MEETING FEDERAL HEALTH MANPOWER STATISTICAL SYSTEM REQUIREMENTS WITH SAS

Steven R. Borbash
West Virginia University

1. Introduction

This paper describes some details of a successful attempt to utilize SAS for two purposes: (1) nearly all data processing work to create and maintain an original health manpower data base, and (2) statistical analysis of the data base. The system described is the West Virginia Health Manpower Statistical System (WVHMSS). It is felt that SAS is ideal for such applications. It is easy to use, result-oriented and relatively economical to run. The WVHMSS application is programmed mainly with SAS-72 supplemented by two special user-written procedures.

Background on the Federal Comprehensive Health Statistics System is presented. Then the design of the WVHMSS is discussed with emphasis on the use of SAS for data processing work.

Financial support for this work was provided by a grant from the West Virginia Regional Medical Program to the West Virginia University Department of Industrial Engineering.

2. The Comprehensive Health Statistics System

The National Center for Health Statistics (NCHS), within the U.S. Department of Health, Education and Welfare launched a new program, the Cooperative Health Statistics System (CHSS) in 1970. This program was authorized through Public Law 91-515. Its objective has been to design and implement a cooperative Federal-State-Local system for producing comparable and uniform health information and statistics in the following seven areas [4].

1. Manpower statistics (inventories and surveys).
2. Health facilities statistics (inventories and surveys).
3. Hospital care statistics (short-stay hospitals).
4. Household interview statistics.
5. Ambulatory care statistics.
6. Long-term care statistics.
7. Vital statistics.

Initial CHSS efforts have been concentrated in the three areas of vital statistics, health manpower statistics and health facilities statistics. In late 1971 a research and development phase was begun to investigate systems design concepts, and in mid-1973 a limited operational phase was set up to establish data collection systems within several states to test these concepts. The design objective is to establish and maintain machine-readable data bases within the states which will provide information necessary for local state and federal health planning activities. These local data bases will be developed and maintained around a standardized core set of data desired at the national level by NCHS. It is hoped that this concept will help eliminate much of the uncoordinated and duplicated data gathering which in the past has produced unreliable results. In return for receiving data from the states and setting standards for the core data set, NCHS plans to subsidize state data system costs to pay for a fair share of the data.

3. The Health Manpower Statistics

In the area of health manpower statistics, the core data base is designed to provide information about the numbers, characteristics and distribution of health care personnel licensed by a given state. At the present time this includes the following licensed occupations and will probably be expanded in the future to cover others.

1. Chiropractors
2. Dental Hygientists
3. Dentists
4. Doctors of Medicine
5. Doctors of Osteopathy
6. Licensed Practical Nurses
7. Nursing Home Administrators
8. Optometrists
9. Pharmacists
10. Physical Therapists
11. Podiatrists
12. Registered Nurses
13. Veterinarians

The federal core data set is identical for all the above professions. It contains the name, address, year of birth and license number as well as a small amount of extra qualifying data for each licensed health care provider. This data set covers 100% of the licenses of each profession listed above, provides a basic inventory of health care providers, and functions as a basis for sampling studies by questionnaire for more detailed data. Collectively, the data sets for the individual licensing agencies make up the data base for a statewide Health Manpower Statistical System (HMSS).

A distinction is made by NCHS between a stage I and stage II core data set. Stage I was as described above, with stage II including all the items in stage I, but in addition having information on type and place of education, specialty, practice setting, etc. Stage II data is the ultimate goal for the federal core data set with stage I being accepted as a starting point for expansion to stage II at a later date.

Establishing and maintaining a health manpower statistical system within a state requires cooperation between the individual licensing agencies, the state agency coordinating the data collection and processing, a data processing center and the NCHS. The stage I core data is obtainable directly (with consent) from the licensing agencies, which are generally required by state law to collect all the required information. This data is often not kept in machine-readable form by the agencies, particularly the smaller ones. The task of establishing and maintaining a statewide machine-readable data base is entrusted to a 'lead agency' within the state, usually the State Department of Health. Licensing agencies are often legislatively required to share their stage I data with the lead agency, or find it economically advantageous to do so because some or all of their data entry and processing costs will be assumed by the lead agency. The lead agency is in a unique position to provide services to the agencies such as statistical summaries, lists of licensees for sale to advertisers, mailing label printing, license printing, etc. Thus, the lead agency can function as a central data processing center providing free or low cost services to the licensing agencies. Costs of operating a stete HMSS are shared between the state government, the federal government and the state licensing agencies on a negotiated basis.

4. The West Virginia Health Manpower Statistical System

The West Virginia Health Manpower Statistical System (WVHMSS) is a cooperative venture between the West Virginia State Department of Health, the licensing agencies in the State of West Virginia and the NCHS. Collection and conversion of agency data to machine-readable form for the stage I inventroy is nearly completed and has been done by the West Virginia State Department of Health and the West Virginia State Computer Center in Charleston, West Virginia.

Software for the system is nearly all SAS-72 based. Funds for software research and development were provided in advance of NCHS assistance by the West Virginia Regional Medical Program through a grant to the West Virginia University Department of Industrial Engineering. The software development and testing is being done at the campus of West Virginia University in Morgantown, West Virginia by processing the agency data sets as they are gathered. The entire software package will be transferred shortly to the State Department of Health for ongoing operations. The West Virginia State Computer Center which supports the SAS-72 package is located adjacent to the State Health Department and will provide convenient data processing services for the WVHMSS as the development efforts of the Industrial Engineering Department are phased out. Tentative future plans for the WVHMSS include expansion of the agency data sets to include the stage II items.

5. WVHMSS Data Processing System Description

The WVHMSS data processing system is subdivided into two sections, a data entry section shown on Fig. 1 and a main processing section shown in Fig. 3. In the data entry section, licensee records from each licensing agency are converted to machine-readable form with the proper format and subjected to a rough visual check. SAS is not used in the data entry section, which relies on COBOL and PL/I programs for special processing.

