data deleted from the file (adapted from the documentation)
- any changes to the ACCESS system
- information on recent types of files incorporated into the ARF
- list of recent projects and the project directors & programmers who have used the file

---

1. Goal: Which problem area does the modification address?

x 1 Underutilization
2 Content & Structure of Data
3 Content & Structure of Documentation
4 Difficulty Accessing Data
5 Cost of ARF Maintenance or Operation

The publicity would increase knowledge about the existence of the ARF and its potential uses. For example, one of the reasons HRSA users gave for not using the We more frequently was that they never knew when the file had been updated. Also they did not know much about how the file was used. The newsletter would provide this basic information routinely.

2. Need: Is the problem that the modification addresses of critical importance?

X 1 Critical
2 Important
3 Useful

The criticality of implementing part or all of this modification depends on how much BHPr wants the ARF to be viewed positively and used. Given our directive, we believe that it is critical rather than important. The different elements of the

proposed modification will affect different constituencies. A newsletter will likely be most useful within HRSA and other DHHS agencies but could include other federal agencies or those outside the federal government. Increased use outside the federal government is most likely to accrue from the marketing activities.

3. What is the impact on the current ARF system programs and procedures?

No impact on the current programs or procedures.

4. What is the impact on end users?

Increased publicity would increase use of the file among both first time and previous users. For example, a newsletter would facilitate use of the file among recipients. They would be better informed about the data in the current file, perhaps more likely to use it, and more likely to discuss it with others who they are now aware of using the file.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - see OTHER
Write & Test Programs
Documentation of System
User Documentation
Other - Costs could include development of a variety of materials other than those now available and establishment of an arrangement with a sales organization. BHPr or ARF contractor staff would develop layouts and prototypes (design) and an approach to developing internal and external distribution lists would have. to be prepared.

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost (This is optional, could be low or high, but something useful can be done for medium costs.)
3 Low Cost

6. What are the implementation costs?

Costs of preparing and printing new sales materials
Costs of developing mailing list for the newsletter
Cost of preparing first newsletter
Fee to sales organization for any activity over and above design and printing of materials

Proposed Modification: Al

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

Staff costs for assembling newsletter.
Revision and reprinting of materials.
Staff costs at professional meetings.
Fee to sales organization for distribution of materials (for example at professional meetings).

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost (Could be low or high, but something useful likely can be done for a medium cost.)
3 Low Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: See Summary Evaluation Forms for Modification Numbers A2. A3. and E4 - E11
2 No:

This modification is likely to result in more first time users relative to the other modifications cited. However, a user group and increases in the amount of training provided will also increase use of ARF. The newsletter and the User Group may be somewhat interchangeable, but because we believe a user group may draw in a small number of regular users in HRSA and may not encourage new users we believe the

newsletter may be more effective. Also, a user group has to decide on a time and place, which may discourage attendance relative to reading the newsletter.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

x 1 Yes:
2 No:

ANY OTHER COMMENTS:

Proposed Modification: Al

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: A2
Name: User Group

SUMMARY OF PROPOSED MODIFICATION: An ARF user group which met from time to time to discuss the uses of the file and any problems encountered and resolved would provide an avenue for information sharing about the ARF, which might lead to file improvements, and is likely to improve the general level of information about the file (and reduce misinformation). A user group could be internal to HRSA or could be under the auspices of a larger group such as the Association of Public Data Users.

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| X | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

The file is used less than it might be because of lack of information, even among those who have sometimes used it.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| | 2 | Important |
| X | 3 | Useful |

The criticality of implementing this change to the ARF depends on how much BHPr wants the ARF to be viewed positively and used. It is more on the critical side of important than on the useful side.

3. What is the impact on the current ARF system programs and procedures?

No immediate, direct effects. Potential for long term feedback on improvements.

4. What is the impact on end users?

Likely to increase the amount of use and increase the appreciation of the file.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - see OTHER
Write & Test Programs -
Documentation of System -
User Documentation -
Other - A plan would identify the group of users who might meet, determine their interests, and commit to a minimum number of meetings to find out if this is a useful tool for increasing ARF use.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

6. What are the implementation costs?

ODAM or ARF contractor staff time to prepare for and be present at the initial meeting of the user group.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

ODAM and ARF contractor staff time to prepare for and be present at regular meetings of the user group.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

Proposed Modification: A2

10. Are Additional Hardware or Software Required for the User Community?

| | | |
|---|---|---|
| | 1 | Yes: |
| x | 2 | No: |

11. Are Alternative Approaches Available to Deal with the Problem?

| | | | |
|---|---|---|---|
| X | 1 | Yes: | See Evaluation Summary Forms for Modification Numbers Al and E4 - E11 |
| | 2 | No: | |

This change may be most attractive to regular users, and prior training (E4 -E11)) may be needed to draw in the less technical potential user.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

| | | | |
|---|---|---|---|
| X | 1 | Yes: | (Partially) - If the only people to attend the user group meetings are those who are already committed users of the ARF, then the problem of underutilization may not be met. |
| | 2 | No: | |

ANY OTHER COMMENTS:

Proposed Modification: A2

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: A3
Name: Electronic Codebook

---

SUMMARY OF PROPOSED MODIFICATION: This modification is the natural extension of most of the other longer range documentation improvements, especially the codebook with summary statistics of variables. The electronic codebook would be a PC-readable format version of this codebook bundled with a keyword/subject search software.

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| x | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Adding all the items described in the earlier modifications could make the codebook too cumbersome to use without the assistance of some keyword/subject search software.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

The difficulty in locating a variable by name and/or by subject matter greatly increases the time and frustration when dealing with the ARF.

3. What is the impact on the current ARF system programs and procedures?

This modification will affect the ARF contractor's creation process to a moderate degree. It will add a few extra steps for the contractor during the creation of the documentation.

4. What is the impact on end users?

This modification should simplify the use of the ARF, and therefore, increase the uses by the user community. At present, a user must sift through two different hundred page documents to find what information ARF has and what that

information includes. As mentioned in the codebook recommendation, many important items, such as cross-references and the mean, the maximum, and the minimum values for variables are missing. If these were added in this codebook format, it would increase the amount of information available to the user in determining variables of interest. The electronic codebook would allow the user to have access to all this information through user-friendly retrieval software. Upon entering the program, the user would be prompted through a series of menus and questions which would reveal exactly what information is available. The user would be able to compare similar variables, and view means, maximum, and minimums for any variables. The system could even include a program generator which would create the code (SAS, COBOL, or parameter field) to extract the data the user would like to analyze.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - determine which language to write the customized software in (FoxBase, dBase, C, etc.), what keys the user would be able to search on (variable name, subject matter, etc.), and how to organize the documentation database

Write & Test Programs - a series of programs must accomplish the following tasks: creation of the documentation database, search and retrieve selected variables and observations, a user interface, and possibly a code generator to extract selected variables from the ARF

Documentation of System - document the above programs

User Documentation - this documentation must be simple and clear enough to allow the most novice PC user to view information about the ARF

Other: In general, development costs for end-user software are always very high since each step must provide complete security to prevent the user from destroying the database or abnormally terminating the program.

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

Adding the data items for keyword/subject matter lookups would probably require some time to determine items associated with each ARF variable and entering these items into machine readable format. The creation of the documentation database should only require the execution of a couple of programs after the data has been entered. The retrieval software, which would be a PC-based, would run at no additional cost.

Proposed Modification: A3

ESTIMATED LEVEL OF COST:

1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

The operational costs should only add a couple of steps to the processing of documentation.

ESTIMATED LEVEL OF COST:

1 High Cost
2 Medium Cost
3 Low Cost
x 4 Minimal

8. Is Additional User Training Required?

X 1 Yes: The users would need to be trained to use this new software. If, however, the software is developed well and is truly user-friendly, the amount of training should be minimal. The only problem will involve informing the users of the new facility.

2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

x 1 Yes: The contractor would need to purchase the chosen programming language (making certain the purchasing agreement allows distribution of executable code).

2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

Xl Yes: See Summary Evaluation Forms for Modification Numbers D1- D3
2 No:

As mentioned before, this modification is the natural extension of the codebook modification. It cannot be implemented unless the codebook is also implemented.

Proposed Modification: A3

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 Yes: MPR has recently made a similar modification for another database. The functioning, user-friendly electronic codebook would greatly improve the usefulness of the ARF documentation.

2 No:

ANY OTHER COMMENTS:

Proposed Modification: A3

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: B1
Name: Correct specific data errors

SUMMARY OF PROPOSED MODIFICATION: This modification calls for the correction of specific data errors which appear on the current version of the ARF. All but one of the errors which users had pointed out to MPR staff during the interviews have already been corrected (prevailing charge identification codes which had been merged incorrectly onto the file, specific prevailing charges indices which have been deleted from the file and replaced by three overall prevailing charge indices, and HMO enrollment data which has been deleted from the file). The current error is with the nursing home facility data.

Some aspects of the data which could be considered erroneous would be: coarse coding of data, duplicates in source data, differences in the collection of population data (P-26 versus LAPI), number of HMOs based on the HMOs headquarters location (these problems are dealt with in modification B3).

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| x | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

This modification calls for the correction of specific data problems. The correction of data problems would increase the trust users have in using the ARF data.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| X | 1 | Critical |
| | 2 | Important |
| | 3 | Useful |

It is critical that the ARF data be accurate.

3. What is the impact on the current ARF system programs and procedures?

The impact of this modification on the current ARF system programs and procedures varies depending on what the data problems are. For example, if incorrectly merged data is the problem, then programs should be checked to be certain the data are being merged correctly; if how the data are measured is the

problem (number of HMOs), perhaps alternative data sources should be explored or the decision to keep the data on the ARF has to be evaluated (if users are misusing the data, then perhaps it should be deleted). It appears that the ARF contractor and ODAM have been responsive to correcting mistakes promptly. One problem which remains is that users are remembering errors they have encountered in the past even though the errors have since been corrected.

4. What is the impact on end users?

Users would have more trust in using ARF data.

5. What are the development costs (list activities for costs should be estimated)?

Plan/Design -
Write & Test Programs - Programs should be checked more carefully for mistakes (NOTE: We have costed out the one error we are aware of under B3, implement better procedures to avoid duplicates.)
Documentation of System -
User Documentation -
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low Cost

6. What are the implementation costs?

There are no implementation costs associated with this modification.