The main processing section is written exclusively in SAS. In this section, further detailed cleanup of the files is done and geographical data (county and district names, latitude and longitude) is added to the records to give a final data set for state use. This final file (the AGENCY.C file) is then used as input for a variety of other SAS programs which perform statistical summaries, print mailing lists and questionnaires, and produce a data set for NCHS (the AGENCY.D file) by abstracting information from the AGENCY.C file. SAS was chosen for the following reasons:

(1) SAS is an easy language to program in, and promised potential savings in programming time and development effort.
(2) Personnel who operate the WVHMSS will probably not be computer professionals. It was felt that SAS programs would be easier for them to understand, use, and modify than programs in another language.
(3) For the sequential file processing used throughout the WVHMSS, SAS was more than adequate.

FIGURE 1
DATA ENTRY SECTION - WVHMSS

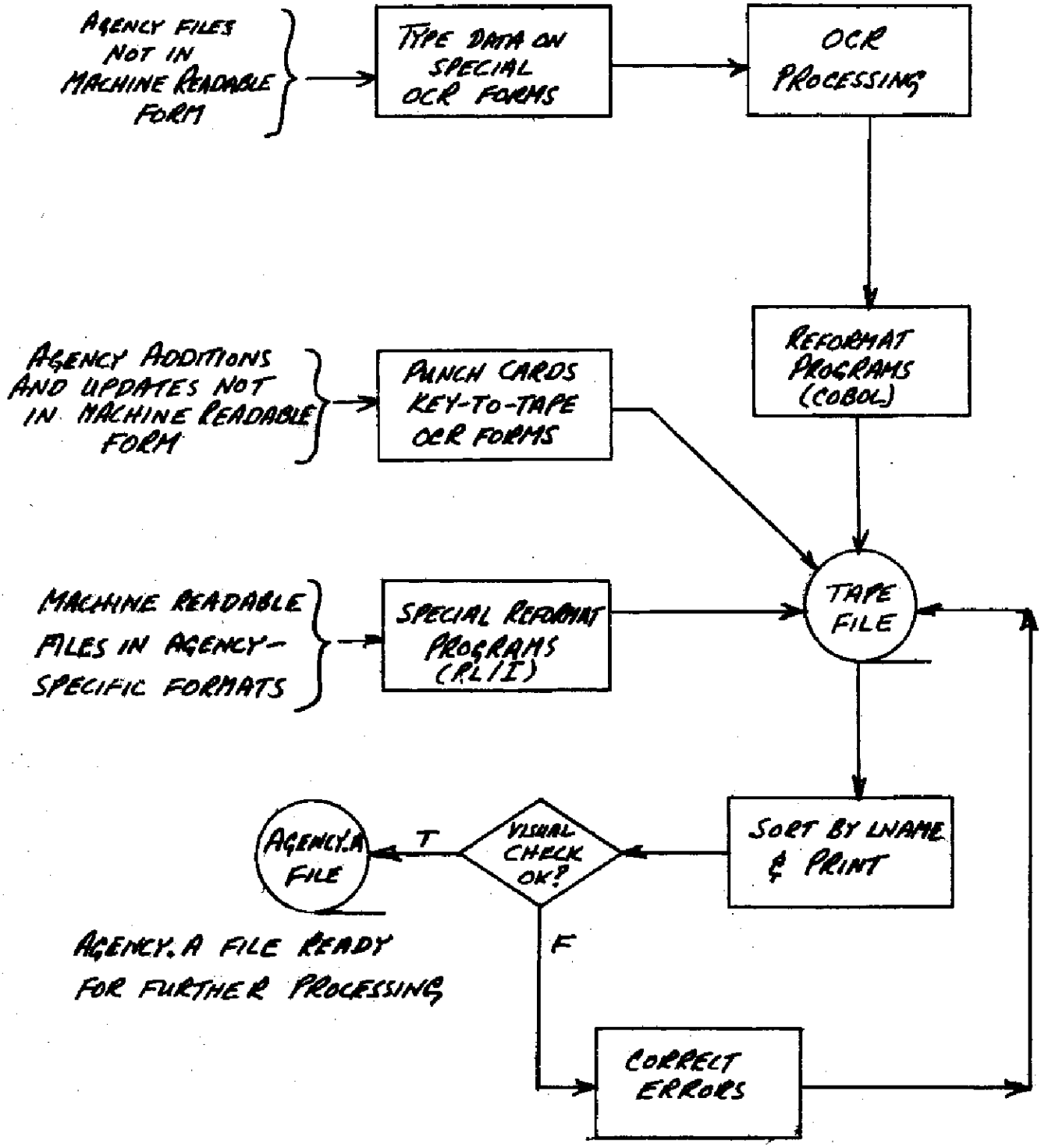


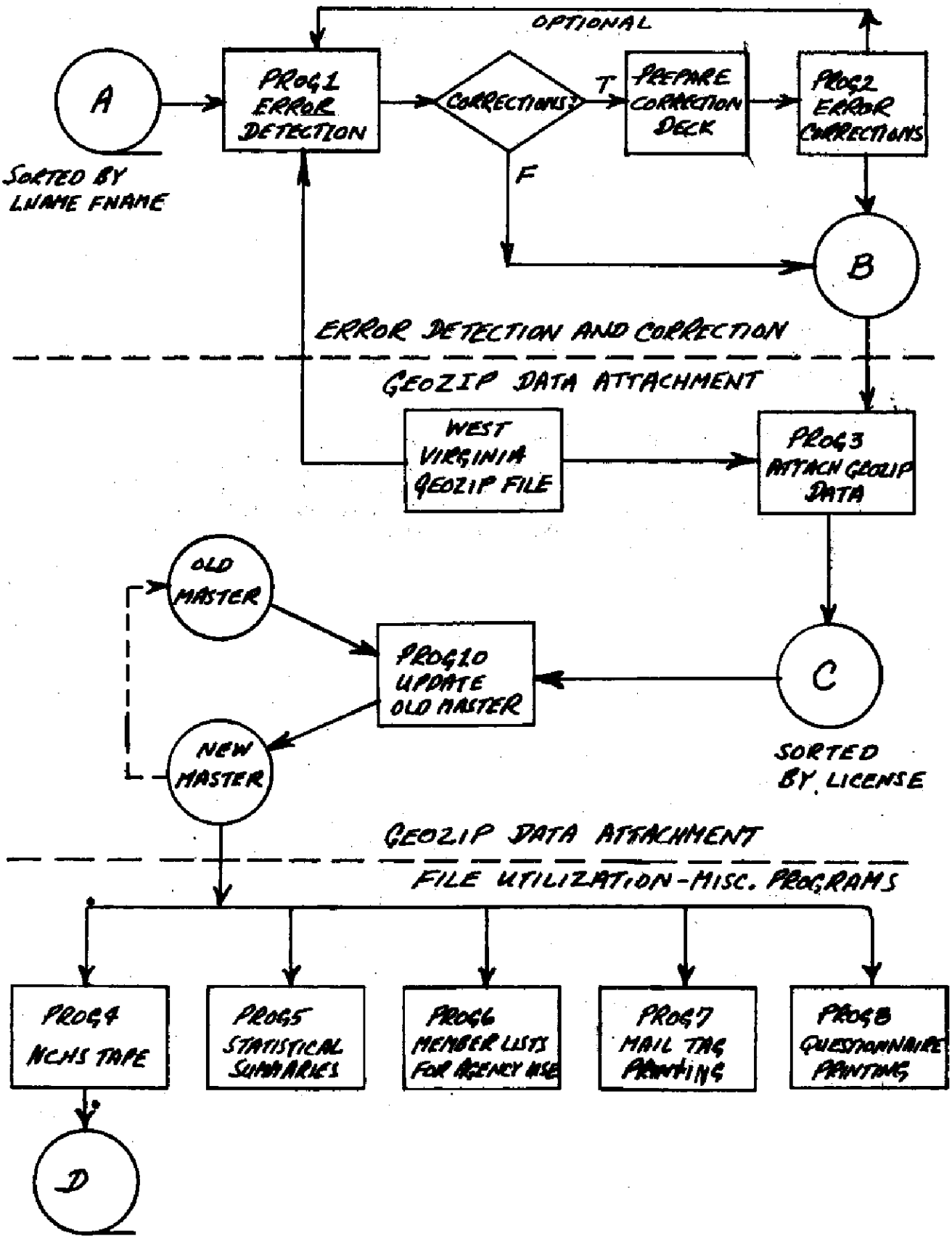
FIGURE 2

SAS INPUT STATEMENT FOR AGENCY.A FILES
(DATA GATHERED FROM LICENSING AGENCIES)

```
INPUT
   LNAME $ 1-17
   FNAME $ 18-33
   STADDR $ 34-53
   CADDR $ 54-68
   STATE $ 69-70
   ZIP 71-75
   ADDRCODE 76
   LICENSE 77-86
   LMO 87-88
   LYR 89-90
   BMO 91-92
   BDAY 93-94
   DYR 95-96;

COMMENT
    FOR FOREIGN COUNTRY RESIDENTS, STATE='FC' AND ZIP IS A
    NEGATIVE THREE-DIGIT FOREIGN COUNTRY CODE.  MIDDLE INITIAL IS
    A PART OF FNAME (BLANK USED FOR A SEPARATOR).  GENERATIONAL
    IDENTIFIER (JR,SR,III,ETC.) IS A PART OF LNAME (DASH USED FOR A
    SEPARATOR).  LMO AND LYR GIVE DATE OR ORIGINAL WV LICENSURE.
//
//
```

FIGURE 3
MAIN SAS PROCESSING SECTION - WVHMSS



The WVHMSS requires only SAS-72, except for PROC FORMS, which is a part of SAS-75. Two non-standard SAS procedures are used with the WVHMSS SAS-72 system. PROC EDIT is used to edit character strings and was developed especially for this project. PROC OUTPUT is used to convert data sets from SAS Internal format (SASFMT) to fixed block format (FBFMT). This PROC is used extensively and was developed independently by Mr. Daniel Chilko who is associated with the West Virginia University Computer Center [2]. Its operation is similar to the PUT statement now available in SAS-75.

5.1 Data Entry Section