ESTIMATED LEVEL OF COST::
1 High Cost
2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

There are no operational costs associated with this modification.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
3 Low Cost

Proposed Modification: B1

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: See Summary Evaluation Forms for Modification Number B3.
2 No:

Improved data editing in the front-end will avoid some errors. For example, implement better procedures to avoid duplicates (B3). One issue which should be examined for this modification is whether or not the ARF contractor examines the source data for errors (duplicates in the source file) before transferring it onto the ARF file. Improved information about error corrections, for example in the newsletter (Al) will improve user knowledge concerning accuracy.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

Xl Yes:
2 No:

ANY OTHER COMMENTS: It is critical that data be correct and that users see it to be correct. Past errors haunt the ARF.

Proposed Modification: B1

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: B2
Name: Expanded Data

---

SUMMARY OF PROPOSED MODIFICATION: Add a variety of data to ARF:

a. Add new data sources
b. Try to obtain complete time series regularly (rather than putting efforts into acquiring annual data for a few types of data, it might make sense to make a trade-off of resources and acquire more data for fewer years)
c. Add new geographic codes that users have requested: current suggestions are Social Security Administration (SSA) state and county codes and New England Consolidated Metropolitan Area (NECMA) codes
d. Expand or revise the existing population, mortality and morbidity data

---

1. Goal: Which problem area does the modification address?

1 Underutilization
x 2 Content & Structure of Data
3 Content & Structure of Documentation
4 Difficulty Accessing Data
5 Cost of ARF Maintenance or Operation

a. Make new data available on ARF by a continual search for new sources (none are currently proposed).
b. If the goal is to provide time-series data, it would be useful to have complete data on as many sources as possible at selected intervals (for the same year) rather than to have some data for some years and not for others.
c. Decrease the amount of time it takes users to link the ARF to other major data sources.
d. Several variables would be added to ARF for the convenience of frequent users.

2. Need: Is the problem that the modification addresses of critical importance?

1 Critical
x 2 Important
3 Useful

a. It is important that ODAM keep up a continual search for new sources of county data that can be added to ARF

Proposed Modification: B2

b. Sophisticated users can handle analyses using data for different years if they have to, making adjustments for changes over time in one source to match the other source. The less sophisticated user finds use of data from different years harder to cope with.

c. Users have already developed "crosswalk" or link files that allow them to link the ARF to other data sets using SSA and NECMA codes. The modification would add the crosswalk (link) file data to the ARF and save users an extra step in data processing.

d. Other HRSA bureaus have slightly different needs than BHPr. Closer coordination on which population, natality, mortality, or morbidity variables are required could make the file more useful to those outside ODAM.

3. What is the impact on the current ARF system programs and procedures?

a. New programs would be needed for cleaning and merging

b. If there were a regular schedule of acquisition (possibly every year for the most used data and every five years for lesser used data), the file could be updated less often, saving costs. The trade-off would be that data might be available, but not on the ARF during the period between regular "all source" updates.

c. New programs would be needed for cleaning and merging

d. New programs would be needed for cleaning and merging

4. What is the impact on end users?

a. File would appeal to more users

b. Less sophisticated users might find the file easier to use for time series, while more sophisticated users might find the delays in updating ARF unacceptable.

c. Will reduce time for some users (especially those using data from HCFA).

d. Will improve the usefulness of the ARF to some users.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - a. Decide which sources were to be added;
b. Decide which are to be updated annually and which on a regular nonannual cycle
c. --
d. Agree on which variables are to be added

Write & Test Programs -
Documentation of System -
User Documentation -
Other -

ESTIMATED LEVEL OF COST:
1 High Cost

Proposed Modification: B2

2 Medium Cost
x 3 Low cost

6. What are the implementation costs?

Initial processing of new data when added to ARF

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

No additional operational costs (once data are added in implementation, that is all that needs to be done). If the changed approach to time-series were introduced this could result in net reduced costs, because the file would be updated less often.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost (Reduced Costs)

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: See Summary Evaluation Form for Modification Number D1
2 No:

If the users had a better guide as to which data were available for a given year, in summary form, they would be less likely to complain about differences in availability of data elements from year to year. One of the changes in the documentation that we are

**Proposed Modification: B2**

proposing is to tabulate the time series data available on the file. (See prototypes in the final report.)

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

| | | |
|---|---|---|
| x | Yes: | points a, c, and d |
| x | No: | point b: BHPr is acquiring virtually all the available data as soon as it is available. Hence the proposed change might simply result in delays in adding available data. An example of an exception would be Osteopathic data by specialty, which is available for every year but not present on the file. |

ANY OTHER COMMENTS: While it would be important to add new data to the file, the data proposed are not critical. Nevertheless, we have budgeted them.

Proposed Modification: B2

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: B3
Name: Data Editing Procedures (Front End)

---

SUMMARY OF PROPOSED MODIFICATION: Modify the ARF "front-end" data editing procedures to improve consistency in handling across sources. Six specific modifications are suggested:

(i) Whenever data from different sources are coded differently with respect to units of the same data element, the ARF system would put the data into common units.

(if) The criteria on the basis of which specific data were included in (or excluded from) the ARF would be explicitly stated (and included in the documentation)

(iii) Re-evaluation of the current missing values policy, which uses zero rather than a specified distinct missing value code, causing confusion to some users, especially in association with the suppression flag field

(iv) Re-evaluation of the policy of maintaining separate observations for five of the Virginia independent cities

(v) Some data in the file is grouped (for example, number of persons in a specific age group or percent of hospitals with occupancy less than 60 percent). Some of the groupings are too coarse, as a result of which important information is lost.

(vi) Review and systematize current procedures for identifying duplicate observations before adding to ARF.

---

1. Goal: Which problem area does the modification address?

1 Underutilization
x 2 Content & Structure of Data
3 Content & Structure of Documentation
4 Difficulty Accessing Data
5 Cost of ARF Maintenance or Operation

The goal is to make it easier for the user to use the file by developing consistent and unambiguous policies for handling data editing in six areas:

(i) By presenting equivalent data in the same units. An example is population data which is currently provided in the units given by the sources, which vary.

(ii) Information on criteria for inclusion or exclusion will help users to decide whether the data in the file meet their own criteria for acceptability for use in analysis.

(iii) uniform missing values

(iv) Uniform handling of data for the five independent Virginia cities which are currently separated under Modified FIPS codes when available from the source and included with the county within which they are located otherwise

(v) Provide more useful information by changing the ranges on grouped data.

(vi) To improve the accuracy of the data reported in the ARF

2. Need: Is the problem that the modification addresses of critical importance?

x 1 Critical
2 Important
3 Useful

The problems are:

- That some users are confused by current data editing policies (lack of explicitness, missing values, and varied editing with different sources)
- ARF users generally take on trust that the data in the file meet certain criteria of acceptability, It would be helpful to have the criteria codified (this would also provide a consistent basis on which ODAM staff could make decisions about inclusion.)
- With respect to Virginia independent cities, the current policy frequently results in users having to do special programming to handle the city observations (particularly when merging with other data sets). Thus one of the ARF's stated goals--of reducing front-end programming for users is defeated.
- Some data are grouped in ways that reduce usability
- Because of the current policy for handling duplicates some data are wrong without users being aware of it. When users become aware of problems, it reduces their trust in the ARF

3. What is the impact on the current ARF system programs and procedures?

Current front-end policy is to group data that are not provided at the level of the county (such as hospital data) and to check that totals from the source file are the same as the totals of merged data (and explain totals that cannot be matched usually because of missing county codes). The six changes described would have varying impacts on ARF system programs and procedures, but they generally imply more review of data content, and greater manipulation than is currently done.

(i) Procedures would be changed in the future to modify source data (this is not a normal ARF procedure).

(ii) No impact on ARF system programs. Procedures would be modified to check whether explicit criteria for source data were met.

(iii) The entire file would have to be converted if this change were made. The programs would be modified so that variables with missing values would be initialized to the new missing value code.

(iv) The current programs would have to be modified.

Proposed Modification: B3

(v) Code would have to be modified in the processing of facility data to the county level. Procedures would change to have more data review before grouping.

(vi) May require additional data editing in the form of duplicate checking for all sources before merging (currently this is not standard practice).

4. What is the impact on end users?

Confidence in the ARF will be increased because of a greater explicitness. Some fields will be clearer to the user and some time will be saved in programming (especially merges with other data sets) and analysis. Data will be more complete.

5. What are the development costs (list activities for which costs should be estimated)?

X Plan/Design -
X Write & Test Programs -
Documentation of System -
X User Documentation -
Other -

(i) Decide which is the predominant unit of measurement for a given type of data available from multiple sources. Check source data for violation of the standard units for given data elements. Write code for a given data source for a given year to amend the units for given data elements.

(ii) Requires staff to write down the criteria used in the past (if formally followed) or to prepare a Iist of criteria that have been followed (more or less) in the past and will be followed in the future.

(iii) Missing value change would require careful evaluation, file length would be affected if more than one byte code was selected. Programs would have to be modified, and user documentation updated

(iv) Policy on Virginia cities would have to be re-evaluated, and programs and user documentation rewritten

(v) More intense pre-grouping data review would be required and previous code for grouping data would have to be modified, and a decision made as to whether previously added data should be recoded and added again

(vi) Write duplicate checking programs

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost
3 Low Cost

Proposed Modification: B3

6. What are the implementation costs?

Run modified programs and corresponding documentation.

ESTIMATED LEVEL OF COST:

1 High Cost
X 2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

For some modifications there would be no costs in addition to implementation costs. For ongoing checking for duplicates and evaluations of missing data and duplicate data from new sources there would be future, ongoing operational costs

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost (between medium and low costs)
x 3 Low Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

1 Yes:
X 2 No:

Proposed Modification: B3

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

| | | |
|---|---|---|
| Xl | Ye-s: | with the exception of the missing values change |
| 2 | No: | The entire file would have to be restructured, and the cost may not be worth the end result of increased usability for the limited number of users who found zero missing value codes a problem. |

ANY OTHER COMMENTS:

Proposed Modification: B3

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: B4
Name: Modify ARF production schedule

SUMMARY OF PROPOSED MODIFICATION: Rather than creating an ARF file on a regular six month cycle, produce the file immediately after the acquisition of major data sets such as the physician, hospital, and population data.

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| x | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

This modification would reduce waiting times for major modifications of the ARF.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| x | 1 | Critical |
| | 2 | Important |
| | 3 | Useful |

The problem is largely a Public Relations problem concerning the appearance of timeliness (or lack thereof) in the ARF data.

3. What is the impact on the current ARF system programs and procedures?

May not make a large impact, since the publication of a new file is frequently held up for major data sets. If the file is produced three times a year instead of twice, it would require additional runs of documentation.