Fig. 1 shows three types of data input to the data entry section. These are: (1) entire agency files not in machine-readable form for initial conversion; (2) additions and updates not in machine-readable form; and (3) machine-readable files in agency-specific formats.

For agency files not in machine-readable form, the information shown in Fig. 2 was typed directly onto special optical character reader (OCR) forms for each licensee. This information was obtained directly from active membership lists and records obtained from the licensing agencies. The typing was done on an IBM selectric typewriter with a special type ball for OCR use. For foreign residents (maintaining an active West Virginia license, but residing in a foreign country), the two character state code was entered as 'FC' and the U.S. zip code was replaced with a negative three digit foreign country code number. (Geographical location and zip coding throughout WVHMSS conforms to U.S. Government standards set forth in references [3] and [5]). After the information is typed onto OCR forms, the forms are processed by an OCR to give a tape file which is then reformatted with a special COBOL program. Additions and updates to the agency files are batched and submitted to a similar conversion process but depending on the batch size may be entered directly onto cards, or keyed onto tape as an alternative to the OCR conversion.

Only one of the WV licensing agencies (the Registered Nurses) had a pre-existing machine-readable membership file. This file was preprocessed with special PL/I programs to put it into the proper WVHMSS format. One typical problem with this file was a 2-character state code embedded in the city address field. This problem was solved by scanning each address field to find the state code which was then extracted and placed in a separate field. This file is the largest in the system (11,000 of the 21,000 total records).

Finally, all of the files in machine-readable form are sorted by LNAME, printed and submitted to a rough visual error check. They are then identified with the AGENCY.A label and sent to the main SAS section of WVHMSS for further processing.

5.2 Main Processing Section

The main processing section can be subdivided into three components: (1) error detection and correction, (2) geozip data attachment, and (3) a

miscellaneous programs section as shown on Fig. 3. These subsections will be discussed separately below.