4. What is the impact on end users?

Better perception of the timeliness of the file. Potentially more use of the file.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - decide after which acquisitions the ARF would be published.
Write & Test Programs -
Documentation of System -

Proposed Modification: B4

User Documentation -
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

6. What are the implementation costs?

There are no additional costs to the current system for implementing this modification.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

Additional costs of running the final ARF file and associated documentation.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
x 2 No:

Proposed Modification: B4

11. **Are** Alternative Approaches Available to Deal with the Problem?

x 1 **Yes:** See Summary Evaluation Forms for Modification Numbers Al and A2

2 No

ODAM can choose between making better information available about the file and making the file more timely. It may be that better information has the better chance of making a difference to how long it takes for the file to be updated with new data.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

1 **Yes:**

x 2 No (Maybe): It appears that ODAM already makes a reasonable effort to link publication of the ARF to acquisition of major data sets. The marginal benefit of producing the file three times a year after the acquisition of new data may not be worth the cost of running the file and distributing information about it.

ANY OTHER COMMENTS: Assessed as useful in light of the difficulties in the way of implementing this change.

Proposed Modification: B4

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: Cl
**Name:** Provide Data At Other Geographic Levels

---

SUMMARY OF PROPOSED MODIFICATION: The county-level basic ARF file would be replaced by a zip code-base file. Data for all zip code areas in a county would be aggregated by the user to obtain county-level data.

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| x | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Some users are interested in performing analyses on a sub-county level.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| | 2 | Important |
| x | 3 | Useful |

The problem affects relatively few users.

3. What is the impact on the current ARF system programs and procedures?

This modification would change the basic concept of the ARF by converting it from a country-based to a zip-code-based file. Substantial changes in processing procedures would be required.

4. What is the impact on end users?

Changing to a zip-code file will simplify research for those interested in sub-county information but will complicate the use of the file for researchers interested in country-level data. Changes in zip codes will complicate multi-year evaluations. The size of the file will increase substantially.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - determine which variables are available on a zip code basis and which are available on the country level

Write & Test Programs - processing programs will need to be modified to allow for zip code level data existing programs

Documentation of System - major modifications

User Documentation - requires documentation of zip codes and changes in codes over time

Other:

ESTIMATED LEVEL OF COST:

x 1 High Cost
2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The costs of the various runs on the ARF would be the implementation costs.

ESTIMATED LEVEL OF COST:

x 1 High Cost
2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

The operational costs should be more than the cost of the current system.

ESTIMATED LEVEL OF COST:

x 1 High Cost
2 Medium Cost
3 Low Cost

Proposed Modification: Cl

8. Is Additional User Training Required?

x 1 Yes: Inform the users of the procedures required to obtain county level data

2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:

X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:

X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

**X** 1 **Yes:** Maintain Current File Structure

2 No:

The obvious alternative to this modification is to retain the current file structure.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

1 Yes:

X 2 No: The implementation of a zip-code file will be expensive and will be useful to a relatively small number of users

ANY OTHER COMMENTS:

Proposed Modification: Cl

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: C2
Name: Create Current Cross-Sectional and Historical Files

---

SUMMARY OF PROPOSED MODIFICATION: In general, the user community can be divided into two groups, users of time-series data or users of cross-sectional data. Currently, the ARF tries to serve both groups but fully satisfies neitber. This modification would provide a different file for each of the two groups.

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Currently, the cross-sectional group has to deal with a bulky file which contains multiply years of data for each variable. Similarly, the time-series group must worry about early years of a time-series being dropped to make room for more cross-sectional data.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| | 2 | Important |
| x | 3 | Useful |

The current file, with over 30,000 characters in each record, is very difficult to use for both groups.

3. What is the impact on the current ARF system programs and procedures?

This modification would add a moderate amount of processing to the ARF contractor's task The contractor would have to process two files, create two sets of documentation, and distribute two files. Since this duel processing could be implemented at one of the later stages of the creation of the ARF, it would not double the amount of work required.

4. What is the impact on end users?

This modification should greatly facilitate the use of the ARF for the user community. Once a user has determined whether he needs the time-series data or the cross-sectional data, he would be able to work with a file containing only the data he might require- The reduced size of the file (or the added variables in the time-series) should allow the user more flexibility in his research.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - determine which variable to include on each file and which programs need to be modified
Write & Test Programs - generally, this step will only entail modifying currently existing programs
Documentation of System - again, moderate modifications
User Documentation - moderate modifications
Other:

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The costs of the various runs on the ARF would be the implementation costs.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

The operational costs should be moderately more than the cost of the current system.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

Proposed Modification: C2

8. Is Additional User Training Required?

x 1 Yes: Inform the users of the availability and contents of the two files.
2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: Maintain Current File Structure
2 No:

The obvious alternative to this modification is to retain the current file structure. Another possibilitv would be to set up an even more complex system of files possibly organized by subject matter. The user could acquire just the file-or files he needs. Naturally, this alternative would require additional user training and a substantial investment for development cost by the ARF contractor.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 Yes: This modification would definitely simplify access to the ARF but some users might be confused about which file to use.
2 No:

ANY OTHER COMMENTS:

Proposed Modification: C2

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: C3
Name: Expand Availability of Profile Reports

---

SUMMARY OF PROPOSED MODIFICATION: At present, ODAM produces profile reports of all counties in the US based on the ARF data every two years. Hard copies of these reports are available upon request. In addition, the software that produces these reports allows ODAM to create different groupings of counties. However, only ODAM and the ARF contractor are able to use this software.

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| X | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

This modification would provide the user community with a much simpler way of accessing the ARF.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| | 2 | Important |
| X | 3 | Useful |

Many non-technical users are unable to access the ARF since they do not know SAS nor do they have access to a programming staff. The only way these users are able to utilize the ARF data is by obtaining printed copies of the Profile Reports.

3. What is the impact on the current ARF system programs and procedures?

To enact this modification, the ARF contractor would need to add a user interface to programs only they and the ODAM staff currently run. This involves prompting the user for all parameters and verifying that the responses are within range.

4. What is the impact on end users?

This modification would allow the user community to generate Profile Reports for selected counties. This would ensure that the users are using the most current

data and obtaining only the Profiles they desire, giving them the flexibility to create different groupings of counties.

5. What are the development costs (list activities for which costs should be estimated)?

| | |
|---|---|
| Plan/Design | - determine what capabilities the users would have and how to prompt them for parameters |
| Write & Test Programs | - adding the user interface and ensuring that the software is completely secure (user can not destroy the data or programs) |
| Documentation of System | - just documenting changes |
| User Documentation | - writing clear and comprehensive user documentation is difficult and time consuming |
| Other: | - Creating user interfaces is generally the most time consuming and expensive part of development. |

ESTIMATED LEVEL OF COST:

1 High Cost
2 Medium Cost
x 3 Low Cost

6. What are the implementation costs?

Implementation cost should be limited to the cost of executing the program, only marginally more than the cost of executing the current program.

ESTIMATED LEVEL OF COST:

1 High Cost
2 Medium Cost
3 Low Cost
x 4 Minimal

7. What are the Operational Costs?

The operational costs could increase substantially from the current costs. If many users have access to this program and they use it on a regular basis, a lot of runs that are not currently being executed could tie up the machine.

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost
3 Low Cost

Proposed Modification: C3

8. Is Additional User Training Required?

x 1 Yes: Some addition user-training would be required; however, this user interface should be rather simple and easy to learn.

2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:

X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

X 1 Yes: The users would need access to the mainframe computer through a terminal. Most, however, already have this access.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

x 1 Yes: See Summary Evaluation Form for Modification Number

2 No:

There are two alternatives to this proposal. First, retain the current system, providing users with hard copies of the Profile Reports. The second is to provide users with other user-friendly, data-delivery methods which simplify access to the ARF to such an extent that the users could create their own customized Profile Reports easily.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

x 1 Yes: This modification is a realistic solution to simplify user access to the data by basically just modifying a currently working program.

2 No:

ANY OTHER COMMENTS:

Proposed Modification: C3

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

**Proposed Modification:** D1
**Name:** Improve Existing Documentation

SUMMARY **OF PROPOSED MODIFICATION:** The existing documentation could be improved through several steps (independent of steps toward producing a complete code-book). These changes include:

(i) Including a dump of one record with the documentation
(ii) Correcting existing errors in the documentation
(iii) Providing complete descriptions of source data, with references
(iv) Providing a contact name for source data
(v) Using and documenting a consistent set of abbreviations
(vi) Providing an explanation of suppression fields and flags
(vii) Providing tables of time-series available in the file

1. Goal: Which problem area does this modification address?

1 Underutilization
2 Content & Structure of Data
X 3 Content & Structure of Documentation
4 Difficulty Accessing Data
5 Cost of ARF Maintenance or Operation

To improve the usability and understanding of the documentation and hence the file.

2. Need: Is the problem that the modification addresses of critical importance?

X 1 Critical
2 Important
3 Useful

Users requested the seven changes listed because they had encountered difficulties using the current documentation. The importance of the changes is as follows:

(i) Dump: Some users can dump data themselves, for others, this would be a useful adjunct to the documentation because it would reveal how the data appear in particular fields, where this is not clear in the documentation
(ii) Correct errors: critical for continued confidence in the file
(iii) Users sometimes need to know the precise source of data in order to check that it has the meaning they are ascribing to it. The current source information is inadequate

Proposed Modification: D1

(iv) A contact person at the source (where acceptable to the source who provides the data) would be of assistance o users. There is a precedent for providing this sort of information in public use files.

(v) The current inconsistency in the use of abbreviations and lack of definitions confuses users

(vi) The user documentation does not explain how the suppression fields and flags are coded

(vii) Users complained that they could not always find the time series data that do exist. Tabulating data with variable names would ease this problem

3. What is the impact on the current ARF system programs and procedures?

Documentation procedures would have to be modified to produce a dump. Other changes would be one-time modifications to the existing user and technical documentation

4. What is the impact on end users?

Will improve usability of the file.

5. What are the development costs (including changes to ARF system programs)?

Plan/Design - Decide which changes to make
Write & Test Programs -
Documentation of System -
User Documentation - Plan one-time changes
Other - Technical documentation--plan one time changes

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low cost

6. What are the implementation costs?

Correct the documentation at the time it is next run.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low cost

7. What are the Operational Costs?

Perform minor updates to the user and technical documentation for new sources as they are added to the ARF.