5.2,1 Error Detection and Correction Subsection

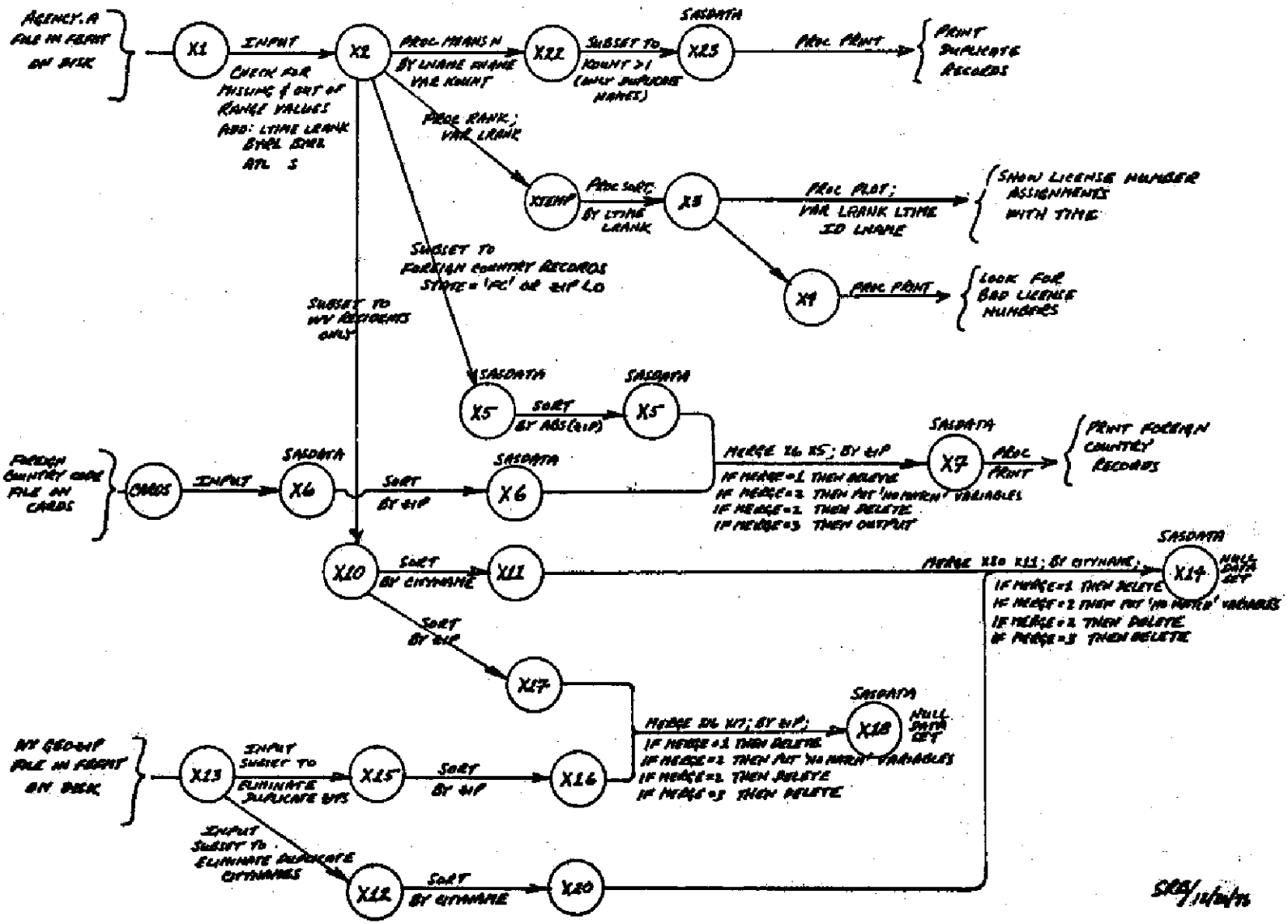
This subsection uses PROG1 and PROG2 to detect and correct obvious errors missed in the data entry (typing) process as well as more subtle errors of logical inconsistency in the files. The output is the AGENCY.B file which is hopefully error free. NCHS requirements specify a bad data tolerance limit of 1% or less for any given file item, and this subsection is designed to provide help in meeting this specification. PROG1 is the largest and most complex of all the programs in the system. Its function is to comb the incoming AGENCY.A file for errors. All agency files are tested for missing and out of range data items, duplicate records, license numbers out of sequence, inconsistencies between LYR (year of initial licensing) and BYR (birth year), as well as improper zip codes and misspelled place names. For the latter two items, the West Virginia GEOZIP File is used as input to PROG1, since it provides a list of all WV zip codes and corresponding place names with acceptable spellings.

An information flow diagram for PROG1 is shown in Fig. 4. Data sets created or used as input are labeled inside circles. Directed lines between circles are labeled to explain how the resultant data set was formed, such as by a SAS PROC which produces an output data set (MEANS, SORT, RANK,etc.), or as the result of an INPUT, SET or MERGE statement. These information flow diagrams, in conjuction with individual program writeups and program listings with COMMENT statements are used to document all the WVHMSS SAS programs. The flow diagrams have proved to be an especially effective documentation tool. For example, refer to the top of Fig. 4. After the AGENCY.A file (X1) is converted to SASFMT from FBFMT with an input statement giving (X2), duplicate records are located by identifying the duplicates with PROC MEANS N NOPRINT OUT = X22; giving (X22). Then (X23) is formed by subsetting all records with KOUNT > 1. These are finally listed for examination with PROC PRINT.

All data sets in the diagram have a separate JCL DD statement except those labeled with SASDATA over their circles on the diagram, such as (X23). These data sets are small and are stored in SASDATA by default. The SPACE parameter in each DD statement is set to accommodate all agency data sets without modification. In PROG1 all data sets except (X1) and (X3) are temporary, in SASFMT, and stored in a 2314 type disk pack. For a file with 10,000 licensees, about 40% of the total disk pack is utilized for temporary storage by PROG1.

The variable LICENSE is the unique identifier for the WVHMSS files. The final AGENCY.C master file is kept sorted by LICENSE and is updated using a SAS MERGE by LICENSE. For this reason, it is important that incorrect and duplicate license numbers be eliminated from the agency files as soon as possible. To identify questionable license numbers, each number is checked against an agency specific range of licenses with program statements as (X2) is built from (X1). Then (XTEMP) is created from (X2) by replacing LICENSE

FIGURE 4 - WYNPSS PROG1
INFORMATION FLOW DIAGRAM



SRS/11/84

by its numerical rank. Out-of-sequence license number identification is based on the assumption that license numbers are assigned sequentially in time by the agencies (true in most cases). The numerical rank of the license numbers and values of LTIME = LMO/12 + LYR both appear in (XTEMP) where LTIME = LMO/12 + LYR. XTEMP is now sorted by LTIME RANK giving (X3) and when a record sequence NUMBER and a DEVIATION = NUMBER-LRANK are added to (X3) giving (X4), it is easy to spot questionable license numbers by printing the file. A printout of a portion of file (X4) is shown in Fig. 5. Any deviation from monotonically non-decreasing values of LRANK indicate that the deviating record has a license number assigned out of sequence or that the time of original licensing (LTIME) is in error. Fractional values of LRANK indicate tied ranks and hence identify duplicate license numbers in the file. A printout of the entire file (X4) is visually scanned for these irregularities and the variables NUMBER and DEVIATION simplify the scanning. It is easier to scan the DEVIATION column to spot deviant values of LRANK then it is to look at LRANK itself. Questionable license numbers are checked with the licensing agency and corrections made where necessary. A plot of LRANK vs. LTIME is useful as in indicator of the general condition of license number sequencing in the file. Such a plot is shown on Fig. 6.

It is especially important that all West Virginia residents have valid zip codes because the counties of residence are attached to the license records with PROG3 by matching the zip codes. To check zip codes (X10) is formed with only WV residents. This file is sorted by zip to given (X17). Then the WV GEOZIP File (X13) is read in from disk and duplicate zips are eliminated to give a zip master file (X15) with no duplicate zips. This is sorted by ZIP to give (X16). Then MERGE X16 X17; BY ZIP; produces data set (X18). The SAS internal variable MERGE = 1, 2 or 3 as shown below depending on whether the zip codes in (X16) (all allowable WV zips) match those in (X17) (all WV residents).

Condition	MERGE
Record in (X16) without a zip match in (X17)	MERGE = 1
Record in (X17) without a zip match in (X16)	MERGE = 2
Record in (X16) with a zip match to record in (X17)	MERGE = 3

When MERGE = 2, there is no match in the master zip code file (X16) for a West Virginia licensee. This is an error condition as opposed to MERGE = 1 (unused zip) or MERGE = 3 (matching zip) which are normal conditions. When MERGE = 2, the PUT statement is used to print the record in question, and the erroneous zip can thus be corrected. All records are deleted before being placed in (X14) which ends up with no observations and is merely a convenience (null) data set required to set up the MERGE. A similar process is used to identify records having WV zips but without matching WV city names (common misspellings are allowed). See data sets (X11), (X12), (X20) and (X14).

The code numbers for foreign countries (negative zips) are checked by attaching the actual country name to the number on the records and then printing these records. Errors are identified by noting no match conditions (MERGE = 2) and by checking correspondence of the listed foreign country with the foreign city name.