Proposed Modification: D1

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes (partly): See Summary Evaluation Form for Modification Number D2
2 No:

A code-book will resolve some of the problems that these modifications are intended to resolve (e.g. suppression flag explanations, lack of understanding of how data are coded).

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 Yes: However, future errors are bound to occur.
2 No:

ANY OTHER COMMENTS:

Proposed Modification: D1

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: D2
Name: Develop Code-book

---

SUMMARY OF PROPOSED MODIFICATION: The documentation would be revised through a series of five actions that would culminate in a codebook. The actions are:

(i) Develop Crosswalk: an indicator would be added to the technical documentation that would show which variables have further discussion in the User documentation. Variable names from the technical documentation would be added to the User documentation.

(ii) Incorporate complete descriptions of variables in the documentation including coding, source, nonresponse, editing, and merge problems

(iii) Provide page references in the technical documentation to information in the user documentation

(iv) Move variable descriptions from the technical documentation (leaving a simple file layout) to the user documentation (coming closer to a codebook)

(v) Coordinate variable order in the user and technical documentation

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Da?;: |
| x | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Develop a codebook that would make it easier to use the user documentation than it is at present.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Users currently have a hard time using the user documentation because not all variables are described in the user documentation, variable names are not given and the variable order is different from that in the technical documentation. Hence, users appear not to refer to the user documentation as much as would be desirable; the closer the documentation comes to a file layout and a codebook, the easier it will be to use, and use will likely increase.

Proposed Modification: D2

3, What is the impact on the current ARF system programs and procedures?

The technical and user documentation programs would have to be substantially revised to include the additional recommended information in the proposed formats.

4. What is the impact on end users?

Users would be able to find the information they needed with less time than it takes at present. For those users who do not take the time with the present documentation, they may make more use of the documentation. All users will be better informed about the file, might make more use of it, and the uses might be more appropriate.

5. What are the development costs (list activities for which costs should be estimated)?

X Plan/Design - develop a cross-walk between the technical and user documentation
X Write & Test Programs - rewrite the programs which produce the technical and user documentation
Documentation of System -
X User Documentation - restructuring of the technical and user documentation
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
X 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

First time programs are run

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low Cost

7. What are the Operational Costs?

Cost of rerunning the documentation each time the file is run in the future. No change from current costs.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low Cost

Proposed Modification: D2

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X2 No

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X2 No

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: See Summary Evaluation Form for Modification Number D1
2 No:

D1 is a short term solution to ease the awkwardness of the current user documentation. In the longer term, a fully integrated electronic codebook would substitute for this modification but requires the work described under D2.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 Yes:
2 No:

ANY OTHER COMMENTS:

Proposed Modification: D2

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: D3
**Name:** Add Summary Statistics to Codebook

SUMMARY OF PROPOSED MODIFICATION: A codebook for the ARF would show name, codes, frequency distributions, minimum, mean and maximum values, and source for each variable.

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| x | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

To facilitate use of the ARF, and to reduce the amount of "looking at" the data required in order to understand variable content.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

The usefulness of the ARF is marred by the difficulty the user may have in understanding the content of variables on the file: summary statistics make clear data content, and provide a cross-check for users to ensure that they are reading the file properly.

3. What is the impact on the current ARF system programs and procedures?

Would require redesign of the documentation preparation procedures and running of summary statistics

4. What is the impact on end users?

Users will need to invest less time to understand and use the file.

5. What are the development costs (list activities for which costs should be estimated)?

X Plan/Design - Determine which variables are to be described by frequency distributions and which by mean, maximum, and minimum values

X Write & Test Programs - write a basic program to produce the descriptive statistics

Documentation of System -
User Documentation -
Other -

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

First time run of summary statistics.

ESTIMATED LEVEL OF COST:

1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

Ongoing running of summary statistics for each new file.

ESTIMATED LEVEL OF COST:

1 High Cost
2 Medium Cost
x 3 Low Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X2 No

Proposed Modification: **D3**

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

1 Yes:
X 2 No:

However, the availability of a codebook (or an electronic codebook) will greatly facilitate use of the ARF. (Modifications D2 and A3)

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 Yes:
2 No:

ANY OTHER COMMENTS:

Proposed Modification: D3

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: El
Name: Distribute on Magnetic Tape (Alternative Formats)

SUMMARY OF PROPOSED MODIFICATION: Under this modification, the ARF would be distributed on magnetic tape(s) in both character and SAS format. Many users have difficulty accessing the ARF data in its existing character format. These users are not comfortable using procedural languages and lack access to trained computer programming staff. Others convert the data to SAS format for analysis. Providing the ARF in SAS format would simplify access and encourage use of the ARF.

1. Coal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

This modification will simplify the use of the ARF.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

This modification would simply and encourage use of the ARF.

3. What is the impact on the current ARF system programs and procedures?

The ARF contractor would have to convert the character version of the ARF to SAS format.

4. What is the impact on end users?

Access for end users desiring to use SAS would be simplified.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - develop an automated process for creating a SAS format version of the ARF

Write & Test Programs - develop conversion software
Documentation of System - this includes documenting the conversion program
User Documentation - the documentation would be expanded to include the output from a SAS PROC CONTENTS on the SAS version of the ARF

Other -

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

6. What are the implementation costs?

An additional conversion step would be added to the end of the ARF processing

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

Additional operational costs would involve creating and distributing the SAS version

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

8. Is Additional User Training Required?

X 1 Yes: The users wishing to use SAS would require SAS training.
2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

10. Are Additional Hardware or Software Required for the User Community?

X 1 Yes: Users would be expected to have access to SAS software.
2 No:

Proposed Modification: El

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: See Summary Evaluation Form for Modification Numbers El through El2

2 No:

Most of the alternatives mentioned elsewhere (bulletin boards, downloading, etc.), address the problem of accessing the data.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

Xl Yes: This modification would be a low cost mechanism which would simplify access.

2 No:

ANY OTHER COMMENTS:

Proposed Modification: **El**

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: E2
Name: Distribute on Alternate Media (Optical Discs)

SUMMARY OF PROPOSED MODIFICATION: Under this modification, the ARF would be distributed on an alternate media, such as optical disk At present, there are three methods of optical storage, the CD-ROM (Compact Disk, Read Only Memory), the WORM drive (Write Once, Read Many times), and the erasable disk. The CD-ROM is the most likely candidate for the ARF due to its relatively inexpensive price (compared to the other two), availability and capabilities. The CD-ROM is usually used for distribution of numerous copies of a data base (often this takes the from of reference libraries).

Using CD-ROM, the end-user would be able to access the complete ARF on a microcomputer without having the resort to downloading data or transferring data from magnetic tape.

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

This modification will utilize the latest technology to improve access to the data on microcomputers.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

The availability of the ARF on CD-ROM will enable microcomputer users with the appropriate hardware to easily access the entire ARF.

3. What is the impact on the current ARF system programs and procedures?

The ARF contractor would have to reformat the ARF to prepare it for "uploading" to the optical disk. Then the contractor would have to arrange to make a master disk and reproduce and distribute copies as needed.

4. What is the impact on end users?

End users desiring to use microcomputers for their analyses would have easy access to the complete ARF.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - determine the best structure for the data (this is especially important since seek times on CD-ROM's are very slow while sequential reads are extremely fast)

Write & Test Programs - develop special retrieval software

Documentation of System - this includes documenting a whole new set of programs

User Documentation - this includes explaining to the users how a obtain the disk and how to use the retrieval software

Other -

ESTIMATED LEVEL OF COST:

x 1 High Cost
2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The costs of creating the master can be very expensive. First, the contractor would have to create a prototype on their machine. Then, they would have to have the CD-ROM master printed (which costs from $2,000 to $5,000). Additional copies of the CD's would be inexpensive.

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

A new master and copies would be required for each release of the ARF.

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost
3 Low Cost

Proposed Modification: E2

8. Is Additional User Training Required?

X 1 **Yes:** The users would need to be trained in the use of the retrieval software.

2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

X 1 Yes: The ARF contractor would need to purchase at least one hard disk with 700Mb of storage to create the prototype. In addition, they would need to purchase an optical disk drive and possibly some additional software.

2 No:

10. Are Additional Hardware or Software Required for the User Community?

X 1 **Yes:** Users would be required to purchase an optical disk drive.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

1 Yes:

X 2 No:

Most of the alternatives mentioned elsewhere (bulletin boards, downloading, etc.), address the problem of accessing the data.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

1 Yes:

X 2 No: This modification would be very expensive to implement and would only solve the problem of access for the few users that possess the required hardware.

ANY OTHER COMMENTS:

Proposed Modification: E2

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: E3
Name: Automated Access to Individual Data Elements/Observations

SUMMARY OF PROPOSED MODIFICATION: Many users prefer to do their analysis on microcomputers. To accommodate these users, software vendors have developed integrated micro-to-mainframe operating environments. Usually tied closely to the vendor's mainframe DBMS, these software packages perform instant retrieval of mainframe data and subsequent entry of the data into a microcomputer spreadsheet or database. Additional functions may also be incorporated into the software, including streamlined upload procedures, the use of mainframe storage as virtual disks, and 3270 full-screen emulation. An example of this automated mainframe-to-microcomputer transfer packages is PC/204 from Computer Corporation of America. Oracle Corporation has developed a link between its DBMS and LOTUS 123 for selected platforms.

Currently, to use the most recent ARF data on a PC, a user must either have a g-track tape unit attached to a PC or download the data from a mainframe facility. Using an automated mainframe-to-microcomputer transfer package would facilitate the downloading process.

The specific proposal is to investigate the use of PC/204 or its Oracle equivalent with Model 204 and Oracle at the Parklawn Computer Center.

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Basically, this modification will provide the users with an alternate method for obtaining the data (instead of having to read it from a magnetic tape).

2 Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| | 2 | Important |
| x | 3 | Useful |

Users often mention the difficulty they have accessing the ARF data, especially on the PC.

Proposed Modification: E3

3. What is the impact on the current ARF system programs and procedures?

The contractor would have to ensure that the most recent copy of the ARF data is available on-line and loaded into the mainframe DBMS. The contractor would utilize some software package which includes both a PC and mainframe component.

4. What is the impact on end users?

This modification should simplify the downloading process for users. It will enable more users to access the ARF more often in the user-friendly PC environment. The user would not have to worry about the commands for downloading the data. The PC software would execute the download after the user has made his data selection.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - The ARF contractor would have to design the Model 204 or Oracle database. Load procedures would have to be specified.
Write & Test Programs - Load programs would have to be developed and tested
Documentation of System - Load programs would have to be documented
User Documentation - the data base and load programs would have to be documented
Other -

ESTIMATED LEVEL OF COST:
X 1 High Cost
2 Medium Cost
3 Low Cost

6. What are the implementation costs?

Loading the complete ARF into Model 204 or Oracle

ESTIMATED LEVEL OF COST:
X 1 High Cost
2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

The operational costs could increase substantially from current costs if many users utilize this process often. In addition, the cost of maintaining the data on-line could be expensive.