FIGURE 6
PROG1 - WYNMSS
FILE PRINT AFTER PROC SORT; BY LTIME LRANK

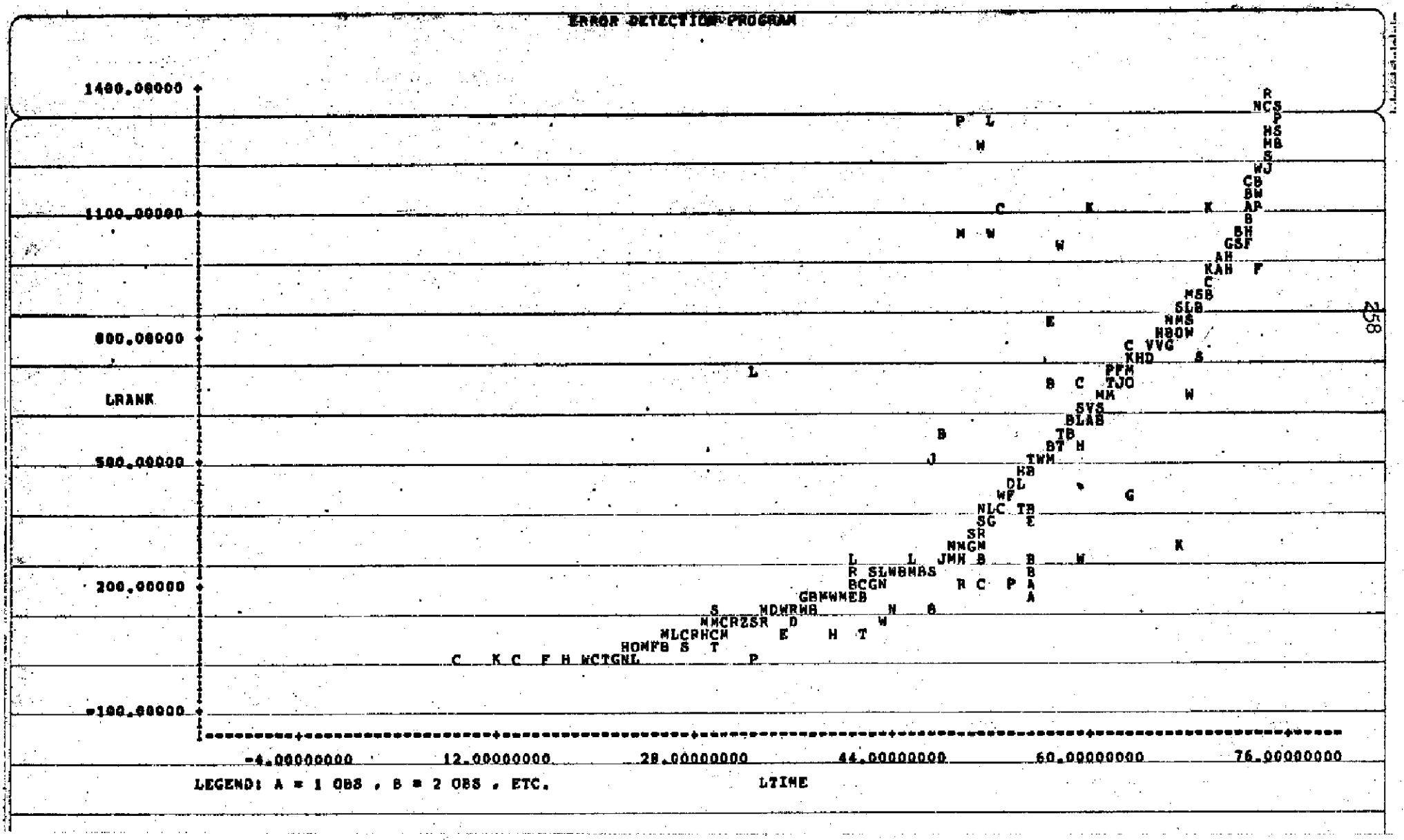
ERROR DETECTION PROGRAM

OBS	LTIME	LRANK	NUM	DEV	LNAME	FNAME	LICENSE
441	53.5833	426.0	441	15.0	STATLER	JACK R	2086
442	53.5833	427.5	442	14.5	GREEN JR	THOMAS C	2087
443	53.5833	427.5	443	15.5	STUMP	MARGARET -	2087
444	53.5833	429.0	444	15.0	DEITZ	CALVIN R	2088
445	53.5833	430.0	445	15.0	DAVIS	ROBERT G	2089
446	53.5833	431.0	446	15.0	THAXTON	H R	2090
447	53.5833	432.0	447	15.0	LESLIE	JAMES K	2091
448	53.5833	433.0	448	15.0	JENNESSEE	CHARLES M	2092
449	53.5833	434.0	449	15.0	FORBES	JOHN H	2093
450	53.5833	435.0	450	15.0	APESOS	CHRISTOPHER -	2094
451	53.5833	436.0	451	15.0	DELLI GATTI	ROBERT E	2095
452	54.1667	375.0	452	77.0	TAYLOR JR	ROY M	2028
453	54.1667	438.0	453	15.0	WALLACE	SARA Z	2098
454	54.1667	439.0	454	15.0	SMITH	BENTON B	2099
455	54.1667	440.0	455	15.0	LEIGHTY	HAROLD D	2100
456	54.1667	441.0	456	15.0	DEWEY	HARRY P	2101
457	54.1667	442.0	457	15.0	HALL	JESSE F	2103
458	54.1667	443.0	458	15.0	KOSTAS	GEORGE A	2104
459	54.1667	444.0	459	15.0	HADDAD	GEORGE -	2105
460	54.5833	165.0	460	295.0	AMICK	WILLIAM R	1722
461	54.5833	199.0	461	262.0	ALLISON	MACK F	1790
462	54.5833	210.0	462	252.0	BALLARD	ROBERT H	1804
463	54.5833	252.0	463	211.0	BARE JR	S E	1979
464	54.5833	351.0	464	113.0	ARRINGTON JR	TED -	1996
465	54.5833	359.0	465	106.0	BAY JR	E H	2004
466	54.5833	445.0	466	21.0	BAUER	SUE -	2107
467	54.5833	446.0	467	21.0	BALSTON III	WINTER B	2109
468	54.5833	447.0	468	21.0	MCNUTT	ALFRED D	2110
469	54.5833	448.0	469	21.0	LEWIS	ROBERT R	2111
470	54.5833	449.0	470	21.0	CARSON	BENJAMIN H	2112
471	54.5833	450.0	471	21.0	RIGGS	JACK A	2113
472	54.5833	451.0	472	21.0	DURANT	FRANCIS C	2114
473	54.5833	452.0	473	21.0	WATZ	ROBERT J	2115
474	54.5833	453.0	474	21.0	GOSNEY	MARTHA -	2117
475	54.5833	454.0	475	21.0	MILLER	WAVERLY L	2118
476	54.5833	455.0	476	21.0	HESS	VERNON L	2119
477	54.5833	456.0	477	21.0	SLAVEN	MARCIA M	2120
478	54.5833	457.0	478	21.0	SHUMATE	WILLIAM J	2121
479	54.5833	458.0	479	21.0	POZEGA	JEAN -	2122
480	54.5833	459.0	480	21.0	HARDMAN	RUDY C	2123
481	54.5833	460.0	481	21.0	HUGGINS	JACK S	2124
482	54.5833	461.0	482	21.0	TURNER	ERNEST -	2125
483	54.5833	462.0	483	21.0	ESHEM	FRANK C	2127
484	54.5833	463.0	484	21.0	SWEARINGEN	JAMES W	2128
485	54.5833	464.0	485	21.0	RAYMOND	DOMINICK J	2129
486	54.5833	465.0	486	21.0	SUPPA JR	SAM -	2130
487	54.5833	466.0	487	21.0	HURL	DONALD E	2131
488	54.5833	467.0	488	21.0	RAIMONDI	FLORENCE E	2132
489	55.0833	347.0	489	142.0	BUTLER JR	R L	1992
490	55.0833	468.0	490	22.0	DOZIER	ROBERT L	2133
491	55.1667	334.0	491	157.0	EVANS	CHARLES H	1976
492	55.1667	469.0	492	23.0	AZAR	LOUIS J	2134
493	55.1667	470.0	493	23.0	BECK	ROBERT M	2135
494	55.1667	471.0	494	23.0	WILKS	DAVY D	2136
495	55.1667	472.0	495	23.0	THORNTON	CLARK M	2137

FIGURE 6

PROG1 - WVHMSS

PLOT OF LTIME VS. LRANK



After PROG1 has been run and errors identified, corrections are made with PROG2, which uses IF-THEN statements to correct license numbers and eliminate duplicate records followed by a MERGE (BY LICENSE) of the AGENCY.A file against a corrections file entered via the card reader. For excessively 'dirty' files, another pass through PROG1 is sometimes made.

5.2.2 Geozip Data Attachment Subsection

Here the corrected AGENCY.B file is augmented by attaching (to the records of West Virginia resident licensees only) the county names and numbers, minor civil division names and numbers, and latitude and longitude of the post office (used for mapping and other locational studies). All these variables are in the GEOZIP file. This is accomplished by first eliminating all duplicate zip records from the GEOZIP file and then sorting it BY ZIP. Next the West Virginia licenseees are selected out of the AGENCY.B file and this WVRES file sorted BY ZIP. The GEOZIP data is then attached with the following statements [1].

```
DATA TOGETHER;
  MERGE GEOZIP WVRES; BY ZIP;

IF MERGE = 1 THEN DELETE;
IF MERGE = 2 THEN PUT 'ERROR' ZIP NAME;
IF MERGE = 2 THEN DELETE;
IF MERGE = 3 THEN OUTPUT;
```

After putting the WVRES file back together with the nonresidents (with SET statements) and sorting by LICENSE, the AGENCY.C file results. This file is used to update an old master (if there is one). The new master AGENCY.C file forms the input for the remaining SAS programs in the WVHMSS.

5.2.3 Miscellaneous Programs Subsection

The AGENCY.C files are utilized by a battery of SAS programs as shown in Fig. 3. PROG5 produces statistical summaries of licensees by state, county and health planning region using PROC FREQ and PROC MEANS. PROC HIST is used to study licensee distribution by age. The numbers of licensees per 10,000 people are computed (using census data) to identify underserved areas of the state. PROG6, PROG7, and PROG8 utilize PROC FORMS to print mailing labels, questionnaires and member lists for agency use.

PROG4 produces the AGENCY.D file for NCHS use. This program is flow charted on Fig. 7. Certain selected variables from the AGENCY.C file are read in. Several new variables are added at this point, such as LICSTATE = '54' for West Virginia and AGENCY = 'XX' for the particular agency. Then PROC EDIT is used to separate the middle initial from the last name field. This is a requirement for the AGENCY.D file. This PROC was written especially for this application. Any character substring can be 'pulled out' of a main string given by the VARIABLE card and the resulting gap closed to one character with left shifts. To remove the middle initials from the FNAME field the

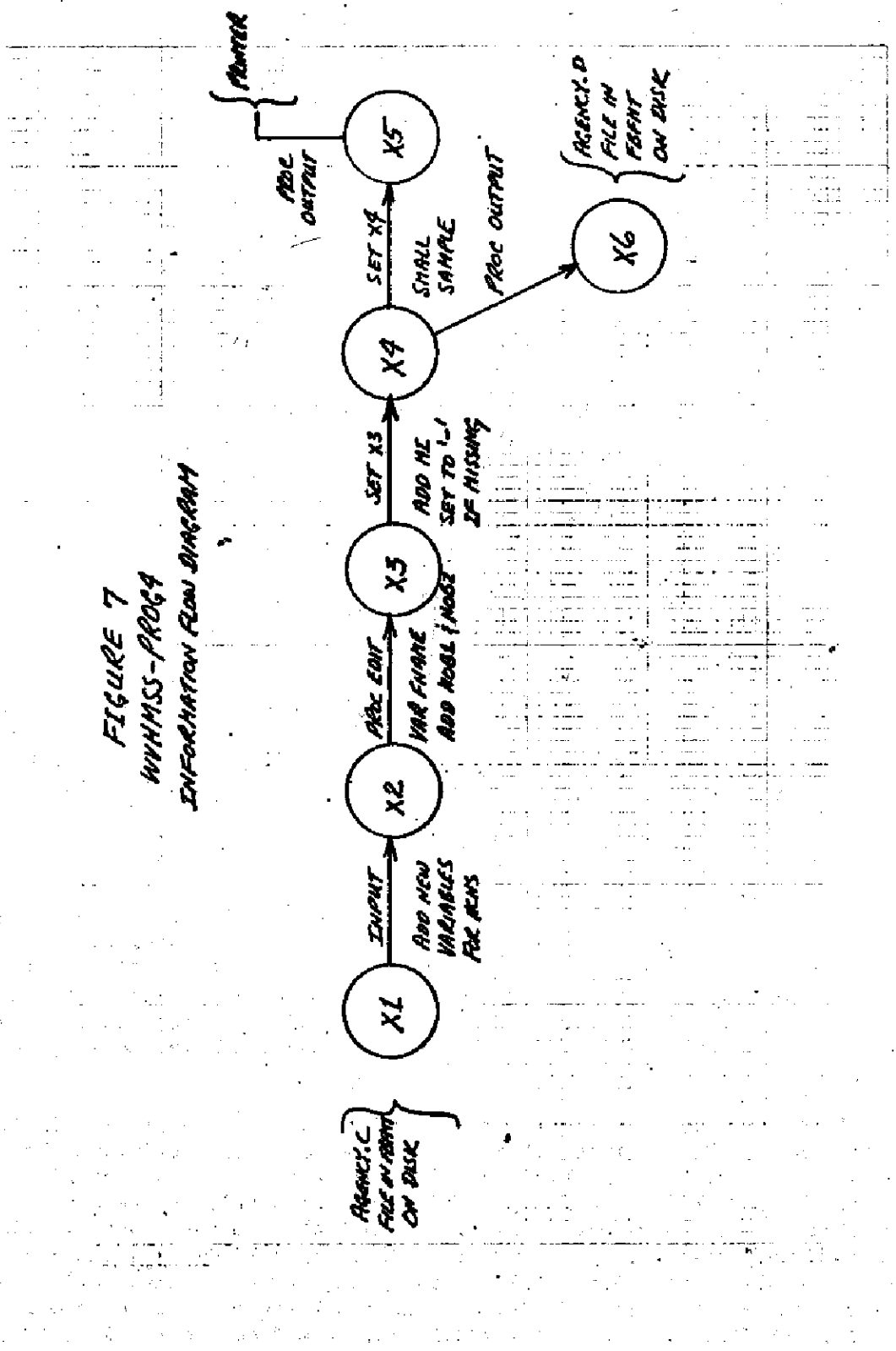


FIGURE 7
WHMIS-PROG4
INFORMATION FLOW DIAGRAM

following is required:

```
PROC EDIT OUT = X3;
VAR FNAME;
PARMCARDS:
   *A*
   *B*
    .
    .
    .
   *Z*
    ;
```