Proposed Modification: E3

ESTIMATED LEVEL OF COST:

x 1 High Cost
2 Medium Cost
3 Low Cost

8. Is Additional User Training Required?

X 1 **Yes:** Some user training would be required to teach the users how to use the software. If, however, the selected software is truly user-friendly, the training should be minimal.

2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

X 1 Yes: The contractor might need to purchase some data transfer software on the mainframe (and the PC version for testing) and possible addition storage on the mainframe.

2 No:

10. Are Additional Hardware or Software Required for the User Community?

X 1 **Yes:** A user would need a modem to download the data, a common item to which many users already have access. The user community would be required to purchase a specific data transfer software to utilize this new feature.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 **Yes:** See Summary Evaluation Form for Modification Numbers El through El2

2 No:

This modification is one of the most expensive delivery method modifications for the ARF contractor to implement but could be one of the simplest for the user community.

12. Is Feasibility of Success Reasonably High? Will the Modification Solve the Problem?

1 **Yes**

x 2 No: This recommendation would probably be too difficult to implement but would greatly simplify the process for those users who are currently having trouble gaining access to the data on the PC. However, costs may be very high. For example, commercial prices for PC/204 are $10,000 for the mainframe component and $500 for each microcomputer. Even considering government purchase discounts, cost is likely to be a serious barrier to use of this system.

Proposed Modification: E3

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: E4
Name: Manual Access to Individual Data Elements/Observations Using Mainframe (Flat File/Non-menu Driven Software)

---

SUMMARY OF PROPOSED MODIFICATION: Under this scenario, which is similar to the current ACCESS system at NIH, individuals would use terminals or their own microcomputer communications software to access the NIH or Parklawn mainframes. They would use the TSO or WYLBUR systems to submit procedural language programs.

A relatively simple front end, like the ACCESS system, would be available to assist users in preparing and submitting their programs.

Downloading, if desired, would be the responsibility of the user.

The ARF would be organized as a flat file in SAS and/or character format. If usage warranted, the data could be stored on-line rather than on tape. This would dramatically speed job turnaround time.

Users would be required to obtain an account at the mainframe data center.

This approach would allow users to obtain access to the latest version of the ARF but would also require some degree of programming sophistication or experience.

---

1. Goal: Which problem areas does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Basically, this modification will provide the users with an alternate method for obtaining the data (instead of having to read it from a magnetic tape).

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Users have mentioned the difficulty they have accessing the most recent ARF data.

Proposed Modification: E4

3. What is the impact on the current ARF system programs and procedures?

The contractor would have to ensure that the most recent copy of the ARF data is available at the mainframe data center.

4. What is the impact on end users?

This modification would allow users to log on to the mainframe, submit a job to create an extract of the data, and download the extract.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - None, other than fine-tuning of existing ACCESS system
Write & Test Programs - write and test any ACCESS system modifications
Documentation of System - document the changes
User Documentation - explain how to access the system using the modified version of ACCESS

Other -

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

6. What are the implementation costs?

Implementation cost should be rather small, with tape storage and file copying costs being the only costs.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

The operational costs could increase substantially from current costs if on-line storage is utilized.

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

Proposed Modification: E4

8. Is Additional User Training Required?

| | | | |
|---|---|---|---|
| Xl | | Yes: | Users would need to be proficient enough to execute the extract programs and, utilizing their PC communication software, download the file they created |
| | 2 | No: | |

9. Are Additional Hardware or Software Required for the ARF Contractor?

| | | | |
|---|---|---|---|
| | 1 | Yes: | |
| X | 2 | No: | However, the contractor might need additional on-line storage on the mainframe. |

10. Are Additional Hardware or Software Required for the User Community?

| | | | |
|---|---|---|---|
| x | 1 | Yes: | A user would need a modem (and communication software) to download the data, a common item to which many users already have access. |
| | 2 | No: | |

11. Are Alternative Approaches Available to Deal with the Problem?

| | | | |
|---|---|---|---|
| X | 1 | Yes: | See Summary Evaluation Form for Modification Numbers El through E12. |
| | 2 | No | |

This modification is the next to simplest approach to alleviate the problems of accessing the data. Most of the following recommendations are more involved approaches to solving the same problem.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

| | | | |
|---|---|---|---|
| | 1 | Yes: | |
| X | 2 | No: | This recommendation would probably not be too difficult to implement but may not simplify the process much for those users who are currently having trouble gaining access to the data on the PC. In addition, this system might not be feasible for a flat file on the mainframe (on-line storage could be very expensive and the extract run could be expensive). |

ANY OTHER COMMENTS:

Proposed Modification: E4

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

**Proposed Modification: E5**
**Name:** Manual Access to Individual Data Elements/Observations Using Mainframe (Flat File/Menu-Driven Software)

---

**SUMMARY OF PROPOSED MODIFICATION: This** method is identical to the previous modification except the user would have access to more elaborate menu-driven software to assist in file extract preparation.

---

1. Goal: Which problem areas does this modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Basically, this modification will provide the users with an alternate method for obtaining the data (instead of having to read it from a magnetic tape).

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Users often mention the difficulty they have accessing the ARF data, especially on the PC.

3. What is the impact on the current ARF system programs and procedures?

The contractor would have to ensure that the most recent copy of the ARF data is available.

4. What is the impact on end users?

This modification would allow users to utilize menu-driven software to create an extract.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - the ARF contractor would have to design the menu- driven software to create the extracts
Write & Test Programs - write and test the software
Documentation of System - document the software
User Documentation - explain how to access the system using the menu-driven software
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The implementation cost would involve installing the menu-driven software.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

The operational costs could increase substantially from current costs if many users utilize this process often. In addition, the cost of maintaining the data on-line could be expensive.

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

8. Is Additional User Training Required?

Xl Yes: Users would need to be trained in the use of the menu-driven software.
2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

Proposed Modification: E5

10. Are Additional Hardware or Software Required for the User Community?

X 1 **Yes:** A user would need a modem (and communication software) to download the data, a common item to which many users already have access.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: See Summary Evaluation Form for Modification Numbers El through E12.

2 No

This modification is the next logical step to the previous modification.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 **Yes:**

2 No:

ANY OTHER COMMENTS:

Proposed Modification: E5

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: E6
Name: Manual Access to Individual Data Elements/Observations Using Mainframe DBMS/Non-menu Driven Software

SUMMARY OF PROPOSED MODIFICATION: This modification is very similar to E4 except it utilizes a DBMS to reduce on-line storage requirements and distributes the access to many files instead of just one. The ARF would be stored using a DBMS such as Model 204 or Oracle. Rather than using a procedural language to create an extract, the user would employ the DBMS-specific language.

1. Goal: Which problem areas does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Basically, this modification will provide the users with an alternate method for obtaining the data (instead of having to read it from a magnetic tape).

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Users often mention the difficulty they have accessing the ARF data, especially on the PC.

3. What is the impact on the current ARF system programs and procedures?

The contractor would have to maintain an on-line DBMS for users to access.

4. What is the impact on end users?

This modification would allow users to log on to the mainframe, submit a job to create an extract of the data, and download the extract.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - the ARF contractor would have to design the file structure for the database
Write & Test Programs - write and test the DBMS programs
Documentation of System - document the programs
User Documentation - explain how to access the system
Other -

ESTIMATED LEVEL OF COST:
X 1 High Cost
2 Medium Cost
3 Low Cost

6. What are the implementation costs?

It would be relatively expensive to perform the initial database load procedures.

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

The operational costs could increase substantially from current costs since on-line storage is required. This cost would be offset slightly by the reduced cost of updating the database files.

ESTIMATED LEVEL OF COST:
X 1 High Cost
2 Medium Cost
3 Low Cost

8. Is Additional User Training Required?

Xl Yes: Users would need to be proficient enough to execute the extract programs and, utilizing their communication software, download the file they created.
2 No:

Proposed Modification: E6

9. Are Additional Hardware or Software Required for the ARF Contractor?

| | | | |
|---|---|---|---|
| X | 1 | Yes: | The contractor might need addition on-line storage on the mainframe. |
| | 2 | No: | |

10. Are Additional Hardware or Software Required for the User Community?

| | | | |
|---|---|---|---|
| Xl | | Yes: | A user would need a modem (and communication software) to download the data, a common item to which many users already have access. |
| | 2 | No: | |

11. Are Alternative Approaches Available to Deal with the Problem?

| | | | |
|---|---|---|---|
| x | 1 | Yes: | See Summary Evaluation Form for Modification Numbers El through El2 |
| | 2 | No: | |

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

| | | | |
|---|---|---|---|
| | 1 | Yes | |
| X | 2 | No: | This recommendation would probably be difficult to implement and may not simplify the process much for those users who are currently having trouble gaining access to the data on the PC. |

ANY OTHER COMMENTS:

Proposed Modification: E6

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: E7
Name: Manual Access to Individual Data Elements/Observations Using Mainframe DBMS/Menu Driven Software

SUMMARY OF PROPOSED MODIFICATION: This method is identical to the previous modification except the user would have access to more elaborate menu-driven software to assist in file extract preparation.

1. Goal: Which problem areas does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Basically, this modification will provide the users with an alternate method for obtaining the data (instead of having to read it from a magnetic tape).

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Users often mention the difficulty they have accessing the ARF data, especially on the PC.

3. What is the impact on the current ARF system programs and procedures?

The contractor would have to ensure that the most recent copy of the ARF data is available.

4. What is the impact on end users?

This modification would allow users to utilize menu-driven software to create an extract.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - the ARF contractor would have to design the menu- driven software to create the extracts
Write & Test Programs - write and test the software
Documentation of System - document the software
User Documentation - explain how to access the system using the menu-driven software
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The implementation cost would involve installing the menu-driven software.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

The operational costs could increase substantially from current costs if many users utilize this process often. In addition, the cost of maintaining the data on-line could be expensive.

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

8. Is Additional User Training Required?

Xl Yes: Users would need to be trained in the use of the menu-driven system.
2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

Proposed Modification: E7

30. Are Additional Hardware or Software Required for the User Community?

x 1 Yes: A user would need a modem (and communication software) to download the data, a common item to which many users already have access.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

Xl Yes: See Summary Evaluation Form for Modification Numbers El through E12.

2 No:

This modification is the next logical step to the previous modification.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

x 1 Yes:

2 No:

ANY OTHER COMMENTS:

Proposed Modification: E7

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: E8
Name: Manual Access to Individual Data Elements/Observations Using Microcomputer (Fiat File/Non-Menu-Driven Software)

---

SUMMARY OF PROPOSED MODIFICATION: This modification is identical to E4 - Manual Access to Individual Data Elements/Observation Using Mainframe (Flat File/Non-Menu-Driven Software), except that a microcomputer would serve as a storage facility rather than a mainframe. A microcomputer operating under SCO Unix or Xenix with the appropriate storage and communications software would perform the same functions as the NIH or Parklawn mainframes. End-users would dial into the system and, using a relatively simple front-end, create extracts for downloading.

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Basically, this modification will provide the users with an alternate method for obtaining the data (instead of having to read it from a magnetic tape).

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Users have mentioned the difficulty they have accessing the most recent ARF data.

3. What is the impact on the current ARF system programs and procedures?

The contractor would have to ensure that the most recent copy of the ARF data is available on the microcomputer server.

4. What is the impact on end users?

This modifkation would allow users to log on to the microcomputer server, submit a job to create an extract of the data, and download the extract.

Proposed Modification: ES

5. What are the development costs (including changes to ARF system programs)?

Plan/Design - The ARF contractor would have to set up the microcomputer server, develop the software interface, and develop data loading procedures

Write & Test Programs - write and test the access software and loading procedures

Documentation of System - document the new programs

User Documentation - explain how to access the system using the new software and hardware

Other -

ESTIMATED LEVEL OF COST:

1 High Cost
X 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

Other than obtaining microcomputer server and software, costs should be low.

ESTIMATED LEVEL OF COST:

1 High Cost
2 Medium Cost
X 3 Low Cost

7. What are the Operational Costs?

Once the hardware purchase is completed, operational costs would be relatively low. A substantial increase in utilization might require additional hardware.

ESTIMATED LEVEL OF COST:

1 High Cost
X 2 Medium Cost
3 Low Cost

8. Is Additional User Training Required?

X 1 Yes: Users would need to be proficient enough to execute the extract programs and, utilizing their PC communication software, download the file they created.

2 No:

Proposed Modification: E8

9. Are Additional Hardware or Software Required for the ARF Contractor?

Xl Yes: Probably a 386 machine with a large hard disk, a multi-communication port, and software.

2 No:

10. Are Additional Hardware or Software Required for the User Community?

X 1 Yes: A user would need a modem (and communication software) to download the data, a common item to which many users already have access.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: See Summary Evaluation Form for Modification Numbers El through E12.

2 No:

This may be a cost-effective approach to providing direct on-line access to the ARF data.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

1 Yes:

X 2 No: This recommendation would probably not be too difficult to implement but may not simplify the process much for those users who are currently having trouble gaining access to the data on the PC.

ANY OTHER COMMENTS:

Proposed Modification: ES

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: E9
Name: Manual Access to Individual Data Elements/Observations Using Microcomputer (Flat File/Menu-Driven Software)

---

SUMMARY OF PROPOSED MODIFICATION: This method is identical to the previous modification except the user would have access to more elaborate menu-driven software to assist in file extract preparation.

---

1. Goal: Which problem areas does this modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Basically, this modification will provide the users with an alternate method for obtaining the data (instead of having to read it from a magnetic tape).

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Users often mention the difficulty they have accessing the ARF data, especially on the PC.

3. What is the impact on the current ARF system programs and procedures?

The contractor would have to ensure that the most recent copy of the ARF data is available.

4. What is the impact on end users?

This modification would allow users to utilize menu-driven software to create an extract.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - the ARF contractor would have to design the menu- driven software to create the extracts
Write & Test Programs - write and test the software
Documentation of System - document the software
User Documentation - explain how to access the system using the menu-driven software
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The implementation cost would involve installing the menu-driven software.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

The operational costs could increase substantially from current costs if many users utilize this process often. In addition, the cost of maintaining the data on-line could be expensive.

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

8. Is Additional User Training Required?

X 1 Yes: Users would need to be trained in the use of the menu-driven software.
2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

X 1 Yes: Probably a 386 machine with a large hard disk, a multi-communication port, and software.
2 No:

Proposed Modification: E9

10. Are Additional Hardware or Software Required for the User Community?

Xl **Yes**: A user would need a modem (and communication software) to download the data, a common item to which many users already have access.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

**X** 1 Yes: See Summary Evaluation Form for Modification Numbers El through El2

2 No

This modification is the next logical step to the previous modification.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

**X** 1 **Yes:**

2 No:

ANY OTHER COMMENTS:

Proposed Modification: E9

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

**Proposed Modification:** El0
**Name:** Manual Access to Individual Data Elements/Observation Using Microcomputer DBMS/Non-menu Driven Software

**SUMMARY OF PROPOSED MODIFICATION:** This modification is very similar to E6 except it utilizes a microcomputer hardware and software. Hardware requirements are described in ES. This modification would require the use of multi-user PC DBMS software such as PC Oracle or Paradox.

1. Coal: Which problem areas does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

Basically, this modification will provide the users with an alternate method for obtaining the data (instead of having to read it from a magnetic tape).

**2.** Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Users often mention the difficulty they have accessing the ARF data, especially on the PC.

3. What is the impact on the current ARF system programs and procedures?

The contractor would have to maintain an on-line DBMS for users to access.

4. What is the impact on end users?

This modification would allow users to log on to the microcomputer server, submit a job to create an extract of the data, and download the extract.

5. What are the development costs (list the activities for which cost should be estimated)?

Plan/Design - the ARF contractor would have to design the file structure for the database
Write & Test Programs - write and test the DBMS programs
Documentation of System - document the programs
User Documentation - explain how to access the system
Other -

ESTIMATED LEVEL OF COST:
x 1 High Cost
2 Medium Cost
3 Low Cost

6. What are the implementation costs?

It would be relatively expensive to perform the initial database load procedures.

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

The operational costs could increase substantially from current costs, since additional storage may need to be purchased. This cost would be offset slightly by the reduced cost of updating the database files.

ESTIMATED LEVEL OF COST:
x 1 High Cost
2 Medium Cost
3 Low Cost

8. Is Additional User Training Required?

Xl Yes: Users would need to be proficient enough to execute the extract programs, and utilizing their PC communication software, download the file they created.
2 No:

Proposed Modification: El0

9. Are Additional Hardware or Software Required for the ARF Contractor?

x 1 Yes: Probably a 386 machine with a large hard disk, a multi-communication port, and software.

2 No:

10. Are Additional Hardware or Software Required for the User Community?

x 1 Yes: A user would need a modem (and communication software) to download the data, a common item to which many users already have access.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

x 1 Yes: See Summary Evaluation Form for Modification Numbers El through E12.

2 No

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

1 Yes:

X 2 No: This recommendation would probably be difficult to implement and may not simplify the process much for those users who are currently having trouble gaining access to the data on the PC.

ANY OTHER COMMENTS:

Proposed Modification: E10

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: **E11**
Name: Manual Access to Individual Data Elements/Observations Using Microcomputer DBMS/Menu-Driven Software

SUMMARY OF PROPOSED MODIFICATION: This method is identical to the previous modification except the user would have access to more elaborate menudriven software to assist in file extract preparation.

1. **Goal:** Which problem areas does the modification address?

   1 Underutilization
   2 Content & Structure of Data
   3 Content & Structure of Documentation
   x 4 Difficulty Accessing Data
   5 Cost of ARF Maintenance or Operation

   Basically, this modification will provide the users with an alternate method for obtaining the data (instead of having to read it from a magnetic tape).

2. Need: Is the problem that the modification addresses of critical importance?

   1 Critical
   x 2 Important
   3 Useful

   Users often mention the difficulty they have accessing the **ARF** data, especially on the PC.

3. What is the impact on the current ARF system programs and procedures?

   The contractor would have to ensure that the most recent copy of the ARF data is available.

4. What is the impact on end users?

   This modification would allow users to utilize menu-driven software to create an extract.

Proposed Modification: **El 1**

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - the ARF contractor would have to design the menu- driven software to create the extracts
Write & Test Programs - write and test the software
Documentation of System - document the software
User Documentation - explain how to access the system using the menu-driven software
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The implementation cost would involve installing the menu-driven software.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

The operational costs could increase substantially from current costs if many users utilize this process often. In addition, the cost of maintaining the data on-line could be expensive.

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

8. Is Additional User Training Required?

X 1 Yes: User would need to be trained in the use of menu-driven software.
2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

X 1 Yes: Probably a 386 machine with a large hard disk, a multi-communication port, and software
2 No:

Proposed Modification: El 1

10. Are Additional Hardware or Software Required for the User Community?

X 1 **Yes:** A user would need a modem (and communication software) to download the data, a common item to which many users already have access.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: See Summary Evaluation Form for Modification Numbers El through E12.

2 No:

This modification is the next logical step to the previous modification.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 **Yes:**

2 No:

ANY OTHER COMMENTS:

Proposed Modification: El 1

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

**Proposed** Modification: E12
Name: Bulletin Board Providing Direct Access to Data Elements

---

SUMMARY OF PROPOSED MODIFICATION: Recently, computer bulletin boar& have become a popular method to distribute data and information electronically. The user is able to, through his computer, call another computer and view various information on that computer. The ARF data could be made available on one of these bulletin boards and the users could download selected data items.

This differs from some of the other proposed modifications in that the user does not have control over the content of the file selected for downloading. File content would be determined by the ARF contractor.

A microcomputer-based server would use Bulletin Board software to control access to files.

---

1. Goal: Which problem areas does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| x | 4 | Difficulty Accessing Data |
| | 5 | Cost of ARF Maintenance or Operation |

The goal is to provide the users with an alternate method for download the ARF data to their PC's.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

As mentioned before, many users have difficulty accessing data, particularly in the PC environment. At present, when a user would like to access the main ARF, he must first read the data tape on a mainframe, select the data he would like to use, and download this data, a rather lengthy process.

3. What is the impact on the current ARF system programs and procedures?

The ARF contractor would load the ARF data onto a server, the machine that users dial to access the data.