Each character substring to be lifted out is enclosed with asterisks on a separate PARMCARD. Only the first substring found is removed. The PROC creates two new variables NOB1 and NOB2 and adds them to the output data set. NOB1 is initialized to the original character string. NOB2 is a character variable initialized to blanks and having length as long as the longest substring on the PARMCARDS. PROC EDIT scans NOB1 with the character strings on successive PARMCARDS until a match is found or all PARMCARDS have been exhausted. If no match is found, NOB1 and NOB2 are left at their initial values. When the first matching substring is found, it is removed from NOB1, placed (left adjusted) into NOB2, and PROC EDIT moves on to the next record.

The final output data set (X6) in PROG4 is produced with PROC OUTPUT. This PROC was designed for general utility purposes, and has been very useful for WVHMSS work. PROC OUTPUT takes SAS-72 records out of SASFMT (variable block length) and places them into FBFMT (fixed block format) with format control on individual variables. The SAS PROC OUTPUT statements used in PROG4 are shown in Fig. 8.

5.3 Operating Costs

Rough operating cost estimating equations for the WVHMSS SAS programs have been derived based on the operating experience for processing about 10,000 records using an IBM 360/75 computer at West Virginia University. They are listed below (K is agency file size in thousands and C is the dollar cost of one run).

PROGRAM	EQUATION	C FOR K=1 (1000 RECORDS)
PROG1	C = 6.5 + (14.0) K	$ 20.50
PROG2	C = 1.5 + (7.5) K	9.00
PROG3	C = 2.5 + (4.0) K	6.50
PROG4	C = 2.0 + (10.0) K	12.00
PROG5	C = 5.0 + (1.0) K	6.00
TOTAL	C = 17.5 + (33.0) K	$ 54.00

PROG1 is the most expensive followed by PROG4 and PROG2. The cost per record varies with file size. For files with K=1 (1000 records), the total cost (PROG1 to PROG5) per record is about $.05. For small files with 100 records, this increases to about $.20 per record. For files with 5000 records

FIGURE 8

RELATIONSHIP BETWEEN AGENCY.A AND AGENCY.D FILES

```
COMMENT
    DATA EXTRACTED FROM AGENCY.C FILE FOR AGENCY.D FILE;

DATA AGENCY_C;
  INPUT
  LNAME_GI $ 1-17
  FNAME_MI $ 18-33
  ZIP $ 71-75
  ADDRCODE $ 76
  LICENSE $ 77-86
  BMO $ 91-92
  BYR $ 95-96
  CNTYNUM $ 112-114
  RESSTATE $ 171-172;

COMMENT
    STRUCTURE OF AGENCY.D FILE;

PROC OUTPUT;
 PARMCARDS;
AGENCY_D(80)/LICSTATE($ 1-2) RECTYPE($ 3-4) AGENCY($ 5-6)
             LICENSE($ 7-15) IDTYPE($ 16) RESSTATE($ 17-18)
             ZIP($ 19-23) CNTYNUM($ 24-26) ADDRCODE($ 27)
             LICDATE($ 28-31) BMO($ 32-33) BYR(34-36)
             LNAME_GI($ 37-53) FNAME($ 54-65) MI($ 66)
             OLDNAME($ 67-80)
             ;

COMMENT
    VARIABLES CREATED IN PROG4 INCLUDE
         LICSTATE = '54' FOR WV
         RECTYPE = '10' FOR NEW RECORDS
         AGENCY = '01' TO '13' DEPENDING ON AGENCY
         IDTYPE = '01' TO '04' FOR TYPE OF LICENSE NUMBER
         LICDATE = 'MOYR' FOR LAST AGENCY RENEWAL DATE
         MI = 'X' MIDDLE INITIAL
         OLDNAME = BLANKS RESERVED FOR NAME CHANGES;
//
//
```

the cost per record drops to $.037. These figures do not include costs incurred in the data entry section, software development costs, or GEOZIP file construction costs.

6. Conclusions

The use of SAS programs as the basis for the West Virginia Health Manpower Statistical System has been discussed. This is a rather ordinary type of data processing operation with strong statistical overtones. For this type of system, SAS-72 with PROC OUTPUT and PROC EDIT supplemented with PROC FORMS from SAS-75 has been very adequate.

The new SAS-75 has many of the features built in which were required as special add-ons for the WVHMSS job, such as the PUT statement, PROC FORMS, etc. This should make SAS-75 a more natural choice in the future for many more data processing applications which are strongly related to establishing or supporting data bases to be used primarily for statistical analyses.

References

[1] Borbash, Steven R., "Map Construction and Geolocational Analyses with SAS and a GEOZIP File", Industrial Engineering Department, Internal Report (Jan., 1976), West Virginia University, Morgantown, W.V. 26506.

[2] Chilko, Daniel M., "SAS Output Procedure", West Virginia University Computer Center Academic Services Instruction 42.24 (Sept., 1973), Morgantown, W.V. 26506.

[3] U.S. Department of Health, Education and Welfare, Geographical Location Codes, DHEW Publication 05-75-11 (Sept., 1975), Office of the Assistant Secretary, Comptroller, Washington, D.C.

[4] U.S. Department of Health, Education and Welfare, Public Health Service, Health Resources Administration, National Center for Health Statistics. News of the Cooperative Health Statistics System, Vol. 1, No. 1 (Jan.-Feb., 1974). National Center for Health Statistics, 5600 Fishers Lane, Rockville, Md. 20852.

[5] U.S. Postal Service, Directory of Post Offices (with ZIP Codes), (March, 1973), Office of Sales Support, U.S. Postal Service, Washington, D.C. 20262.