Proposed Modification: El2

4. What is the impact on end users?

The user community would benefit greatly from this modification. Without ordering the tape, and waiting for its delivery, a user would be able to dial-up and extract data. Thus, he could access the data much more quickly and produce more timely results.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - determine best hardware and software configuration
Write & Test Programs - write software for retrieve selected data items
Documentation of System - document this software
User Documentation - describe to the user how to access the system
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The only cost would be reformatting the data and loading it onto the Bulletin Board.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

With the exception of hardware, software, and training requirements listed below, the operational cost should be minimal.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
3 Low Cost
x 4 Minimal Cost

8. Is Additional User Training Required?

X 1 Yes: In order to make this modification work, the ARF contractor will have to invest some time and effort into training the users. The

purpose of the training would be two-fold: 1) inform users of the new method for accessing the data; 2) training the users in how to actually use this method.

2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

x 1 Yes: Most likely, the contractor would need to dedicate a machine with a high-speed modem and a large hard disk to function as the server. In addition, the contractor might be required to purchase data transfer and/or data retrieval software.

2 No:

10. Are Additional Hardware or Software Required for the User Community?

X 1 Yes: Once again, the only hardware and software requirements for the users is a PC with a modem and communication software to use that modem, which many users already possess.

2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

x 1 Yes: See Summary Evaluation Form for Numbers El through Ell.

2 No:

This is one of the mid-range proposals for simplifying access to the ARF data.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

x 1 Yes: If this recommendation works well, it would greatly improve the user communities ability to access the ARF. However, it requires substantial investments in equipment, development, and user training.

2 No:

ANY OTHER COMMENTS:

# SUMMARY EVALUATION OF PROYOSEG MODIFICATION TO ARF

Proposed **Modification: F1**
Name: Store Data in Compressed Format

SUMMARY OF PROPOSED MODIFICATION: Computerized data systems such as the ARF may store data in either character or binary formats. In character format, data is stored as alphanumeric characters and must be translated to a binary format in order to be processed. Binary data is already stored in a format recognizable by the computer and therefore does not require translation.

Binary storage is available only for numeric integer or floating point data. Large numbers require less storage space in standard **binary** format. Packed binary data formats allow for the efficient storage of smaller numbers. Character formats ease file development and debugging. Standard dump utilities may be used to provide printouts of data. Customized computer programs may be required to provide similar printouts when data is stored in binary format. In addition, there is little compatibility between the binary formats of the various mainframe, minicomputer, and microcomputer machines while almost all machines recognize ASCII codes for character data.

When making a decision on storage format, the tradeoff is between the "ease of use" and system compatibility of character versus the processing and storage efficiency of binary.

The current ARF processing system produces a character format tape. Most of the intermediate system files are also stored in character format. The proposed modification to the **ARF** system is to convert, where appropriate, the intermediate system files to binary format. The "appropriateness" of the conversion would be determined by the amount of the processing and storage savings which would be realized by the conversion.

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| X | 5 | Cost of ARF Maintenance or Operation |

This modification should reduce the **cost,** space required, and execution times of many **ARF** system programs.

2. Need: Is the problem that the **modification** addresses of critical importance?

1 Critical

Proposed Modification: **F1**

2 Important
x 3 Useful

This proposal addresses the efficiency, and therefore the cost, of the current system.

3. What is the impact on the current ARF system programs and procedures?

The ARF contractor would need to change the input and output statements in most of the programs. In addition, an extra program would most probably be required at the beginning and end of the processing to convert the file into binary format and then back to character format.

4. What is the impact on end users?

This modification will have no impact on the user community.

5. What are the development costs (list activities for which costs should be estimated)?

| | |
|---|---|
| Plan/Design | - decide which files to store in compressed format |
| Write & Test Programs | - write simple program to convert to/from compressed format and modify all programs using that file |
| Documentation of System | - document the file structure and formats of each variable |
| User Documentation | - none |
| Other | |

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low Cost

6. What are the implementation costs?

The implementation cost include the cost of the conversion programs.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low Cost

Proposed Modification: F1

7. What are the Operational Costs?

The operational cost should be marginally less than the cost of running the current system.

ESTIMATED LEVEL OF COST:

| | | |
|---|---|---|
| | 1 | High Cost |
| | 2 | Medium Cost |
| | 3 | Low Cost |
| x | 4 | Minimal Cost |

8. Is Additional User Training Required?

| | | |
|---|---|---|
| | 1 | Yes: |
| X | 2 | No: |

9. Are Additional Hardware or Software Required for the ARF Contractor?

| | | |
|---|---|---|
| | 1 | Yes: |
| X | 2 | No: |

10. Are Additional Hardware or Software Required for the User Community?

| | | |
|---|---|---|
| | 1 | Yes: |
| X | 2 | No: |

11. Are Alternative Approaches Available to Deal with the Problem?

| | | | |
|---|---|---|---|
| x | 1 | Yes: | Maintain Current System. |
| | 2 | No: | |

The alternative is to retain the current system.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

| | | | |
|---|---|---|---|
| x | 1 | Yes: | The question is "How many years does it take for the saving in operational cost to exceed the development cost?" |
| | 2 | No: | |

ANY OTHER COMMENTS: This modification would be a relatively low-cost and simple modification. However, the modifications are not likely to result in major cost savings.

Proposed Modification: F1

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: F2
Name: Expand the Use of Flowcharts in System Documentation

---

SUMMARY OF PROPOSED MODIFICATION: System flowcharts, which present the flow of data and processing through the use of special symbols, are valuable system development, maintenance, and documentation tools. The flowcharts help in organizing and streamlining the many pieces of a complex system during the development phase; provide detailed information on the internal processing performed within each system module, which is important during system maintenance and modification; and provide convenient, pictorial documentation of the system. System flowcharts are also valuable in describing the system to staff who may not have been involved in the development process.

The current ARF system documentation includes a relatively loosely organized set of system flowcharts. Many of the charts are drawn by hand and do not contain detailed input and output references.

The proposed modification is to expand the number and detail of the flowcharts contained in the ARF system documentation. The use of microcomputer-based flowchart software such as EasyFlow 2 (retail price $150) will enable the ARF contractor to develop and maintain a comprehensive set of system flowcharts.

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficult-y Accessing Data |
| x | 5 | Cost of ARF Maintenance or Operation |

The goal of this modification is to reduce future costs of maintaining and modifying the ARF system by improving ARF system documentation.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| X | 1 | Critical |
| | 2 | Important |
| | 3 | Useful |

Future costs will be reduced by improving ARF system documentation.

Proposed Modification: F2

3. What is the impact on the current ARF system programs and procedures?

The ARF contractor would need to review and, where appropriate, modify the system flowcharts for each processing step. The ARF contractor will need to update system flowcharts as future processing is performed.

4. What is the impact on end users?

This modification will have no impact on the user community.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - review all processing
Write & Test Programs - not applicable
Documentation of System - create flowcharts for entire process
User Documentation - not applicable
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

6. What are the implementation costs?

All flowcharts would need to be created a clean format (probably using the software EasyFlow).

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

Update system flowcharts as files are processed.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
3 Low Cost
x 4 Minimal Cost

Proposed Modification: F2

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

X 1 Yes: The contractor probably needs to purchase software for creating flowcharts (i.e. EasyFlow 2).
2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: Maintain Current System.
2 No:

The alternative is to retain the current system.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

Xl Yes: This modification should be relatively easy to implement and would enhance the documentation of the system.
2 No:

ANY OTHER COMMENTS:

Proposed Modification: F2

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: F3
**Name:** Reexamine/Streamline Merge Procedures

---

SUMMARY OF PROPOSED MODIFICATION: A new ARF data file is created by adding data from a transaction file to data from the previous version of the ARF. Selected data from the previous ARF is excluded from the new version in order to adhere to mainframe operating system constraints on record size. Each transaction record contains information about the value of one datum for one county. Thus each transaction record contains a county identifier, a variable identifier, and the value of the variable.

For several of the source data, numerous variables are added for each county. County and variable identifiers are duplicated in each of the transaction records. By altering the processing to allow multiple variables per transaction record, substantial savings in storage and processing costs may be realized.

The proposed modification is to modify the transaction processing procedures for selected files, such as the AHA and AMA data files. Existing transaction processing would be replaced with a procedure which utilized multiple variables per transaction record.

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| x | 5 | Cost of ARF Maintenance or Operation |

This modification should reduce the cost, space required, and execution times for adding data sets with large numbers of variables.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

The processing costs associated with AHA and AMA data would decrease.

3. What is the impact on the current ARF system programs and procedures?

The ARF contractor would modify the EXTRACT and UPDATE programs for source data files with numerous variables, probably executing a unique program for each of these unique sources (thereby eliminating the need to place variable identifiers on the transaction record for each variable). Currently, they execute the same basic programs for all sources.

4. What is the impact on end users?

This modification should have no impact on the user community.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - determine which source files are large enough to justify writing custom merge programs
Write & Test Programs - write program to add large data sets
Documentation of System - document these programs
User Documentation - none
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low Cost

6. What are the implementation costs?

The implementation cost should be just the cost of running the programs one time.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low Cost

7. What are the Operational Costs?

The operational cost should be marginally less than the cost of running the current system.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
3 Low Cost
X 4 Minimal Cost

Proposed Modification: F3

8. Is Additional User Training Required?

| | | |
|---|---|---|
| | 1 | Yes: |
| X | 2 | No: |

9. Are Additional Hardware or Software Required for the ARF Contractor?

| | | |
|---|---|---|
| | 1 | Yes: |
| X | 2 | No: |

10. Are Additional Hardware or Software Required for the User Community?

| | | |
|---|---|---|
| | 1 | **Yes:** |
| X | 2 | No: |

11. Are Alternative Approaches Available to Deal with the Problem?

| | | | |
|---|---|---|---|
| X | 1 | **Yes:** | Maintain Current System. |
| | 2 | No: | |

The alternative is to retain the current system.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

| | | |
|---|---|---|
| Xl | Yes: | The question is "How many years does it take for the saving in operational cost to exceed the development cost?" |
| 2 | No: | |

ANY OTHER COMMENTS:

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: F4
Name: Utilize Procedural Languages More Extensively

SUMMARY OF PROPOSED MODIFICATION: In general, a program written in a lower level language is more efficient than the same program written in a higher level language. However, the writing of the program in the lower level language is often much more time consuming. Many of the procedures the ARF contractor performs each year in SAS could be done more efficiently in a procedural language such as COBOL As the ARF contractor transfers more of the ARF system processing to a microcomputer environment, the time advantages of procedural language programs will become more important.

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| x | 5 | Cost of ARF Maintenance or Operation |

This modification should reduce the cost of creating the ARF.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

The modification will reduce processing time.

3. What is the impact on the current ARF system programs and procedures?

The ARF contractor would replace SAS programs with COBOL programs for any program run on a regular (yearly or less) basis.

4. What is the impact on end users?

This modification should have no impact on the user community.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - determine which programs to rewrite
Write & Test Programs - write program
Documentation of System - document these programs
User Documentation - none
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The implementation cost should be just the cost of running the programs one time.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

7. What are the Operational Costs?

The operational cost should be marginally less than the cost of running the current system.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
3 Low Cost
x 4 Minimal Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

1 Yes:
X 2 No:

Proposed Modification: F4

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 **Yes:** Maintain Current System.
2 No:

The alternative is to retain the current system.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 Yes: The question is "How many years does it take for the saving in operational cost to exceed the development cost?"
2 No:

ANY OTHER COMMENTS:

Proposed Modification: F4

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: F5
Name: Expand Computerized Data Entry, Editing, and Checking Procedures

---

SUMMARY OF PROPOSED MODIFICATION: ARF source data is a combination of computerized data files and printed information from a variety of sources. Data errors are most likely to be introduced into the ARF from sources requiring manual data entry by the ARF contractor. Until recently, most of the data entry was performed using simple text editors, with repeated manual review of data dumps to correct data entry errors. As the ARF contractor has migrated their processing to a microcomputer environment, they have made use of more sophisticated data entry technologies, primarily the use of SAS screens and menus. The SAS programs perform immediate range checks.

The proposed recommendation is to expand the use of computerized procedures for data entry, editing, and checking.

Numerous software packages specifically designed for data entry are available and may prove superior to the custom SAS programs which have been and which are currently being developed by the ARF contractor.

The remaining manual review of data may be augmented by more elaborate and comprehensive computerized editing and checking procedures. The proposal is to review existing editing and checking procedures and automate them as much as possible. Such editing and checking procedures may include computerized algorithms which flag or edit datum which differ in some significant way to previous values or with values for similar counties.

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| x | 5 | Cost of ARF Maintenance or Operation |

This modification should speed manual data entry, editing, and checking and result in fewer data errors.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Proposed Modification: F5

These modifications would improve the most error-prone part of the current system and reduce costs associated with manual data entry.

3. What is the impact on the current ARF system programs and procedures?

The ARF contractor would enter data using the selected software and utilize computerized algorithms to supplement manual editing and checking procedures.

4. What is the impact on end users?

This modification should have no impact on the user community.

5. What are the development costs (list activities for which costs should be e&mated)?

Plan/Design - determine which software to use; design computerized data entry routines; design computerized checking and editing algorithms
Write & Test Programs - data entry and editing/checking routines
Documentation of System - data entry and editing/checking routines
User Documentation - add descriptions of editing/checking process where relevant
Other - Describe

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The implementation cost should be the time required to enter the data.

ESTIMATED LEVEL OF COST:
1 High Cost
x 2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

The operational cost should be less than the cost of running the current system, both to enter and to process the data.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
3 Low Cost

Proposed Modification: F5

x 4 Minimal Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

X 1 Yes: The contractor would have to purchase the data entry software.
2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: Maintain Current System.
2 No:

The alternative is to retain the current system.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 Yes: The data entry component of this modification should be relatively easy to implement and should help the contractor save money almost immediately. Implementation of the editing/checking procedures will require more time due to the more involved nature of identifying and specifying the required algorithms.
2 No:

ANY OTHER COMMENTS:

Proposed Modification: F5

SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

---

Proposed Modification: F6
**Name:** Implement Artificial Intelligence-based Edit/Checking Procedures

---

**SUMMARY OF PROPOSED MODIFICATION:** This proposed modification is an extension of F5 (Expand Computerized Data Entry, Editing, and Checking Procedures). Under this modification, formal editing and checking rules and relationships would be specified and impiemented using Artificial Intelligence methodologies. The intention would to dramatically reduce the amount of manual data review required.

---

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| x | 5 | Cost of ARF Maintenance or Operation |

This modification should speed manual data entry, editing, and checking and result in fewer data errors.

**2. Need:** Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| | 2 | Important |
| x | 3 | Useful |

These modifications would reduce the amount of manual data review required.

**3.** What is the impact on the current ARF system programs and procedures?

The ARF contractor would utilize AI-based algorithms to replace manual editing and checking procedures.

4. What is the impact on end users?

This modification should have no impact on the user community.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - design AI-based checking and editing algorithms
Write & Test Programs - AI editing/checking routines

Documentation of System - AI editing/checking routines
User Documentation - add descriptions of editing/checking process where relevant

Other -

ESTIMATED LEVEL OF COST:

x 1 High Cost
2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The implementation cost should be the time required to enter the new algorithms.

ESTIMATED LEVEL OF COST:

1 High Cost
x 2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

The operational cost should be less than the cost of running the current system, both to enter and to process the data.

ESTIMATED LEVEL OF COST:

1 High Cost
2 Medium Cost
x 3 Low Cost
4 Minimal Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

X 1 Yes: The contractor would have to purchase the AI software.
2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

Proposed Modification: F6

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: Maintain Current System
2 No:

The alternative is to retain the current system.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

1 Yes:
X 2 No: The AI algorithms will be very difficult to design. AI algorithms are most useful in situations involving problems which are repeated often. This is not the case with the ARF system. Although it is a relatively straight forward process to identify data problems, it is often difficult and time-consuming to determine the source and preferred correction to the error. These problems are often not repeated. Therefore, it is likely that many of the algorithms developed based on past experience will have little relevance to future problems,

ANY OTHER COMMENTS

Proposed Modification: F6

# SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed Modification: F7
Name: Data Base Management System

SUMMARY OF PROPOSED MODIFICATION: Currently, each time the ARF contractor adds or modifies any variable on the basic ARF, he reads and writes the entire file (about 90 MB). Instead of this process, the contractor could keep a number of separate files, organized by source and subject matter. To add a new variable, the contractor would add the variable to one of the much smaller subfiles. In addition, the subfiles could reduce the amount of storage required by not containing records for counties with no data on a particular subject.

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| x | 5 | Cost of ARF Maintenance or Operation |

This modification would be enacted to reduce maintenance and operation costs.

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| X | 2 | Important |
| | 3 | Useful |

Efficiency does not seem to be a problem but can always be improved.

3. What is the impact on the current ARF system programs and procedures?

The ARF contractor would have to change many of their programs drastically. The EXTRACT/UPDATE procedure to modify the big, flat file would now have to be adapted for numerous small files. In addition, the contractor might need to add an additional program at the end of the process to create a big, flat file (to deliver the ARF in its current form).

4. What is the impact on end users?

There would be no impact on end users.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - decide which variables and sources to include in which files
Write & Test Programs - write and test all the programs to create and update these files
Documentation of System - document the entire new system
User Documentation - none
Other -

ESTIMATED LEVEL OF COST:
x 1 High Cost
2 Medium Cost
3 Low Cost

6. What are the implementation costs?

Numerous programs would have to be run to create the various related files.

ESTIMATED LEVEL OF COST:
x 1 High Cost
2 Medium Cost
3 Low Cost

7. What are the Operational Costs?

The cost of maintaining the system should be less than the cost of maintaining the current system.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
x 3 Low Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

X 1 Yes: The contractor might need to purchase some DBMS software to manage the files.
2 No:

Proposed Modification: F7

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
X 2 No:

11. Are Alternative Approaches Available to Deal with the Problem?

Xl Yes: Maintain Current System.
2 No:

The alternative to this modification is to retain the current file structure.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 Yes: It might not be worth the cost of converting to a new system.
2 No:

ANY OTHER COMMENTS:

Proposed Modification: F7

## SUMMARY EVALUATION OF PROPOSED MODIFICATION TO ARF

Proposed **Modification:** F8
**Name:** Moving More Processing to the PC

SUMMARY OF PROPOSED MODIFICATION: The idea of this modification is to move as much processing as possible to the PC to save on mainframe computing time.

1. Goal: Which problem area does the modification address?

| | | |
|---|---|---|
| | 1 | Underutilization |
| | 2 | Content & Structure of Data |
| | 3 | Content & Structure of Documentation |
| | 4 | Difficulty Accessing Data |
| x | 5 | Cost of ARF Maintenance or Operation |

2. Need: Is the problem that the modification addresses of critical importance?

| | | |
|---|---|---|
| | 1 | Critical |
| x | 2 | Important |
| | 3 | Useful |

Efficiency does not seem to be a problem but can always be improved.

3. What is the impact on the current ARF system programs and procedures?

The contractor would move processing to the PC whenever possible. Processing on the PC is somewhat different than on the mainframe. In general, more user-friendly software is available on the PC and there is no cost for computer runs. However, disk space can be a problem and long runs can take much longer.

4. What is the impact on end users?

There would be no impact on end users.

5. What are the development costs (list activities for which costs should be estimated)?

Plan/Design - determine which applications are portable to the PC
Write & Test Programs - modify (or completely rewrite) mainframe programs to run on the PC
Documentation of System - often more difficult on the PC because copies of batch run are not always printed out

Proposed Modification: F8

User Documentation - none
Other -

ESTIMATED LEVEL OF COST:
1 High Cost
X 2 Medium Cost
3 Low Cost

6. What are the implementation costs?

The cost should be minimal.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low Cost

7. What are the Operational Costs?

The cost should be significantly less than the cost of the current system.

ESTIMATED LEVEL OF COST:
1 High Cost
2 Medium Cost
X 3 Low Cost

8. Is Additional User Training Required?

1 Yes:
X 2 No:

9. Are Additional Hardware or Software Required for the ARF Contractor?

X 1 **Yes:** The contractor might have to buy software and hardware to move some of the applications to the PC.
2 No:

10. Are Additional Hardware or Software Required for the User Community?

1 Yes:
x 2 No:

Proposed Modification: F8

11. Are Alternative Approaches Available to Deal with the Problem?

X 1 Yes: Retain the current system.
2 No:

The DBMS modification (F7) and the microcomputer delivery methods (E8-E11) are related to this modification.

12. Is Feasibility of Success Reasonably High? Is the Modification Likely to Solve the Problem?

X 1 Yes: This modification is already being implemented. More applications will be moved to the PC if they would save money. Therefore, by definition, this modification will achieve its goal to some degree.
2 No:

ANY OTHER COMMENTS:

Proposed Modification: F8

**MATHEMATICA**
Policy Research, Inc.

PO. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535

600 Maryland Avenue S.W., Suite 550
Washington, DC 20024-25 12
(202) 484-9220